

# Efficient Estimation of Quadratic Finite Population Functions in the Presence of Auxiliary Information

Randy R. SITTER and Changbao WU

---

By viewing quadratic and other second-order finite population functions as totals or means over a derived synthetic finite population, we show that the recently proposed model calibration and pseudoempirical likelihood methods for effective use of auxiliary information from survey data can be readily extended to obtain efficient estimators of quadratic and other second-order finite population functions. In particular, estimation of a finite population variance, covariance, or variance of a linear estimator can be greatly improved when auxiliary information is available. The proposed methods are model assisted in that the resulting estimators are asymptotically design unbiased irrespective of the correctness of a working model but very efficient if the working model is nearly correct. They have a number of attractive features, which include applicability to a general sampling design, incorporation of information on possibly multivariate auxiliary variables, and the ability to entertain linear or nonlinear working models, and they result in nonnegative estimates for certain strictly positive quantities such as variances. Several existing estimators are shown to be special cases of the proposed general methodology under a linear working model.

KEY WORDS: Generalized regression estimator; Model-assisted approach; Model calibration; Pseudoempirical likelihood; Survey sampling; Variance estimation.

---

## 1. INTRODUCTION

The problem of estimating a finite population mean or total in the presence of auxiliary information has been extensively discussed in survey sampling. Although a purely model-based prediction approach has been used by some researchers, the model-assisted approach has gained much popularity in recent literature. Several general procedures have been proposed, including generalized regression estimators (Cassel, Särndal, and Wretman 1976; Särndal 1980), calibration estimators (Deville and Särndal 1992), empirical likelihood methods (Chen and Qin 1993; Chen and Sitter 1999; Zhong and Rao 2000), and, more recently, the model calibration (MC) and model-calibrated pseudoempirical likelihood (PEML) methods (Wu and Sitter 2001).

Estimation of quadratic or other higher-order finite population functions is also important. For example, efficient estimators for finite population variances, covariances between two response variables, or variances of linear estimators are highly desirable. Shah and Patel (1996) presented several examples to illustrate why the estimation of population variances and covariances might be useful in their own right. However, because of the relative complexity of these functions, it is not obvious that one can obtain more efficient estimators for these higher-order population quantities when certain auxiliary information is available from survey data.

The literature in this area can be roughly categorized as either of the following, depending on the auxiliary information available: case 1, in which the population total  $X$  or mean  $\bar{X}$  of a single auxiliary variable  $x$  is known, or case 2, in which the population variances and covariances,  $S_{ij}$ , of one or several auxiliary variables are known. In case 1, a common strategy used by several authors is to restrict to a specific class of estimators, say  $v_h = h(\bar{X}/\bar{x})v$ , where  $\bar{x}$  is the sample mean of the

$x$  variable,  $h(\cdot)$  is a smooth function satisfying  $h(1) = 1$ , and  $v$  is a conventional variance estimator not using any auxiliary information. The objective is then to find an "optimal" estimator within this class (see Das and Tripathi 1978; Deng and Wu 1987, and references cited therein). Zhang (1996) discussed the estimation of a population variance of  $y$ ,  $\sigma_0^2$ , with population mean  $\mu_0$  known in the context of infinite populations (iid observations). His method can be viewed as a special case of Chen and Qin's (1993) empirical likelihood method. Datta and Ghosh (1993) introduced a hierarchical Bayes procedure for estimating the strata variances in the presence of auxiliary information.

In case 2, a similar strategy has been used by considering certain classes of estimators that incorporate the auxiliary information in a particular way. Examples include a ratio- or product-type estimator,  $v_g = (S_x^2/s_x^2)^g v$ , a regression-type estimator,  $v_{reg} = v + \hat{\beta}^2(S_x^2 - s_x^2)$ , and some multivariate generalizations of this kind, where  $S_x^2$  and  $s_x^2$  are the population and sample variances of variable  $x$ . When both  $\bar{X}$  and  $S_x^2$  are known, a more general class, say  $v_u = u(\bar{X}/\bar{x}, S_x^2/s_x^2)v$ , may be considered, where  $u(\cdot, \cdot)$  is a smooth function satisfying  $u(1, 1) = 1$ . The work of Isaki (1983), Singh, Horn, and Yu (1998) and Théberge (1999) deserves special consideration.

Motivated by the successful use of ratio and regression estimators for estimation of population means and totals, assuming simple random sampling with replacement, Isaki (1983) proposed a multivariate ratio estimator,  $\hat{S}_{ym}^2 = s_y^2 \sum_{i=1}^k W_i S_i^2 / s_i^2$ , and regression estimator,  $\hat{S}_{yr}^2 = s_y^2 + \sum_{i=1}^k B_i (S_i^2 - s_i^2)$ , for the finite population variance,  $S_y^2$ , where  $S_i^2$  and  $s_i^2$  are the known population variance and sample variance of auxiliary variable  $x_i$ ,  $i = 1, \dots, k$ , with  $0 < W_i < 1$  and  $\sum_{i=1}^k W_i = 1$ . With extra assumptions from a superpopulation model,  $W_i$  and  $B_i$  are then chosen by minimizing the model variances  $V_{\xi}(\hat{S}_{ym}^2)$  and  $V_{\xi}(\hat{S}_{yr}^2)$ .

