Proceedings of Statistics Canada Symposium 2002
Modelling Survey Data for Social and Economic Research

OPTIMAL CALIBRATION AND EMPIRICAL LIKELIHOOD METHODS IN SURVEY SAMPLING

C. Wu[1]

ABSTRACT

The calibration method has gained much popularity in recent literature on survey sampling, and calibration estimators are routinely computed by many survey organizations. The choice of calibration variables for all existing approaches, however, remains ad hoc. In this article we show that the model-calibration estimator for the finite population mean, which was proposed by Wu & Sitter (2001) through an intuitive argument, is indeed optimal among a class of calibration estimators. We further present optimal calibration estimators for the finite population distribution function, the population variance, variance of a linear estimator and other quadratic finite population functions under a unified framework. A limited simulation study shows that the improvement of these optimal estimators over the conventional ones can be substantial. The question of when and how can auxiliary information be used for both the estimation of the population mean using a generalized regression estimator and the estimation of its variance through calibration is addressed clearly under the proposed general methodology. Some fundamental issues in using auxiliary information from survey data are also addressed under the context of optimal estimation.

Some key words*:* Asymptotic design variance; Auxiliary information; Model calibration; Optimal estimation; Superpopulation.

1. INTRODUCTION

The notion of calibration estimators was introduced by Deville & Särndal (1992) in the context of using auxiliary information from survey data. Suppose U = {1, 2, ... , N} is the set of labels for the finite population. Let (y_i, x_i) be the values of the study variable y and the vector of auxiliary variables x attached to the ith unit. The question is how to effectively estimate $\bar{Y} = N^{-1}\sum_{i=1}^{N} y_i$ using the known population totals $X = \sum_{i=1}^{N} x_i$ at the estimation stage. Let s = {1, 2, ..., n} be the set of sampled units under a general sampling design, p, and $\pi_i = P(i \in s)$ be the first order inclusion probabilities. The conventional calibration estimator for $\bar{Y}$ is defined as $\hat{\bar{Y}}_C = N^{-1}\sum_{i\in s} w_i y_i$, where the w_i's are modified from the basic design weights $d_i = 1/\pi_i$ by minimizing a distance measure Φ_s between the w_i's and the d_i's subject to constraints

$$\sum_{i\in s} w_i x_i = \sum_{i=1}^{N} x_i \tag{1.1}$$

The most commonly used distance measure is the chi-squared distance

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

where the *qi's* are known positive constants uncorrelated with the d_i's. Alternative distance measures can also be considered. See Deville & Särndal (1992) for a detailed discussion.

There are two basic components in the construction of calibration estimators: a distance measure and a set of calibration equations. The choice of a distance measure is less critical since the resulting estimators are all

[1] Department of Statistics and Actuarial Science, University of Waterloo, cbwu@uwaterloo.ca

asymptotically equivalent to the one using a chi-squared distance with certain choice of q_i (Deville & Särndal, 1992). Calibration equations (1·1) are routinely used by many survey organizations and are referred to as benchmark constraints. Benchmark constraints are often imposed in practice for two reasons: (i) the surveyor believes that the weights which give perfect estimates for the auxiliary variables should also give a good estimate for the study variable; (ii) the auxiliary information is only available at the aggregate level, i.e. only X is known. Statistics practitioners in areas such as demography sometimes insist on benchmarking over lots of variables to match the known totals from census at the risk of worsening the efficiency of the estimators. On the other hand, if complete auxiliary information x_1, …, x_N is known, which is often the case in many survey problems, a very compelling question to ask would be "what is the best calibration equation to use in the construction of a calibration estimator?".
Let $ui = u(x_i)$, $i = 1, \ldots, N$, where u(·) is a real function. If we replace (1·1) by

$$\sum_{i \in s} w_i u(x_i) = \sum_{i=1}^{N} u(x_i), \tag{1.2}$$

then the question becomes "what kind of choice of $u(\cdot)$ will make $\hat{\bar{Y}}_C$ most efficient"? Note that the benchmark constraints (1·1) consist of k equations, where k is the number of components in x, while constraint (1·2) only has one equation using the single data-reduction variable $u = u(x)$. It has been shown by Wu & Sitter (2001, part (3) of Theorem 1) that the conventional calibration estimator for the finite population mean or total using (1·1) is identical to the calibration estimator using (1·2) with a special of $u(\cdot)$. The single equation constraint (1·2) with unspecified $u(\cdot)$ is more general than the fixed k equation benchmark constraints.

It is well-known that in survey sampling a uniformly minimum variance unbiased estimator does not exist under the design-based framework. Indeed the only choice of $u(\cdot)$ that results in a $\hat{\bar{Y}}_C$ with minimum variance is $u(x_i) \equiv y_i$, and this of course is practically useless.

The model-assisted optimal estimators using the criterion of minimum expected design variance $E_\xi\left\{V_p\left(\hat{\bar{Y}}\right)\right\}$ under a superpopulation have been discussed by several authors. See, for example, the work by Godambe (1955), Godambe & Thompson (1973), Cassel, Särndal & Wretman (1976), and Isaki & Fuller (1982). The expected design variance was also termed as "anticipated variance" by Isaki & Fuller (1982). Note that E_p and V_p refer to the expectation and the variance under the sampling design, p, and E_ξ and V_ξ denote the expectation and the variance under a superpopulation model, ξ.

In this article, we use a similar criterion. Calibration estimators belong to the class of non-linear estimators and their exact design variance or mean square error doesn't have a closed form. A natural replacement for optimality considerations is to minimize the model expectation of the asymptotic design variance $E_\xi\left\{AV_p\left(\hat{\bar{Y}}\right)\right\}$, where AV_p represents the design-based asymptotic variance. Since the bias $B_p\left(\hat{\bar{Y}}_C\right) = E_p\left(\hat{\bar{Y}}_C - \bar{Y}\right)$ of a calibration estimator $\hat{\bar{Y}}_C$ satisfies $B_p\left(\hat{\bar{Y}}_C\right) = o\left(n^{-1/2}\right)$ and $V_p\left(\hat{\bar{Y}}_C\right) = O\left(n^{-1}\right)$, minimizing $E_\xi\left\{AV_p\left(\hat{\bar{Y}}_C\right)\right\}$ is equivalent to minimizing $E_\xi\left\{E_p\left(\hat{\bar{Y}}_C - \bar{Y}\right)^2\right\}$ asymptotically.