In the case of a single auxiliary variable,  $x$ , Singh et al. (1998) proposed a high-level calibration approach for variance

---

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

estimation. To estimate the Yates–Grundy (YG)-type variance  $V_{YG} = 2^{-1} \sum_{i=1}^N \sum_{j=1}^N (\pi_i \pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)^2$ , their proposed estimator is  $\hat{V}_{SS} = 2^{-1} \sum_{i=1}^n \sum_{j=1}^n \Omega_{ij}(y_i/\pi_i - y_j/\pi_j)^2$ , where  $\Omega_{ij}$  are modified from  $D_{ij} = (\pi_i \pi_j - \pi_{ij})/\pi_{ij}$  by minimizing a distance measure between  $\Omega_{ij}$  and  $D_{ij}$  subject to the constraint

$$\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \Omega_{ij} \left( \frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2, \quad (1)$$

where  $\pi_i = P(i \in s)$  and  $\pi_{ij} = P(i, j \in s)$  are the first- and second-order inclusion probabilities and  $s$  denotes the set of sampled units. The right side of (1) is  $V_{YG}(\hat{X}_{HT})$  and is assumed known. Note that the  $D_{ij}$  may take negative values, which makes it difficult to interpret a distance measure like, say,  $\sum \sum (\Omega_{ij} - D_{ij})^2 / (D_{ij} Q_{ij})$ . It is also unclear whether this method can be extended to handle multiple auxiliary variables.

Treating calibration as a purely algebraic problem, Théberge (1999) extended the calibration estimators of Deville and Särndal (1992) to a bilinear parameter through the use of Kronecker products. The general form of the resulting estimator is quite complicated and usually involves  $n^2 \times n^2$  matrix operations, where  $n$  is the sample size, and thus is computationally awkward. This same estimator cannot be extended further to handle general second-order finite population quantities that are not in bilinear form. The method can result in negative estimates for known positive quantities.

The less fruitful and more sparse literature on this topic is due mainly to the fact that the relationship between a high-order function of the response variables and the auxiliary variables is almost certainly nonlinear. Most existing methods for using auxiliary information are developed based on an assumed linear working model, either explicitly or implicitly (see Wu and Sitter 2001 for a discussion). Such an assumption makes it virtually impossible to directly extend the methodology to the estimation of higher-order population functions.

In this article we develop efficient estimators for quadratic and other second-order finite population functions using the recently proposed MC and model-calibrated PEML methods (Wu and Sitter 2001). The procedure was originally proposed for estimation of finite population means, totals, and distribution functions. The most significant feature of this approach is that it effectively uses auxiliary information at the estimation stage under a general sampling design and a general working model, linear or nonlinear, with single or multiple auxiliary variables. The approach is model assisted in that the resulting estimators are asymptotically design unbiased irrespective of the correctness of the model, but high efficiency will be achieved if the working model is nearly correct. By viewing quadratic and other second-order finite population functions as totals or means of a derived synthetic finite population, we show that the MC and the model-calibrated PEML methods can be readily extended to obtain efficient estimators of high-order finite population functions. In particular, estimation of a finite population variance, a covariance between two response variables, the variance of a linear estimator, and other

related functions can be greatly improved when certain auxiliary information is available.

In Section 2 we briefly describe the MC and PEML methods, with discussions on the use of calibration equations for finite populations, and then extend them to the estimation of quadratic and other second-order finite population quantities and summarize some asymptotic properties. In Section 3 we derive several new estimators for a finite population variance or covariance and for the variance of a linear estimator using the general estimation strategy of Section 2 and a linear working model relating  $\mathbf{y}$  to  $\mathbf{x}$ . We show that  $\hat{S}_{yr}^2$  of Isaki (1983),  $\hat{V}_{SS}$  of Singh et al. (1998), and the extended calibration estimator of Théberge (1999) follow naturally as special cases of the proposed estimators. The PEML method also provides nonnegative estimates for certain known positive quantities. In Section 4 we demonstrate through a limited simulation study the improved performance of the proposed estimators over their conventional counterparts. Variance estimation for the generalized regression estimator (GREG) requires special treatment, and we discuss this in Section 5. We give some concluding remarks in Section 6.

## 2. EFFICIENT ESTIMATION OF QUADRATIC FINITE POPULATION FUNCTIONS

Let  $U = \{1, 2, \dots, N\}$  be the set of labels of the finite population. Associated with unit  $i$  are values of response variables,  $\mathbf{y}_i$ , and covariates,  $\mathbf{x}_i$ , both vector valued. In the case of scalar variables, we use  $y$  and  $x$ . In the most general case, we assume that the values  $\mathbf{x}_1, \dots, \mathbf{x}_N$  are known for the entire finite population (referred to as complete auxiliary information) but that  $\mathbf{y}_i$  is known only if the  $i$ th unit is selected in the sample,  $s$ . However, under a linear working model we only need that certain second-order summary statistics of  $\mathbf{x}$  be known. Throughout, we assume that all  $\pi_i$  and  $\pi_{ij}$  are strictly positive and let  $d_i = 1/\pi_i$  and  $d_{ij} = 1/\pi_{ij}$ .

A quadratic finite population function can be defined as  $Q = \sum_{i=1}^N \sum_{j=1}^N a_{ij} y_i y_j$ , where  $a_{ij}$  are known constants. More generally, let

$$T = \sum_{i=1}^N \sum_{j=i+1}^N \phi(\mathbf{y}_i, \mathbf{y}_j),$$

where  $\phi(\cdot, \cdot)$  is a symmetric function (a kernel of degree 2 for a  $U$  statistic). The quadratic form  $Q$  defined earlier and the bilinear parameter discussed by Théberge (1999) are both special cases of  $T$ . Practically useful examples include the finite population variance of  $y$ ,  $S_y^2 = (N-1)^{-1} \sum_{i=1}^N (y_i - \bar{Y})^2$ ; the finite population covariance between two response variables  $y$  and  $z$ ,  $C_{yz} = (N-1)^{-1} \sum_{i=1}^N (y_i - \bar{Y})(z_i - \bar{Z})$  (see Sec. 3 for alternative formulations of  $S_y^2$  and  $C_{yz}$ ); and the YG form of the variance,  $V_{YG} = \sum_{i=1}^N \sum_{j=i+1}^N (\pi_i \pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)^2$ , of the Horvitz–Thompson (HT) estimator of a population total,  $\hat{Y}_{HT} = \sum_{i \in s} d_i y_i$ . Other quantities of interest, such as the correlation coefficient between  $y$  and  $z$  and the regression coefficient of  $y$  on  $z$ , are smooth functions of several quadratic or other second-order functions.