In §2, we show that the model-calibration estimator for the finite population mean, which was proposed by Wu & Sitter (2001) through an intuitive argument, is indeed optimal among a class of calibration estimators in the sense of minimum expected asymptotic design variance under a superpopulation model and any regular sampling design. The result provides a unified framework for constructing optimal calibration estimators for the finite population distribution function, the population variance, variance of a linear estimator and other quadratic finite population functions. Optimal calibration estimators for the distribution function are presented in §3, and estimators for a general

second-order finite population quantity using optimal calibration are constructed in §4. Also in §4, the question of when and how can auxiliary information be used for both the estimation of the population mean using a generalized regression estimator and the estimation of its variance through calibration is addressed clearly under the unified framework. The optimal pseudo empirical maximum likelihood estimators, which are asymptotically equivalent to the optimal calibration estimators, are particularly useful in estimating the distribution function, the population variance and other known non-negative quantities. Results of a limited simulation study on the design-based finite sample performance of these optimal estimators with comparison to the conventional ones are reported in §5. Some fundamental issues in using auxiliary information from survey data are also addressed under this framework, and these together with some concluding remarks are given in §6.

2. THE OPTIMALITY OF THE MODEL-CALIBRATION ESTIMATOR

For asymptotic set-up, we assume there is a sequence of finite populations, indexed by v. The population size and sample size for the with population are denoted by N_v and n_v. As $v \to \infty$, $N_v \to \infty$ and $n_v \to \infty$. All limiting processes should be understood to mean $v \to \infty$. The index v will be suppressed to simplify notation. For a detailed formulation of this asymptotic framework, see Isaki & Fuller (1982). We consider situations where complete auxiliary information $x_1, \cdots, x_N$ is available.

Suppose $y_1, y_2, \cdots, y_N$ is a random sample from a superpopulation ξ such that

$$E_\xi(y_i \mid x_i) = \mu(x_i, \theta), V_\xi(y_i \mid x_i) = \{v(x_i)\}^2 \sigma^2, i = 1,2,..., N. \tag{2.1}$$

Here θ and σ^2 are model parameters and θ is possibly vector-valued, μ(·, ·) and v(·) are known functions, $y_1, y_2, ..., y_N$ are conditionally independent of each other for the given x_i's.

Let $\hat{\bar{Y}}_C$ be a calibration estimator of $\bar{Y}$ when $C = \{u(x_1), u(x_2), ...\}$ is used in *(1·2)*. Let ***L*** be the set of sequences $C = \{u(x_1), u(x_2), ...\}$ for all conceivable functions u(·) such that $N^{-1}\sum_{i=1}^{N}\{u(x_1)\}^2 \to c \neq 0$ as $N \to \infty$. This finite second moment condition on the sequence C Є ***L*** is not very restrictive and is needed in the proofs. We assume $\{\mu(x_1, \theta), \mu(x_2, \theta), ...\}$ Є ***L*** and $\{v(x_1), v(x_2), ...\}$ Є ***L***.

A sampling design is said to be regular if the design results in a fixed sample size, has inclusion probabilities π_i and π_{ij} independent of the response variable y, and satisfies

(C1) $\max_{i \in s} nd_i / N = O(1)$.

(C2) $N^{-1}\sum_{i \in s} d_i u_i - N^{-1}\sum_{i=1}^{N} u_i = O_p\left(n^{-1/2}\right)$ for any sequence $(u_1, u_2, ...)$ Є ***L***.

Condition (C1) simply states that no basic design weight is disproportionally large. Condition (C2) follows if the Horvitz-Thompson estimator for $\bar{u}_N = N^{-1}\sum_{i=1}^{N} u_i$ is asymptotically normally distributed.

Theorem 1. Among the class of calibration estimators $\hat{\bar{Y}}_C$ *with* $C = \{u(x_1), u(x_2), ...\}$ Є ***L****, the choice of* $C = \{\mu(x_1, \theta), \mu(x_2, \theta), ...\}$ *minimizes* $E_\xi\left\{AV_p\left(\hat{\bar{Y}}_C\right)\right\}$ *under the model (2·1) and any regular sampling design.*

Proof: See the Appendix.

In practice, the model parameter θ will have to be replaced by a sample-based estimate, $\hat{\theta}$. It can be shown that replacing θ by a design-consistent estimator $\hat{\theta}$ will not change the estimator $\hat{\bar{Y}}_C$ asymptotically. The resulting

estimator $\hat{\bar{Y}}_{MC}$ was termed by Wu & Sitter (2001) the model-calibration estimator of $\bar{Y}$. Wu & Sitter (2001) proposed $\hat{\bar{Y}}_{MC}$ through an intuitive argument and showed that it is a more general and efficient way to construct calibration estimators when complete auxiliary information is available. The optimality of $\hat{\bar{Y}}_{MC}$, however, was not investigated in that paper.

This optimal calibration approach can be applied to the pseudo empirical likelihood method (Chen & Sitter, 1999) to obtain an optimal estimator with attractive features. Let $\hat{\bar{Y}}_{EC}$ be the pseudo empirical maximum likelihood estimator of $\bar{Y}$ (Chen & Sitter, 1999) obtained by calibrating over $C = \{u(x_1), u(x_2), ...\}$. That is, $\hat{\bar{Y}}_{EC} = \sum_{i \in s} \hat{p}_i y_i$, where the $\hat{p}_i$'s maximize the pseudo empirical log-likelihood function $l(p) = \sum_{i \in s} d_i \log(p_i)$ subject to constraints

$$\sum_{i \in s} p_i = 1 \, (0 < p_i < 1), \; \sum_{i \in s} p_i u(x_i) = \frac{1}{N} \sum_{i=1}^{N} u(x_i). \tag{2.2}$$

The model-calibrated pseudo empirical maximum likelihood estimator $\hat{\bar{Y}}_{ME}$ of $\bar{Y}$ is obtained when $C = \{\mu(x_1, \hat{\theta}), \mu(x_2, \hat{\theta}), ...\}$ is used in the constraints *(2·2).*

Theorem 2. Among the class of pseudo empirical maximum lielihood estimators $\hat{\bar{Y}}_{EC}$ *for* $\bar{Y}$ *where C Є L, the choice of* $C = \{\mu(x_1, \theta), \mu(x_2, \theta), ...\}$ *minimizes* $E_\xi \left\{ AV_p \left(\hat{\bar{Y}}_{EC} \right) \right\}$ *under the model (2.1) and any regular sampling design.*

Proof: See the Appendix.