If we arrange all of the pairs  $(ij)$  with  $i < j$  in a sequence, and denote this by  $\alpha = 1, 2, \dots, N^*$ , where  $N^* = N(N-1)/2$ ,

then we may express  $T$  as

$$T = \sum_{\alpha=1}^{N^*} t_{\alpha},$$

where  $t_{\alpha} = \phi(\mathbf{y}_i, \mathbf{y}_j)$  for  $\alpha = (ij)$ . Thus  $T$  can be viewed as a population total defined on a synthetic finite population  $U^* = \{1, 2, \dots, N^*\}$  with characteristic of interest  $t_{\alpha}$ . The corresponding sample of pairs would be  $s^* = \{\alpha = (ij) : i < j \text{ and } i, j \in s\}$ . Let  $n^* = n(n-1)/2$  be the number of pairs in  $s^*$ . The ‘‘first-order’’ inclusion probabilities over this synthetic population are  $\pi_{\alpha}^* = \pi_{ij}$  for unit  $\alpha = (ij)$ . Let  $d_{\alpha}^* = 1/\pi_{\alpha}^* = 1/\pi_{ij}$ . Note that in many applications, the diagonal terms  $\phi(\mathbf{y}_i, \mathbf{y}_i) = 0$ . If not then these terms can also be included in  $T$ , which amounts to changing  $N^*$  to  $N(N+1)/2$  and  $n^*$  to  $n(n+1)/2$  and using  $\pi_{ii} = \pi_i$ .

## 2.1 The Model-Calibration and Generalized Difference Estimators

If one were to now directly apply the calibration method of Deville and Särndal (1992) to estimate  $T$  viewed as a population total over the derived population, then one would use  $\hat{T}_C = \sum_{\alpha \in s^*} w_{\alpha} t_{\alpha}$ , where the  $w_{\alpha}$ 's minimize a distance measure between the  $w_{\alpha}$ 's and the  $d_{\alpha}^*$ 's subject to constraint  $\sum_{\alpha \in s^*} w_{\alpha} c_{\alpha} = \sum_{\alpha=1}^{N^*} c_{\alpha}$ , for some scalar auxiliary variable, say  $c_{\alpha}$ . If so, then  $\hat{T}_C$  would be the same or asymptotically equivalent to the GREG estimator treating  $c$  as the single auxiliary variable (Deville and Särndal 1992). Thus, although the estimator  $\hat{T}_C$  will be approximately design-unbiased for any fixed constants  $c_{\alpha}$ , because it is a GREG-type estimator (see Wu and Sitter 2001), the estimator will be most efficient if  $c_{\alpha}$  is chosen such that the relationship between  $t_{\alpha}$  and  $c_{\alpha}$  is nearly linear. If the linear correlation coefficient between  $t_{\alpha}$  and  $c_{\alpha}$  is 1 or  $-1$ , then  $\hat{T}_C = T$ .

One could alternately use a difference estimator (Cassel et al. 1976), which is given by  $\hat{T}_D = \sum_{\alpha \in s^*} d_{\alpha}^* t_{\alpha} + \sum_{\alpha=1}^{N^*} c_{\alpha} - \sum_{\alpha \in s^*} d_{\alpha}^* c_{\alpha}$ . In general,  $\hat{T}_D$  should be used with caution, because it performs well only if  $c_{\alpha}$  is a good proxy for  $t_{\alpha}$ . When  $e_{\alpha} = t_{\alpha} - c_{\alpha}$  is more variable than  $t_{\alpha}$  itself, we will have  $\text{var}(\hat{T}_D) > \text{var}(\hat{T}_{HT})$ , where  $\hat{T}_{HT} = \sum_{\alpha \in s^*} d_{\alpha}^* t_{\alpha}$ . Asymptotically, we also have  $\text{var}(\hat{T}_C) \leq \text{var}(\hat{T}_D)$ , so  $\hat{T}_C$  is preferred in most situations.

Calibrating on an arbitrary known sequence of constants,  $c_{\alpha}$ , will result in a calibration estimator  $\hat{T}_C$  that is asymptotically at least as good as the HT estimator,  $\hat{T}_{HT}$ . The question then becomes: What is the ‘‘best’’ choice for these constants? To determine how auxiliary information can be best used at the estimation stage, two related questions need to be answered: (1) How much auxiliary information is available? and (2) How are the auxiliary variables related to the variables of interest? Auxiliary information is usually available at one of two different levels: some summary statistics at the population level or complete auxiliary information at the unit level. The second question cannot be answered without resorting to a superpopulation model. Blindly calibrating on individual auxiliary variables generally is not a good approach.

Under a general sampling design and a general working model, auxiliary information can be effectively used through the fitted values. This idea has been developed in detail by

Wu and Sitter (2001) using the MC and model-calibrated PEML methods. We describe the methods as we extend them to the present context.

Suppose that the relationship between  $y_i$  and  $\mathbf{x}_i$  can be depicted by a semiparametric model through the first- and second-order moments,

$$E_{\xi}(y_i | \mathbf{x}_i) = \mu(\mathbf{x}_i, \boldsymbol{\theta}), \quad V_{\xi}(y_i | \mathbf{x}_i) = v_i^2 \sigma^2, \\ i = 1, 2, \dots, N, \quad (2)$$

where  $\boldsymbol{\theta} = (\theta_0, \dots, \theta_p)'$  and  $\sigma^2$  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  $E_{\xi}$  and  $V_{\xi}$  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. For ease of presentation, we restrict to scalar  $y_i$  at this point.