The model-calibrated pseudo empirical maximum likelihood estimator for $\bar{Y}$ is asymptotically equivalent to the model-calibration estimator and is optimal under the same context. Simple algorithms for computing the estimator $\hat{\bar{Y}}_{ME}$ have been developed by Chen *et al.* (2002). The most attractive feature of the estimator $\hat{\bar{Y}}_{ME}$, however, is the intrinsic properties of the weights: $\hat{p}_i > 0$ and $\sum_{i \in s} \hat{p}_i = 1$. This is particularly useful when the method is extended to estimate the distribution function and other known nonnegative quantities. The optimal calibration approach also provides a unified framework for the efficient estimation of the distribution function, the population variance and other quadratic finite population functions. This is detailed in §3 for the distribution function and in §4 for the estimation of variance and other quadratic functions.

3. OPTIMAL CALIBRATION ESTIMATORS FOR THE DISTRIBUTION FUNCTION

The finite population distribution function $F_Y(t) = N^{-1} \sum_{i=1}^{N} I(y_i \le t)$ is also a finite population mean defined over an indicator variable $z_i = I(y_i \le t)$. Here $z_i = 1$ if $y_i \le t$ and 0 otherwise. Without using any auxiliary information, estimation of $F_Y(t)$ is a special case of estimating the population mean and is usually straightforward. In the presence of auxiliary information, special attention needs be given to the following:

(a) While benchmark constraints (1·1) calibrated directly over the x variables sometimes are justifiable for the estimation of $\bar{Y}$, this consistency requirement is not needed for the estimation of $F_Y(t)$. Efficiency will be the primary concern.

(b) It is the indicator variable $z_i = I(y_i \le t)$ that we have to work with. There is also an issue of local efficiency (particular value of t) versus global efficiency (an arbitrary t) in estimating $F_Y(t)$.

(c) It is desirable that an estimator of $F_Y(t)$, say $\hat{F}_Y(t)$, is itself a distribution function, so quantile estimates can be obtained through direct inversion of $\hat{F}_Y(t)$.

Many techniques for estimating $\bar{Y}$, when applied directly to the estimation of $F_Y(t)$, will produce unsatisfactory results. For instance, in the case of a scalar x variable, a regression-type estimator for $F_Y(t)$ will have the form $\hat{F}_{RE}(t) = \hat{F}_Y(t) + \{F_X(t) - \hat{F}_X(t)\}\hat{B}$, where $\hat{F}_Y(t)$ and $\hat{F}_X(t)$ are Horvitz-Thompson type estimators for $F_Y(t)$ and $F_X(t) = N^{-1}\sum_{i=1}^{N} I(x_i \le t)$. The $\hat{B}$ is the estimated slope of regressing $I(y_i \le t)$ on $I(x_i \le t)$. $\hat{F}_{RE}(t)$ suffers from several drawbacks. The obvious one is that $\hat{F}_{RE}(t)$ is not a distribution function and it can take values outside of [0, 1].

The model-calibrated pseudo empirical likelihood method can be readily applied here to obtain estimators of $F_Y(t)$ which are not only efficient but also themselves genuine distribution functions. The optimal calibration variable $\mu(x_i, \theta)$ should now be replaced by $g(x_i, t) = E_\xi\{I(y_i \le t) \mid x_i\} = P(y_i \le t \mid x_i)$. Two types of working models can be considered to get $g(x_i, t)$: models that relate the y_i to the x_i or models that relate the indicator variable $I(y_i \le t)$ to the x_i.

Under the commonly used regression model,

$$y_i = x_i'\theta + v(x_i)\varepsilon_i,\ i = 1,2,\ldots,N, \tag{3.1}$$

where the ε_i's are independent and identically distributed random variates with mean 0 and variance σ^2. Let G(·) be the cumulative distribution function of the ε_i's. We have

$$g(x_i, t) = P(y_i \le t \mid x_i) = G\{(t - x_i'\theta)/v(x_i)\}.$$

As in the mean case, the model parameter θ will have to be replaced by a sample-based design-consistent estimate in applications.

Note that $g_i = g(x_i, t)$ are probabilities, an alternative modeling process is to use a generalized linear model for the binary observations $I(y_i \le t)$. For example, we may use a logistic regression model

$$\log\left(\frac{g_i}{1-g_i}\right) = x_i'\theta, \tag{3.2}$$

with the usual variance function $V(g) = g(1 - g)$. Under the model (3·2) we have $g(x_i, t) = \exp(x_i'\theta)/\{1+\exp(x_i'\theta)\}$. Let $\hat{F}_{EC}(t) = \sum_{i\in s}\hat{p}_i I(y_i \le t)$ where the $\hat{p}_i$'s maximize $l(p)$ subject to constraint (2·2) with $C = \{u(x_1), u(x_2), \ldots\}$.

Theorem 3. The pseudo empirical maximum likelihood estimator $\hat{F}_{ME}(t)$ calibrated over $\{g(x_1, t), g(x_2, t), \ldots\}$ is optimal among the class of estimators $\hat{F}_{EC}(t)$ with C Є ***L*** *under the working model (3·1) or (3·2) and any regular sampling design.*

Proof: The result follows directly from Theorem 2 if one replaces y_i *by* $I(y_i \le t)$ and $\mu(x_i, \theta)$ *by* $g(x_i, t)$.

The design-based properties and small sample performance of these estimators as well as the related quantile estimation problem have been investigated in a separate paper by Chen & Wu (2002). We will term the estimator of $F_Y(t)$ calibrated directly over the *x* variables (Chen & Sitter, 1999) as the conventional calibration estimator. As we will see in §5 that the optimal calibration estimators are much more efficient.

It should be noted that the two working models (3·1) and (3·2) discussed above are not compatible with each other. Optimality of the resulting estimator, therefore, is meaningful under the chosen model. It should also be noted that the optimal calibration variable $g(x_i, t)$ depends on t. *No* single set of weights $\hat{p}_i$ will produce an optimal estimator for an arbitrary t. Chen & Wu (2002) suggest to use a fixed t_o in $g(x_i, t)$ while the resulting weights are used for any t in $\hat{F}_{ME}(t)$. Under this treatment $\hat{F}_{ME}(t)$ *is* a genuine distribution function. Chen & Wu (2002) demonstrate through a simulation study that the so-constructed $\hat{F}_{ME}(t)$ *is* very efficient for values of t in a wide neighborhood of t_o. The actual value of t_o can be easily determined to maximize the efficiency of the resulting estimator when certain neighborhood of t_o is of interest.

4. OPTIMAL ESTIMATION OF VARIANCE AND OTHER QUADRATIC FUNCTIONS

Estimation of variance and other second-order finite population quantities using auxiliary information has been addressed by many survey researchers. Various techniques, such as regression, ratio and calibration estimation, have been attempted. See Sitter & Wu (2002) for a literature review. A common weakness of these approaches is the ad hoc argument of applying certain techniques, which were originally developed for estimating $\bar{Y}$, to estimate the variance or other second-order population parameters without a common framework that unifies the two types of finite population parameters.

The optimal model-calibration and the model-calibrated pseudo empirical likelihood methods can be extended to handle variances and other second-order finite population parameters through a batch approach. For parameters in a general form of $Q = \sum_{i=1}^{N}\sum_{j=i+1}^{N}\phi(y_i, y_j)$, which include the population variance $S^2 = N^{-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2 = \{N(N-1)\}^{-1}\sum_{i=1}^{N}\sum_{j=i+1}^{N}(y_i - y_j)^2$ and the variance of the Horvitz-Thompson estimator $V_p(\hat{Y}_{HT}) = \sum_{i=1}^{N}\sum_{j=i+1}^{N}(\pi_i\pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)^2$ as special cases, a unified estimation strategy can be developed as follows.

Q can be viewed as a total over a synthetic finite population, i.e. $Q = \sum_{\alpha=1}^{N^*} t_\alpha$ where $\alpha = (ij) = 1,2,\cdots,N^*$, $t_\alpha = \phi(y_i, y_j)$ for $\alpha = (ij)$, and $N^* = N(N-1)/2$ is the total number of pairs. The sample data over the synthetic population consists of all the pairs from the original sample: $s^* = \{(ij) : i < j, i, j \in s\}$. The "first-order" inclusion probabilities under this setting are $\pi_{ij} = P(i, j \in s)$, and the "basic design weights" are $d_{ij} = 1/\pi_{ij}$. The term $\mu(x_i, \theta) = E_\xi(y_i|x_i)$ should now be replaced *by* $E_\xi\{\phi(y_i, y_j)|x_i, x_j\}$.

Using the original pair index (ij), the model-calibration estimator of Q is defined as $\hat{Q}_{MC} = \sum_{i\in s}\sum_{j>i} w_{ij}\phi(y_i, y_j)$ where the weights w_{ij} minimize the modified chi-squared distance measure

$$\Phi_{s^*} = \sum_{i\in s}\sum_{j>i}(w_{ij} - d_{ij})^2/(d_{ij}q_{ij})$$

subject to

$$\sum_{i\in s}\sum_{j>i} w_{ij}E_\xi\{\phi(y_i, y_j)|x_i, x_j\} = \sum_{i=1}^{N}\sum_{j=i+1}^{N}E_\xi\{\phi(y_i, y_j)|x_i, x_j\}. \tag{4.1}$$

Let $\hat{Q}_C$ be a calibration estimator of Q when $C^* = \{u(x_i, x_j), i, j = 1, 2, \cdots\}$ *is* used in (4·1) as the calibration variable. Let $\boldsymbol{L}^*$ be the set of all possible sequence $C^* = \{u(x_i, x_j), i, j = 1, 2, \cdots\}$ satisfying a finite second moment condition similar to the one in defining $\boldsymbol{L}$. If we re-define the regular sampling design by replacing the d_i's in (C1) and *(C2)* of *§2* by d_{ij} with suitable reformulation, we have the following result.

Theorem 4. *Among the class of calibration estimators* $\hat{Q}_C$ *with* $C^* = \{u(x_i, x_j), i, j = 1, 2, \cdots\} \in \boldsymbol{L}^*$, *the model-calibration estimator* $\hat{Q}_{MC}$ *attains the minimum value of* $E_\xi\{AV_p(\hat{Q}_C)\}$ *under the model (2·1) and any regular sampling design.*

Proof: The result of Theorem 1 does not apply directly here due to a weak correlation among the sequence of $t_\alpha = \phi(y_i, y_j)$, $\alpha = (ij) = 1, 2, \cdots, N^*$: t_α and $t_{\alpha'}$ are not independent of each other under the mode (2·1) if $\alpha = (ij)$ and $\alpha' = (lm)$ have one index in common. Since the total number of pairs $(t_\alpha, t_{\alpha'})$ with possible non-zero covariance is of order $O(N^3) = O\{(N^*)^{3/2}\}$ and the total number of zero-covariance pairs is of order $O\{(N^*)^2\}$, using a similar notation as in the proof of Theorem 1, it can be shown that $E_p\{V_\xi(T_1)\} = O\{(N^*)^{-1}\}$ and $E_p\{V_\xi(T_2)\} = O\{(n^*)^{-1}(N^*)^{-1/2}\}$. The rest of the proof follows directly from that of Theorem 1.

The pseudo empirical (log) likelihood function is modified to accommodate all the pairs (ij) using the d_{ij}'s. Let

$$l^*(p) = \sum_{i\in s}\sum_{j>i} d_{ij}\log(p_{ij}).$$

The model-calibrated pseudo empirical maximum likelihood estimator of Q is defined as $\hat{Q}_{ME} = N^*\sum_{i\in s}\sum_{j>i}\hat{p}_{ij}\phi(y_i, y_j)$ where the $\hat{p}_{ij}$'s maximize $l^*(p)$ subject to

$$\sum_{i\in s}\sum_{j>i} p_{ij} = 1(p_{ij} > 0), \sum_{i\in s}\sum_{j>i} p_{ij} E_\xi\{\phi(y_i, y_j) \mid x_i, x_j\} = \frac{1}{N^*}\sum_{i=1}^{N}\sum_{j=i+1}^{N} E_\xi\{\phi(y_i, y_j) \mid x_i, x_j\}. \quad (4.2)$$

A theorem that is parallel to Theorem *2* regarding the optimality of $\hat{Q}_{ME}$ can be similarly established.