Let  $\hat{\boldsymbol{\theta}}_N$  be an estimate of  $\boldsymbol{\theta}$  based on the entire finite population, and let  $y_i^* = \mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}}_N)$ . If the working model (2) is appropriate, then  $y_i^*$  should be an ‘‘ideal’’ choice for approximating  $y_i$ ; that is,  $y_i^*$  should have higher linear correlation with  $y_i$  than other ad hoc choices if the model is adequate and the auxiliary information is really informative.

A design-based estimator  $\hat{\boldsymbol{\theta}}$  for  $\hat{\boldsymbol{\theta}}_N$  can be obtained from the sample data (Wu and Sitter 2001). Let  $\hat{y}_i = \mu(\mathbf{x}_i, \hat{\boldsymbol{\theta}})$ ,  $i = 1, 2, \dots, N$ , be the fitted values from model (2). The MC estimator for the finite population total  $Y = \sum_{i=1}^N y_i$  is defined as  $\hat{Y}_{MC} = \sum_{i \in s} w_i y_i$ , where the weights  $w_i$  minimize an average distance between  $w_i$  and the basic design weights  $d_i$  and  $\Phi_s$ , subject to constraints

$$\sum_{i \in s} w_i = N, \quad \text{and} \quad \sum_{i \in s} w_i \hat{y}_i = \sum_{i=1}^N \hat{y}_i. \quad (3)$$

Under a linear regression working model with an intercept, the first constraint in (3) is redundant. The resulting  $\hat{Y}_{MC}$  under a chi-squared distance measure coincides with the usual GREG. The estimated regression coefficient  $\hat{B}$  between  $y_i$  and  $\hat{y}_i$  is always equal to 1 (thm. 1 of Wu and Sitter 2001). If the working model is nonlinear or the unknown population quantity to be estimated is a nonlinear function (including quadratic functions), then the  $\hat{B}$  will not necessarily be 1. Under such situations, adding the first constraint usually would improve the estimate, which is equivalent to finding a GREG-type estimator by regressing  $y_i$  over  $\hat{y}_i$  with an intercept. (See the simulation results reported in Wu and Sitter 2001.)

The MC method can be easily extended to the present context for estimating  $T$  by using fitted values  $\hat{t}_{\alpha} = \phi(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j)$  for  $t_{\alpha}$  and treating  $d_{\alpha}^* = 1/\pi_{ij}$  as the basic design weights, where  $\alpha = (ij)$ . Recall that  $d_{ij} = 1/\pi_{ij}$  and use the original pair index  $(ij)$ ; then the MC estimator of the quadratic function  $T$  is defined as  $\hat{T}_{MC} = \sum \sum_{(ij) \in s^*} w_{ij} \phi(\mathbf{y}_i, \mathbf{y}_j)$ , where the  $w_{ij}$ 's minimize an average distance measure between  $w_{ij}$  and  $d_{ij}$  subject to

$$\frac{1}{N^*} \sum_{i \in s} \sum_{j > i} w_{ij} = 1 \quad \text{and} \\ \sum_{i \in s} \sum_{j > i} w_{ij} \phi(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j) = \sum_{i=1}^N \sum_{j=i+1}^N \phi(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j). \quad (4)$$

Under the simple chi-squared distance measure

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

where the  $q_{ij}$ 's are known positive weights unrelated to  $d_{ij}$ , the resulting estimator is given by

$$\hat{T}_{MC} = \sum_{i \in s} \sum_{j > i} d_{ij} \phi(\mathbf{y}_i, \mathbf{y}_j) + \left\{ \sum_{i=1}^N \sum_{j=i+1}^N \phi(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j) - \sum_{i \in s} \sum_{j > i} d_{ij} \phi(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j) \right\} \hat{B}, \quad (5)$$

where  $\hat{B} = C(u, v) / C(u, u)$ ,  $C(u, v) = \sum_{(ij) \in s^*} d_{ij} q_{ij} (u_{ij} - \bar{u})(v_{ij} - \bar{v})$ ,  $u_{ij} = \phi(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j)$ ,  $v_{ij} = \phi(\mathbf{y}_i, \mathbf{y}_j)$ ,  $\bar{u} = \sum_{(ij) \in s^*} d_{ij} q_{ij} u_{ij} / \sum_{(ij) \in s^*} d_{ij} q_{ij}$ , and  $\bar{v}$  and  $C(u, u)$  are defined similarly. Note that the first term on the right side of (5) is the usual HT estimator,  $\hat{T}_{HT}$ . When there is no auxiliary information, the constraints used in (4) will have to be removed. The unconstrained minimization of  $\Phi_{s^*}$  results in  $w_{ij} = d_{ij}$ , and thus in this case  $\hat{T}_{MC}$  reduces to  $\hat{T}_{HT}$ . For simplicity of presentation, we consider a single  $y$  variable with model (2) in the following theorem, but the results hold for the general case of vector response variables,  $\mathbf{y}$ . Under proper asymptotic settings, assuming  $T = O_p(N^*)$ , we have the following.