As usual, any model parameters appearing in the constraints *(4·1)* or *(4·2)* will be replaced by sample-based design-consistent estimates. We discuss some of the details of the proposed estimators for the population variance S^2 in *§4.1*. The question of when and how can auxiliary information be used for both the estimation of population mean using a generalized regression estimator and the estimation of its variance through calibration is answered under this unified approach in *§4.2*.

4·1 Estimation of the population variance

Note that the population variance can be re-written as $S^2 = \{N(N-1)\}^{-1}\sum_{i=1}^{N}\sum_{j=i+1}^{N}(y_i - y_j)^2$. Under the model *(2·1)*, $E_\xi\{(y_i - y_j)^2 \mid x_i, x_j\} = \{\mu(x_i, \theta) - \mu(x_j, \theta)\}^2 + \sigma^2\{v^2(x_i) + v^2(x_j)\}$, and this should be used in the constraint *(4·1)* for optimal estimation. One can also sufficiently replace *(4·1)* by the following two equations.

$$\sum_{i\in s}\sum_{j>i} w_{ij}\{\mu(x_i, \theta) - \mu(x_j, \theta)\}^2 = \sum_{i=1}^{N}\sum_{j=i+1}^{N}\{\mu(x_i, \theta) - \mu(x_j, \theta)\}^2 \quad (4.3)$$

$$\sum_{i\in s}\sum_{j>i} w_{ij}\{v^2(x_i) + v^2(x_j)\} = \sum_{i=1}^{N}\sum_{j=i+1}^{N}\{v^2(x_i) + v^2(x_j)\}. \quad (4.4)$$

In many applications $v(x_i) \equiv 1$, in this case the second calibration equation *(4·4)* becomes $\sum_{i\in s}\sum_{j>i} w_{ij} = N^*$. The resulting estimator $\hat{S}^2_{MC}$ reduces to the one proposed by Sitter & Wu *(2002)*. Under a linear working model where $\mu(x_i, \theta) = x_i'\theta$, this estimator has a neat form of $\hat{S}^2_{MC} = \hat{S}^2_{HT} + \hat{\theta}'(S_x^2 - S_x^2)\hat{\theta}\,\hat{B}$, where

$\hat{S}_{HT}^2 = \{N(N-1)\}^{-1}\sum_{i\in s}\sum_{j>i} d_{ij}(y_i - y_j)^2$, $S_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \bar{X})(x_i - \bar{X})'$ $s_x^2 = \{N(N-1)\}^{-1}\sum_{i\in s}\sum_{j>i} d_{ij}(x_i - x_j)(x_i - x_j)'$, and $\hat{B}$ is the estimated regression coefficient of regressing $v_{ij} = (y_i - y_j)^2$ over $u_{ij} = \hat{\theta}'(x_i - x_i)(x_i - x_j)'\hat{\theta}$.

The model-calibrated pseudo empirical maximum likelihood estimator is more useful in this context. Note that under the model (2·1) the constraints (4·2) can be replaced by

$$\sum_{i\in s}\sum_{j>i} p_{ij} = 1 \quad (p_{ij} > 0) \tag{4.5}$$

$$\sum_{i\in s}\sum_{j>i} p_{ij}\{\mu(x_i,\theta) - \mu(x_j,\theta)\}^2 = \frac{1}{N*}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\{\mu(x_i,\theta) - \mu(x_j,\theta)\}^2 \tag{4.6}$$

$$\sum_{i\in s}\sum_{j>i} p_{ij}\{v^2(x_i) + v^2(x_j)\} = \frac{1}{N*}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\{v^2(x_i) + v^2(x_j)\} \tag{4.7}$$

When $v(x_i) \equiv 1$, the last equation (4·7) reduces to (4·5), and the resulting estimator $\hat{S}_{ME}^2$ also reduces to the one proposed by Sitter & Wu (2002). A simple and stable algorithm for computing the weights $\hat{p}_{ij}$ is described in Sitter & Wu (2002). Since $\hat{p}_{ij} > 0$, the model-calibrated pseudo empirical maximum likelihood estimator is always positive which is desirable for practical applications.

Under a model with non-homogeneous variance, including the constraint (4.4) or (4.7) will usually improve the efficiency of the resulting estimators, as shown by a simulation study reported in §5.

4·2 Variance estimation for the generalized regression estimator

The generalized regression estimator for the population total (or mean) is one of the most popular techniques in using auxiliary information from surveys. Assuming the totals *X* are known, the generalized regression estimator for *Y* is computed as $\hat{Y}_{GR} = \hat{Y}_{HT} + (X - \hat{X}_{HT})'\hat{\theta}$, where $\hat{Y}_{HT} = \sum_{i\in s} d_i y_i$ and $\hat{X}_{HT} = \sum_{i\in s} d_i x_i$ are the conventional Horvitz-Thompson estimators, $\hat{\theta}$ is the estimated regression coefficient of *y* over *x*. Its asymptotic design variance is given by

$$AV_p(\hat{Y}_{GR}) = \sum_{i=1}^{N}\sum_{j=i+1}^{N}(\pi_i\pi_j - \pi_{ij})\left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j}\right)^2 ,$$

where $e_i = y_i - x_i'\theta_N$ and θ_N is the finite population regression coefficient that is estimated by $\hat{\theta}$.

The question of when and how can auxiliary information be used for both the estimation of the population total using a generalized regression estimator and the estimation of its variance can now be answered clearly under the optimal model-calibration approach. Note that $AV_p(\hat{Y}_{GR})$ has the form of Q with $\phi(y_i, y_j) = (\pi_i\pi_j - \pi_{ij})(e_i/\pi_i - e_j/\pi_j)^2$. Under the model (3.1) which is the one that motivated the generalized regression estimator, the optimal calibration variable that should be used in (4·1) is

$$E_\xi\{\phi(y_i, y_j) \mid x_i, x_j\} \approx (\pi_i\pi_j - \pi_{ij})\left\{\frac{v^2(x_i)}{\pi_i^2} + \frac{v^2(x_j)}{\pi_j^2}\right\}\sigma^2 .$$

Here we have used the fact that $E_\xi(e_i) = 0$. It is now clear that if the model (3·1) has a homogeneous variance structure,

i.e. $v(x_i) \equiv 1$, the calibration variable will be independent of the x_i's. The same auxiliary information cannot be used to improve the variance estimation for the generalized regression estimator. On the other hand, under a linear regression model with non-homogeneous variance, there will be room for improvement. The constraint that should be used to construct the model-calibration estimator is given by

$$\sum_{i\in s}\sum_{j>i} w_{ij}\left(\pi_i\pi_j-\pi_{ij}\right)\left\{\frac{v^2(x_i)}{\pi_i^2}+\frac{v^2(x_j)}{\pi_j^2}\right\}=\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(\pi_i\pi_j-\pi_{ij}\right)\left\{\frac{v^2(x_i)}{\pi_i^2}+\frac{v^2(x_j)}{\pi_j^2}\right\}.$$

Similar constraint should be used when one estimates $AV_p\left(\hat{Y}_{GR}\right)$ using the model calibrated pseudo empirical likelihood method.

5. A SIMULATION

The optimality of the proposed estimators is established under the true model with large samples. In this section a limited simulation study has been conducted to investigate the design-based finite sample performance of these estimators with comparison to the conventional calibration estimators.

In the simulation, a fixed finite population of size $N = 2000$ was generated from the regression model $y_i=\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}\varepsilon_i$ where the x_{1i}'s and the x_{2i}'s follow a standard gamma distribution and a standard log-normal distribution, respectively. Both auxiliary variables take non-negative values and are skewed to the right which is quite common in real applications. Note that $v(x_i) = x_{1i}$ is used in the model. The β_i's are all conveniently set to be 1. The ε_i's are independent and identically distributed as $N(0,\sigma^2)$. Four different values of σ^2 are chosen such that the finite population correlation coefficient p between y and $\beta_0+\beta_1 x_1+\beta_2 x_2$ are 0.9, 0.8, 0.7 and 0.6, respectively. At each simulation run, a simple random sample of size n = 100 is taken first from the finite population, the model parameters (β_0, β_1, β_2) and σ^2 are estimated using the usual least square method, and various estimators are computed from the sample data. The process is repeated B = 1000 times.

Under the linear regression model, the optimal model-calibration estimator of the population mean calibrated over $u_i = x_i'\hat{\beta}$ reduces to the conventional calibration estimator using constraints (1.1) (Wu & Sitter, 2001) and hence is not included in the simulation.

For the distribution function $F_Y(t)$, four estimators are computed: the optimal model-calibrated empirical likelihood estimator $\hat{F}_{ME1}(t)$ under the regression model (3.1), the optimal model-calibrated empirical likelihood estimator $\hat{F}_{ME2}(t)$ under the logistic regression model (3·2), the pseudo empirical maximum likelihood estimator $\hat{F}_{CC}(t)$ under the conventional calibration equation $\sum_{i\in s} p_i x_i = \bar{X}$, and the baseline Horvitz-Thompson estimator $\hat{F}_{HT}(t)$. All estimators are computed at five different population quantiles t_α, with α = 0.10, 0.30, 0.50, 0.70, and 0.90, and the optimal weights $\hat{p}_i$ are computed using the particular value of t_α. The performance of an estimator $\hat{F}(t)$ is evaluated by the Relative Percentage Bias (RB%) and the Relative Efficiency (RE) defined as

$$RB\% = \frac{1}{B}\sum_{b=1}^{B}\frac{\hat{F}_b(t)-F_Y(t)}{F_Y(t)} \text{ and } RE+\frac{MSE\left(\hat{F}_{HT}(t)\right)}{MSE\left(\hat{F}(t)\right)},$$

where $MSE\left(\hat{F}(t)\right)=B^{-1}\sum_{b=1}^{B}\left[\hat{F}_b(t)-F_Y(t)\right]^2$ and $\hat{F}_b(t)$ is computed from the bth simulation run. The Horvitz-Thompson estimator is used for baseline comparison.

Table 1. *Relative Efficiency of Estimators for the Distribution Function*

α	0.10	0.30	0.50	0.70	0.90
$\hat{F}_{ME1}(t)$	1.15	1.79	2.17	1.99	2.18
$\hat{F}_{ME2}(t)$	1.07	1.43	1.72	1.73	1.85
$\hat{F}_{CC}(t)$	1.04	1.12	1.27	1.39	1.53
$\hat{F}_{HT}(t)$	1.00	1.00	1.00	1.00	1.00

Table 1 summarizes the simulated relative efficiency for the population with $p = 0.80$. Results from other values of p demonstrate similar pattern as in Table 1 with reduced relative efficiency for all three estimators as p decreases. The simulated relative percentage biases are all within 2%. The optimal estimator $\hat{F}_{ME1}(t)$ under the regression model outperforms all others at all population quantiles and the gain over the conventional calibration estimator $\hat{F}_{CC}(t)$ can be very substantial. The estimator $\hat{F}_{ME2}(t)$ under a logistic regression model performs reasonably well, but it is slightly less efficient compared with $\hat{F}_{ME1}(t)$ which is computed under the true regression model. One advantage of using $\hat{F}_{ME2}(t)$ is that the variance function *v(x)* which plays a crucial role under the regression model is not an issue here, and the logistic regression model provides an attractive alternative for many real world applications.

As for the population variance, we compute the model-calibration estimator $\hat{S}^2_{MC1}$ and the model-calibrated empirical likelihood estimators $\hat{S}^2_{ME1}$ where the single constraint (4·3) or (4·6) is used; when two constraints (4·3) and (4·4), or (4·6) and (4·7), are both used, the resulting estimators are denoted by $\hat{S}^2_{MC2}$ and $\hat{S}^2_{ME2}$. The relative percentage bias and relative efficiency are similarly defined and comparisons are made with the baseline Horvitz-Thompson estimator $\hat{S}^2_{HT}$.