*Theorem 1.* a. Under the regularity conditions (a)–(c) specified in the Appendix, the MC estimator  $\hat{T}_{MC} = \hat{T}_{HT} + O_p(N^*/\sqrt{n^*})$  and is therefore an asymptotically design unbiased estimator of  $T$ .

b. With an extra condition (d) also given in the Appendix, the asymptotic design-based variance of  $\hat{T}_{MC}$  is given by

$$\text{var}(\hat{T}_{MC}) \doteq \frac{1}{2} \sum_{i=1}^N \sum_{j=i+1}^N \sum_{l=1}^N \sum_{m=l+1}^N (\pi_{ij} \pi_{lm} - \pi_{ijlm}) \left( \frac{E_{ij}}{\pi_{ij}} - \frac{E_{lm}}{\pi_{lm}} \right)^2,$$

where  $\pi_{ijlm}$  are the fourth-order inclusion probabilities,  $E_{ij} = v_{ij} - u_{ij} B_N$ ,  $B_N = C_N(u, v) / C_N(u, u)$ ,  $C_N(u, v) = \sum_{i=1}^N \sum_{j=i+1}^N d_{ij} q_{ij} (u_{ij} - \bar{u}_N)(v_{ij} - \bar{v}_N)$ ,  $u_{ij} = \phi(y_i^*, y_j^*)$ ,  $v_{ij} = \phi(y_i, y_j)$ ,  $y_i^* = \mu(\mathbf{x}_i, \boldsymbol{\theta}_N)$ ,  $\bar{u}_N = \sum_{i=1}^N \sum_{j=i+1}^N u_{ij} / N^*$ , and  $C_N(u, u)$  and  $\bar{v}_N$  are defined similarly.

c.  $\text{var}(\hat{T}_{MC})$  can be consistently estimated by

$$\text{var}(\hat{T}_{MC}) \doteq \frac{1}{2} \sum_{i=1}^n \sum_{j=i+1}^n \sum_{l=1}^n \sum_{m=l+1}^n \frac{\pi_{ij} \pi_{lm} - \pi_{ijlm}}{\pi_{ijlm}} \left( \frac{e_{ij}}{\pi_{ij}} - \frac{e_{lm}}{\pi_{lm}} \right)^2,$$

where  $e_{ij} = \phi(y_i, y_j) - \phi(\hat{y}_i, \hat{y}_j) \hat{B}$ .

*Proof.* See the Appendix.

The asymptotic variance of  $\hat{T}_{MC}$  involves inclusion probabilities up to the fourth order. A comprehensive algebraic comparison between  $\text{var}(\hat{T}_{MC})$  and

$$\text{var}(\hat{T}_{HT}) \doteq \frac{1}{2} \sum_{i=1}^N \sum_{j=i+1}^N \sum_{l=1}^N \sum_{m=l+1}^N (\pi_{ij} \pi_{lm} - \pi_{ijlm}) \left( \frac{v_{ij}}{\pi_{ij}} - \frac{v_{lm}}{\pi_{lm}} \right)^2$$

is not possible; however, the relationship between  $\text{var}(\hat{T}_{MC})$  and  $\text{var}(\hat{T}_{HT})$  is similar to that of  $\text{var}(\hat{Y}_{GR})$  and  $\text{var}(\hat{Y}_{HT})$ ,

where  $\hat{Y}_{GR}$  is the GREG estimator for the population total  $Y$ . The  $E_{ij}$ 's are the fitted residuals of the “response variable”  $t_\alpha = v_{ij}$  over “covariate”  $u_{ij}$ . The fact that the  $E_{ij}$ 's are less variable than the  $v_{ij}$ 's implies that  $\text{var}(\hat{T}_{MC})$  will be smaller than  $\text{var}(\hat{T}_{HT})$  for most commonly used sampling designs. The variance reduction depends on the correlation coefficient  $\rho_{v,u}$  between  $v_{ij}$  and  $u_{ij}$ . The two extreme cases are (a)  $|\rho_{v,u}| = 1$ , where  $E_{ij} = 0$  and  $\text{var}(\hat{T}_{MC}) = 0$ , and (b)  $\rho_{v,u} = 0$ , where  $E_{ij} = v_{ij} - \bar{v}$  and  $\text{var}(\hat{T}_{MC}) = \text{var}(\hat{T}_{HT})$ . This relationship can be seen more clearly under simple random sampling, where it is easy to show that

$$\text{var}(\hat{T}_{MC}) = \text{var}(\hat{T}_{HT})(1 - \rho_{v,u}^2) + O(N^*/n^*).$$

Note that  $\text{var}(\hat{T}_{HT}) = O(N^*/n^*)$ , the MC estimator of  $T$ , will perform at least as good as the HT estimator, with the variance reduction depending on  $\rho_{v,u}$ .

If  $\mathbf{x}$  provides relevant information in explaining  $y$  through a model like (2)—that is, the correlation between  $y_i$  and  $y_i^* = \mu(\mathbf{x}_i, \boldsymbol{\theta}_N)$  is high, and, consequently, a high correlation between  $v_{ij}$  and  $u_{ij}$ —then the gain from using  $\hat{T}_{MC}$  over  $\hat{T}_{HT}$  can be substantial. Moreover, construction of  $\hat{T}_{MC}$  requires no extra step in the modeling stage. The same fitted values  $\hat{y}_i$  are used for the estimation of any quadratic and other second-order population functions. Indeed, the structure of (5) is identical to the generalized regression estimator applied to a single response variable,  $t_\alpha = v_{ij}$ , and a single auxiliary variable,  $t_\alpha^* = \phi(y_i^*, y_j^*)$ . Computation for  $\hat{T}_{MC}$  requires no extra matrix manipulation after the initial modeling and thus is extremely simple.

It is interesting to notice that if we let  $\hat{B} = 1$  in (5), the resulting estimator can be viewed as a generalized difference (GD) estimator as discussed earlier. We denote this estimator by  $\hat{T}_{GD}$ . A theorem similar to Theorem 1 can also be restated in terms of  $\hat{T}_{GD}$ .

$\hat{T}_{GD}$  is computationally simpler than  $\hat{T}_{MC}$  and can perform very well under correctly specified working model where the correlation between  $t_\alpha$  and  $t_\alpha^*$  is high. It may also perform poorly under misspecified superpopulation model, where the fitted values are off target. The MC estimator, on the other hand, is more robust. The foregoing discussions regarding  $\hat{T}_C$  and  $\hat{T}_D$  all apply here to make parallel statements about  $\hat{T}_{MC}$  and  $\hat{T}_{GD}$ .

## 2.2 The Pseudoempirical Maximum Likelihood Estimator

The model-calibrated PEMPL estimator for the finite population mean  $\bar{Y}$  is defined as  $\hat{Y}_{EL} = \sum_{i \in s} \hat{p}_i y_i$ , where the  $\hat{p}_i$ 's maximize the PEMPL

$$\hat{l}(\mathbf{p}) = \sum_{i \in s} d_i \log p_i, \quad (6)$$

subject to

$$\sum_{i \in s} p_i = 1, \quad \sum_{i \in s} p_i u_i = 0 \quad (0 < p_i < 1), \quad (7)$$

where  $u_i = \hat{y}_i - N^{-1} \sum_{i=1}^N \hat{y}_i$ . Despite the underlying nonparametric likelihood motivations (Chen and Sitter 1999),  $\hat{Y}_{EL}$

asymptotically equivalent to the MC estimator  $\hat{Y}_{MC} = N^{-1}\hat{Y}_{MC}$  (Wu and Sitter 2001). But the most practically important feature of  $\hat{Y}_{EL}$  is the intrinsically positive weights  $\hat{p}_i > 0$ . This is particularly appealing when we extend the technique to estimating certain nonnegative quadratic finite population functions. One may also alternatively view the PEML in (6) as a distance measure; then the relationship between  $\hat{Y}_{MC}$  and  $\hat{Y}_{EL}$  follows naturally.

Paralleling the development of  $\hat{T}_{MC}$ , we define the model-calibrated PEML estimator of  $T$  as  $\hat{T}_{EL} = N^* \sum \sum_{(ij) \in s^*} \hat{p}_{ij} \phi(\mathbf{y}_i, \mathbf{y}_j)$  where the  $\hat{p}_{ij}$ 's maximize

$$\hat{l}(\mathbf{p}) = \sum_{i \in s} \sum_{j > i} d_{ij} \log p_{ij},$$

subject to constraints

$$\sum_{i \in s} \sum_{j > i} p_{ij} = 1, \quad \sum_{i \in s} \sum_{j > i} p_{ij} \phi(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j) = \frac{1}{N^*} \sum_{i=1}^N \sum_{j=i+1}^N \phi(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j) \quad (p_{ij} > 0). \quad (8)$$

*Theorem 2.* Under the regularity conditions specified in the Appendix, the PEML estimator of  $T$ ,  $\hat{T}_{EL}$ , is asymptotically equivalent to the MC estimator  $\hat{T}_{MC}$  under the uniform weight  $q_{ij} = 1$ ; that is,  $\hat{T}_{EL} = \hat{T}_{MC} + o_p(N^*/\sqrt{n^*})$ . Thus  $\hat{T}_{EL}$  is also an asymptotically design-unbiased estimator of  $T$  with the same asymptotic variance as  $\hat{T}_{MC}$ .

*Proof.* See the Appendix.

One advantage of  $\hat{T}_{EL}$  besides its likelihood-based motivation is that the weights,  $\hat{p}_{ij}$ , are always positive, which may not be true for the weights of the MC estimator,  $w_{ij}$ . This property might be very attractive when all of the terms in  $T$  are themselves positive and therefore positive weights will ensure positive estimation.

The Lagrange multiplier method can be used to show that

$$\hat{p}_{ij} = \frac{d_{ij}^*}{1 + \lambda b_{ij}} \quad \text{for } (ij) \in s^*,$$

where  $d_{ij}^* = d_{ij} / \sum_{i \in s} \sum_{j > i} d_{ij}$ ,  $b_{ij} = \phi(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j) - (N^*)^{-1} \times \sum_{i=1}^N \sum_{j=i+1}^N \phi(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j)$ , and the scalar Lagrange multiplier,  $\lambda$ , is the solution to

$$g(\lambda) = \sum_{i \in s} \sum_{j > i} \frac{d_{ij} b_{ij}}{1 + \lambda b_{ij}} = 0. \quad (9)$$

The simple, stable algorithm described by Chen, Sitter, and Wu (2002) can be modified to solve (9). The modified algorithm comprises only two steps:

1. Compute  $b_L = \min\{b_{ij} : (ij) \in s^*\}$ , and  $b_U = \max\{b_{ij} : (ij) \in s^*\}$ . If  $b_L > 0$  or  $b_U < 0$ , then the PEML estimator does not exist; otherwise, proceed to the next step.
2. Find the solution to  $g(\lambda) = 0$  in the interval  $(-1/b_U, -1/b_L)$  using the bisection method:
  - a. Let  $L = -1/b_U$ ,  $R = -1/b_L$ , and  $\epsilon = 10^8$ .
  - b. Let  $M = (L + R)/2$ . If  $|g(M)| \leq \epsilon$ , then stop and report  $\lambda = M$ ; otherwise, let  $L = M$  if  $g(M) > 0$ , or let  $R = M$  if  $g(M) < 0$ .
  - c. Repeat step b until  $|g(M)| \leq \epsilon$  and report  $\lambda = M$ .

Chen et al. (2002) extend this algorithm to allow one to restrict the range of the resulting weights. A similar extension could be done in this context.