Table 2. *Relative Efficiency of Estimators for the Population Variance*

p	$\hat{S}^2_{MC1}$	$\hat{S}^2_{ME1}$	$\hat{S}^2_{MC2}$	$\hat{S}^2_{ME2}$	$\hat{S}^2_{HT}$
0.90	6.36	5.45	7.05	4.27	1.00
0.80	2.27	2.42	3.17	2.34	1.00
0.70	1.25	1.56	2.67	1.95	1.00
0.60	0.89	1.02	2.01	1.83	1.00

The simulated relative efficiency of each estimators for all four values of p are reported in Table 2. The absolute values of the simulated relative percentage bias are all less than 4% and are not reported here to save space. The estimators $\hat{S}^2_{MC1}$ and $\hat{S}^2_{ME1}$ perform extremely well when the single calibration variable $\mu(x_i, \hat{\beta}) = x_i' \hat{\beta}$ is a strong predictor of the response variable (i.e. high value of p), but it could perform badly when such a relationship is weak (eg. $p = 0.60$). The optimal estimators $\hat{S}^2_{MC2}$ and $\hat{S}^2_{ME2}$ which use auxiliary information from both the mean function $\mu(x_i, \hat{\beta})$ and the variance function $v(x_i)$ perform well for all cases, and their loss of efficiency due to the reduced value of p is less dramatic.

6. CONCLUDING REMARKS

In recent literature, there is a significant amount of work on using auxiliary information from surveys at the estimation stage. It has been shown that an effective way of doing this is to use a constrained minimization (such as the calibration method) or a constrained maximization (such as the empirical likelihood method) of an objective function. For most existing approaches, however, the choice of the calibration variables used in those constraints remains ad hoc. This is particularly the case when the finite population distribution function or a second-order population quantity such as the population variance is to be estimated.

The proposed optimal calibration approach requires specification of the mean function $\mu(x_i, \theta)$ and/or the variance function $v(x_i)$ from the model. A general discussion on model building and diagnostics using complex survey data is beyond the scope of this paper and requires further research. In many applications, the parametric linear regression model (3·1) will likely be used. In the case of a single x variable, Breidt & Opsomer (2000) used nonparametric smoothing technique to find the model expectations of the response variable. Extending the method to multiple x variables seems possible.

It should be noted that the model-assisted optimal calibration estimators take advantage of the model but are not excessively dependent on the model's accuracy. The proposed estimators are most efficient under a model that adequately describes the survey population and remain design consistent even if the model is misspecified. The optimality results, which are established under the conceptually assumed true model, provide practical guidance for the construction of calibration estimators under the working model.

An important feature of the results presented in this paper is that the optimality of the model-calibration or the model-calibrated pseudo empirical maximum likelihood estimators is independent of the sampling design. This is in contrast to the results of Godambe & Thompson (1973), or Cassel *et al.* (1976), where an optimal estimator is paired to a particular sampling design. The independence of the optimality of an estimator to sampling design is practically appealing when such an estimator is to be constructed at the estimation stage.

Our results also provide a unified framework for optimal estimation of the population mean or total, the distribution function, the population variance, variance of a linear estimator or other second-order finite population quantities. Some fundamental issues in using auxiliary information from surveys can now be addressed more clearly following this optimal calibration approach:

(i) The effective use of auxiliary information from survey data depends on both the parameters to be estimated and the actual relationship between the response variable and the covariates. Blindly calibrating over auxiliary variables is usually not a good approach.

(ii) The benchmark constraints used in (1·1) are justifiable if the relationship between y and x is close to linear and the parameter of interest is the population mean or total. In this case the resulting (conventional) calibration estimator of $\bar{Y}$ is identical to the optimal model-calibration estimator obtained using $\hat{\mu}_i = x_i' \hat{\theta}$ as the calibration variable (Wu & Sitter, 2001). So benchmarking implies efficient estimation.

(iii) If the relationship between y and x is linear, knowing $\bar{X}$ is "sufficient" for efficient estimation of the population mean $\bar{Y}$ or the total Y. If the relationship is nonlinear, or the parameters of interest involve a nonlinear function, complete auxiliary information and/or more advanced modeling are essential for "optimal" estimation.

(iv) The variance function $v(x_i)$ from model (2·1) does not play a role in the construction of optimal calibration estimators for the population mean or total. This, however, is not the case for the optimal estimation of the finite population distribution function, the population variance or other second-order population quantities where $v(xi)$ is equally important as the mean function $\mu(x_i, \theta)$.

(v) Auxiliary information can sometimes be triply used at the design stage, the estimation of the population mean or total using a generalized regression estimator, and the estimation of its variance through calibration. Such situations can be identified under the optimal calibration approach.

For cases where complete auxiliary information is required for optimal estimation but such an information is not available, the optimal calibration approach can be combined with two-phase sampling where the large first phase sample measured over the covariates is treated as "complete" auxiliary information. Some practical issues and the gain of efficiency from using this approach are currently under investigation.

ACKNOWLEDGEMENT

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The author thanks Professor Randy R. Sitter, Dr. Steve Drekic, the Associate Editor and two referees for helpful comments and suggestions that greatly improved the paper.

APPENDIX: PROOFS OF THEOREMS 1 AND *2*

Proof of Theorem 1: Without loss of generality, we consider the chi-squared distance measure with the weights q_i satisfying $N^{-1}\sum_{i=1}^{N} q_i^2 = O(1)$ and $q_i \geq q$ for some constant $q > 0$. It can be easily show that minimizing Φ_s subject to *(1·2)* leads to

$$\hat{\bar{Y}}_C = \frac{1}{N}\sum_{i\in s} d_i y_i + \frac{1}{N}\left(\sum_{i=1}^{N} u_i - \sum_{i\in s} d_i u_i\right)\hat{B},$$

where $u_i = u(x_i)$, $\hat{B} = \left(\sum_{i\in s} d_i q_i u_i y_i\right)/\left(\sum_{i\in s} d_i q_i u_i^2\right)$.

Under a regular sampling design, $AV_p\left(\hat{\bar{Y}}_C\right) = V_p(T)$, where

$$T = \frac{1}{N}\sum_{i\in s} d_i y_i + \frac{1}{N}\left(\sum_{i=1}^{N} u_i - \sum_{i\in s} d_i u_i\right)B_N ,$$

and $B_N = \left(\sum_{i=1}^{N} u_i q_i y_i\right)/\left(\sum_{i=1}^{N} q_i u_i^2\right)$.