### 3. ESTIMATING VARIANCES AND COVARIANCES UNDER A LINEAR WORKING MODEL

In this section we consider the estimation of  $S_y^2 = (N-1)^{-1} \sum_{i=1}^N (y_i - \bar{Y})^2$ ,  $C_{yz} = (N-1)^{-1} \sum_{i=1}^N (y_i - \bar{Y})(z_i - \bar{Z})$ , and  $V_{YG} = \sum_{i=1}^N \sum_{j=i+1}^N (\pi_i \pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)^2$  using the proposed estimation strategy. We consider linear regression working models for both  $y$  and  $z$ , that is,  $E_\xi(y_i) = \mathbf{x}'_i \boldsymbol{\beta}$  and  $E_\xi(z_i) = \mathbf{x}'_i \boldsymbol{\gamma}$ . The fitted values for  $y_i$  and  $z_i$  are  $\hat{y}_i = \mathbf{x}'_i \hat{\boldsymbol{\beta}}$  and  $\hat{z}_i = \mathbf{x}'_i \hat{\boldsymbol{\gamma}}$ , where  $\hat{\boldsymbol{\beta}}$  is the design-based estimator for the regression coefficients  $\boldsymbol{\beta}$ ,

$$\hat{\boldsymbol{\beta}} = \left\{ \sum_{i \in s} d_i \mathbf{x}_i \mathbf{x}'_i \right\}^{-1} \sum_{i \in s} d_i \mathbf{x}_i y_i,$$

and  $\hat{\boldsymbol{\gamma}}$  is defined similarly with  $y_i$  replaced with  $z_i$ . Note that  $S_y^2$  and  $C_{yz}$  can be reexpressed as

$$S_y^2 = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N (y_i - y_j)^2$$

and

$$C_{yz} = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N (y_i - y_j)(z_i - z_j).$$

It is then straightforward to show that the MC estimators for  $S_y^2$ ,  $C_{yz}$ , and  $V_{YG}$  are given by

$$\begin{aligned} \hat{S}_{MC}^2 &= \hat{S}_{HT}^2 + \hat{\boldsymbol{\beta}}' (S_x^2 - s_x^2) \hat{\boldsymbol{\beta}} \hat{B}_1, \\ \hat{C}_{MC} &= \hat{C}_{HT} + \hat{\boldsymbol{\beta}}' (S_x^2 - s_x^2) \hat{\boldsymbol{\gamma}} \hat{B}_2, \end{aligned}$$

and

$$\hat{V}_{MC} = \hat{V}_{HT} + \hat{\boldsymbol{\beta}}' (U_{YG}^2 - \hat{U}_{HT}^2) \hat{\boldsymbol{\beta}} \hat{B}_3,$$

where

$$\hat{S}_{HT}^2 = \frac{1}{N(N-1)} \sum_{i \in s} \sum_{j > i} d_{ij} (y_i - y_j)^2,$$

$$\hat{C}_{HT} = \frac{1}{N(N-1)} \sum_{i \in s} \sum_{j > i} d_{ij} (y_i - y_j)(z_i - z_j),$$

$$\hat{V}_{HT} = \sum_{i \in s} \sum_{j > i} d_{ij} (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2,$$

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{X}})(\mathbf{x}_i - \bar{\mathbf{X}})',$$

$$s_x^2 = \frac{1}{N(N-1)} \sum_{i \in s} \sum_{j > i} d_{ij} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)',$$

$$U_{YG}^2 = \sum_{i=1}^N \sum_{j=i+1}^N (\pi_i \pi_j - \pi_{ij}) \left( \frac{\mathbf{x}_i}{\pi_i} - \frac{\mathbf{x}_j}{\pi_j} \right) \left( \frac{\mathbf{x}_i}{\pi_i} - \frac{\mathbf{x}_j}{\pi_j} \right)',$$

$$\hat{U}_{HT}^2 = \sum_{i \in s} \sum_{j > i} d_{ij} (\pi_i \pi_j - \pi_{ij}) \left( \frac{\mathbf{x}_i}{\pi_i} - \frac{\mathbf{x}_j}{\pi_j} \right) \left( \frac{\mathbf{x}_i}{\pi_i} - \frac{\mathbf{x}_j}{\pi_j} \right)',$$

and  $\hat{B}_1$ ,  $\hat{B}_2$ , and  $\hat{B}_3$  are defined similarly as  $\hat{B}$  in (5) with  $u_{ij} = \hat{\beta}'(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)'\hat{\beta}$  and  $v_{ij} = (y_i - y_j)^2$  for  $\hat{B}_1$ ,  $u_{ij} = \hat{\beta}'(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)'\hat{\gamma}$  and  $v_{ij} = (y_i - y_j)(z_i - z_j)$  for  $\hat{B}_2$ , and  $u_{ij} = \hat{\beta}'[(\pi_i\pi_j - \pi_{ij})(\mathbf{x}_i/\pi_i - \mathbf{x}_j/\pi_j)(\mathbf{x}_i/\pi_i - \mathbf{x}_j/\pi_j)]'\hat{\beta}$  and  $v_{ij} = (\pi_i\pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)^2$  for  $\hat{B}_3$ .

The generalized difference estimators  $\hat{S}_{GD}^2$ ,  $\hat{C}_{GD}$ , and  $\hat{V}_{GD}$  are obtained by letting  $\hat{B}_1 = \hat{B}_2 = \hat{B}_3 = 1$ , as before.

The PEML estimators for  $S_y^2$ ,  $C_{yz}$ , and  $V_{YG}$  are given by

$$\hat{S}_{EL}^2 = \frac{1}{2} \sum_{i \in s} \sum_{j > i} \hat{p}_{ij} (y_i - y_j)^2,$$

$$\hat{C}_{EL} = \frac{1}{2} \sum_{i \in s} \sum_{j > i} \hat{p}_{ij} (y_i - y_j)(z_i - z_j),$$

and

$$\hat{V}_{EL} = N^* \sum_{i \in s} \sum_{j > i} \hat{p}_{ij} (\pi_i\pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2,$$

where  $N^* = N(N-1)/2$ ,  $\hat{p}_{ij} = d_{ij}^*/(1 + \lambda b_{ij})$ , and  $\lambda$  is the solution to (9). For  $\hat{S}_{EL}^2$ ,  $b_{ij} = \hat{\beta}'\{(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)' - 2S_x^2\}\hat{\beta}$ ; for  $\hat{C}_{EL}$ ,  $b_{ij} = \hat{\beta}'\{(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)' - 2S_x^2\}\hat{\gamma}$ ; and for  $\hat{V}_{EL}$ ,  $b_{ij} = \hat{\beta}'\{(\pi_i\pi_j - \pi_{ij})(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)' - (N^*)^{-1}U_{YG}^2\}\hat{\beta}$ .