Let $\mu_i = \mu(x_i, \theta)$, $\bar{\mu} = E_\xi(\bar{Y}) = N^{-1}\sum_{i=1}^{N}\mu_i$, $B_\xi(T) = E_\xi(T) - \bar{\mu}$. Since $E_p(T) = \bar{Y}$, $V_p(T) = E_p\left(T - \bar{Y}\right)^2$, it is straightforward to show that

$$E_\xi\{V_p(T)\} = E_p\{V_\xi(T)\} + E_p\left[\{B_\xi(T)\}^2\right] - V_\xi(\bar{Y}).$$

Note that E_ξ and V_ξ are conditional for the given x_i's. Let $U^2 = N^{-1}\sum_{i=1}^{N} q_i u_i^2$, $D = N^{-1}\left(\sum_{i=1}^{N} u_i - \sum_{i\in s} d_i u_i\right)$. We can rewrite T as $T_1 + T_2$, where $T_1 = N^{-1}\sum_{i\in s} d_i y_i$, $T_2 = DU^{-2}N^{-1}\sum_{i=1}^{N} q_i u_i y_i$. We have

$$E_p\{V_\xi(T)\} = E_p\{V_\xi(T_1)\} + E_p\{V_\xi(T_2)\} + 2E_p\{COV_\xi(T_1, T_2)\},$$

where $COV_\xi(T_1, T_2)$ denotes the covariance under the model. It can be seen that

$$E_p\{V_\xi(T_1)\} = \frac{1}{N^2}\sum_{i=1}^{N} d_i v^2(x_i)\sigma^2 = O\left(\frac{1}{N}\right),$$

$$E_p\{V_\xi(T_2)\} = \{E_p(D^2)\}U^{-4}\frac{1}{N^2}\sum_{i=1}^{N} q_i^2 u_i^2 v^2(x_i)\sigma^2 = O\left(\frac{1}{nN}\right).$$

Here we have used the facts that $\max_{i\in s} nd_i / N = O(1)$, $N^{-1}\sum_{i=1}^{N} v^2(x_i) = O(1)$, $E_p(D^2) = V_p\left(N^{-1}\sum_{i\in s} d_i u_i\right) = O(n^{-1})$, $N^{-1}\sum_{i=1}^{N} q_i^2 = O(1)$, $q_i \geq q > 0$ for all *i*, and $N^{-1}\sum_{i=1}^{N} q_i u_i^2 \to c^* \neq 0$.

It also follows from $|COV_\xi(T_1, T_2)| \leq \{V_\xi(T_1)\}^{1/2}\{V_\xi(T_2)\}^{1/2}$ that

$$\{E_p|COV_\xi(T_1, T_2)|\}^2 \leq E_p\{V_\xi(T_1)\}E_p\{V_\xi(T_2)\},$$

which implies $E_p\{COV_\xi(T_1, T_2)\} = O(n^{-3/2})$. When *n is* large, the leading term in $E_p\{V_\xi(T)\}$ is $E_p\{V_\xi(T_1)\}$, which is independent of the choice of the sequence, *C*. The term $V_\xi(\bar{Y})$ is also independent of *C*.

For the term $E_p\left[\{B_\xi(T)\}^2\right]$, note that

$$B_\xi(T) = \frac{1}{N}\sum_{i\in s} d_i(\mu_i - u_i B) - \frac{1}{N}\sum_{i=1}^{N}(\mu_i - u_i B),$$

where $B = \sum_{i=1}^{N} q_i u_i \mu_i / \sum_{i=1i}^{N} q_i u_i^2$. It follows that $E_p\{B_\xi(T)\} = 0$ and

$$E_p\left[\{B_\xi(T)\}^2\right] = V_p\{B_\xi(T)\} = V_p\left\{N^{-1}\sum_{i\in s} d_i(\mu_i - u_i B)\right\} = O(n^{-1}).$$

Minimizing $E_\xi\left\{AV_p\left(\hat{\bar{Y}}_C\right)\right\}$ amounts to minimizing $E_p\left[\{B_\xi(T)\}^2\right]$. The choice of $C = (\mu_1, \mu_2, ...)$ results in *B = 1* and $E_p\left[\{B_\xi(T)\}^2\right] = 0$.

Proof of Theorem 2: By Theorem 1 of Chen & Sitter (1999), we have

$$\hat{\bar{Y}}_{ME} = \left(\sum_{i\in s} d_i\right)^{-1}\sum_{i\in s} d_i y_i + \left\{\frac{1}{N}\sum_{i=1}^{N} u(x_i) - \left(\sum_{i\in s} d_i\right)^{-1}\sum_{i\in s} d_i u(x_i)\right\}\hat{B} + o_p\left(n^{-1/2}\right),$$

where $\hat{B}$ is similarly defined as in Theorem 1 with $q_i = 1$.

The term $T_1^* = \left(\sum_{i\in s} d_i\right)^{-1}\sum_{i\in s} d_i y_i$ is a ratio type estimator and its design-based variance $V_p\left(T_1^*\right)$ is not the same as $V_p(T_1)$, where $T_1 = N^{-1}\sum_{i\in s} d_i y_i$. However, since $\sum_{i\in s} d_i$ is a constant under the superpopulation model, the conclusion about $E_p\{V_\xi(T_1)\}$ in Theorem 1 can also be restated here in terms of T_1^*. The remaining part of the proof is similar to the proof of Theorem 1 and is omitted.

REFERENCES

Breidt, F.J. and Opsomer, J.D. (2000). Local polynomial regression estimators in survey sampling. *The Annals of Statistics.* 28, 1026-1053.

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 Sitter, R.R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*. 9, 385-406.

Chen**,** J., Sitter, R.R. and Wu, C. (2002). Using empirical likelihood method to obtain range restricted weights in regression estimators for surveys. *Biometrika*, 89, 230-237.

Chen, J. and Wu, C. (2002). Estimation of distribution function and quantiles using the model-calibrated pseudo empirical likelihood method. *Statistica Sinica***,** 12, 1223-1239.

Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. J. *Am. Statist. Assoc.* 87, 376-82.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. J. *Roy. Statist.* Soc.. Ser. B, 17, 267-278.

Godambe, V.P. and Thompson, M.E. (1973). Estimation in sampling theory with exchangeable prior distributions. *The Annals of Statistics.* 1, 1212-1221.

Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. J. *Am. Statist. Assoc.* 77, 89-96.

Sitter, R.R. and Wu, C. (2002). Efficient estimation of quadratic finite population functions in the presence of auxiliary information. J. *Am. Statist. Assoc.*, 97, 535-543.

Wu, C. and Sitter, R.R. (2001). A model-calibration approach to using complete auxiliary information from survey data. J. *Am. Statist. Assoc.* 96, 185-93.