Note that in the case of a linear working model, despite the motivation of using the predicted values for each unit and thus implicitly desiring complete auxiliary information, one needs only knowledge of certain population quantities of the  $\mathbf{x}$  variables, such as  $S_x^2$  or  $U_{YG}^2$ . We can also establish the relationship of these proposed general estimators with those in the literature that have been developed for some more restrictive cases:

**Case 1.** Vector  $\mathbf{x}$  variables with a general sampling design and a bilinear parameter. A practically useful result of Théberge's extended calibration estimator for a bilinear parameter (form. 22 of Théberge 1999, p. 639) is identical to the GD estimator proposed in this article under a linear working model.

**Case 2.** Vector  $\mathbf{x}$  variables under simple random sampling. Under simple random sampling,  $s_x^2$  reduces to the usual sample variance-covariance matrix. Both  $\hat{S}_{MC}^2$  and  $\hat{S}_{GD}^2$  can be viewed as more general forms of the estimator  $\hat{S}_{yr}^2$  proposed by Isaki (1983) without his model assumption (2). Isaki's estimator  $\hat{V}_G$  is also in the same spirit as  $\hat{V}_{MC}$  or  $\hat{V}_{GD}$ .

**Case 3.** A scalar  $x$  variable with a general sampling design. If we drop the constraint  $\sum_{(ij) \in s^*} w_{ij} = N^*$ , assume  $\pi_i\pi_j - \pi_{ij} > 0$ , and let  $q_{ij} = Q_{ij}/(\pi_i\pi_j - \pi_{ij})$ , then  $\hat{V}_{MC}$  will be identical to  $\hat{V}_{SS}$  of Singh et al. (1998). Including the constraint  $\sum_{(ij) \in s^*} w_{ij} = N^*$  will usually improve the estimate. In this setting,

$$\hat{S}_{MC}^2 = \hat{S}_{HT}^2 + (S_x^2 - s_x^2)\hat{C}_1,$$

$$\hat{C}_{MC} = \hat{C}_{HT} + (S_x^2 - s_x^2)\hat{C}_2,$$

$$\hat{S}_{GD}^2 = \hat{S}_{HT}^2 + \hat{\beta}^2(S_x^2 - s_x^2),$$

and

$$\hat{C}_{GD} = \hat{C}_{HT} + \hat{\beta}\hat{\gamma}(S_x^2 - s_x^2),$$

where  $\hat{\beta}$  and  $\hat{\gamma}$  are the estimated slopes of the two regression lines with  $y$  and  $z$  as response variables,  $\hat{C}_1$  and  $\hat{C}_2$  are defined

as  $\hat{B}$  in (5) with  $u_{ij} = (x_i - x_j)^2$  and  $v_{ij} = (y_i - y_j)^2$  for  $\hat{C}_1$  and  $u_{ij} = (x_i - x_j)^2$  and  $v_{ij} = (y_i - y_j)(z_i - z_j)$  for  $\hat{C}_2$ . Both  $\hat{S}_{MC}^2$  and  $\hat{C}_{MC}$  are independent of  $\beta$  and  $\gamma$ . The difference estimator  $\hat{S}_{GD}^2$  was also discussed by Shah and Patel (1996). The MC estimator  $\hat{S}_{MC}^2$  reduces to the ratio estimator  $\hat{S}_R^2 = \hat{S}_{HT}^2(S_x^2/s_x^2)$  if we drop the constraint  $\sum_{i \in s} \sum_{j > i} w_{ij} = N^*$  from (4) and choose  $q_{ij} = (x_i - x_j)^{-2}$  in the distance measure, assuming  $x_i - x_j \neq 0$  for all  $(i, j)$ .

For the PEML in the case of a single auxiliary variable  $x$ ,  $b_{ij} = (x_i - x_j)^2 - 2S_x^2$  for both  $\hat{S}_{EL}^2$  and  $\hat{C}_{EL}$ , so the  $\hat{p}_{ij}$ 's are the same for both estimators. They are also independent of the regression coefficients  $\beta$  and  $\gamma$ . Note that  $\hat{S}_{EL}^2$  always takes nonnegative values. The same can be said of  $\hat{V}_{EL}$  if the design satisfies  $\pi_i\pi_j - \pi_{ij} \geq 0$  for all  $(i, j)$ .

**Case 4.** A scalar  $x$  variable with simple random sampling. In this simple case,  $\hat{S}_{HT}^2$ ,  $\hat{C}_{HT}$ , and  $s_x^2$  all reduce to the usual sample variances and covariances,  $s_y^2$ ,  $s_{yz}$ , and  $s_x^2$ . Also,  $\hat{S}_{GD}^2$  becomes  $\hat{S}_{yre}^2 = s_y^2 + \hat{\beta}^2(S_x^2 - s_x^2)$ , as proposed by Isaki (1983).

It should be noted that in both cases 3 and 4,  $\hat{C}_1 \neq \hat{\beta}^2$  and  $\hat{C}_2 \neq \hat{\beta}\hat{\gamma}$ . With the uniform weights  $q_{ij} = 1$ ,  $\hat{C}_1$  and  $\hat{C}_2$  are actually the estimated slopes of the regression line of  $(y_i - y_j)^2$  on  $(x_i - x_j)^2$  and of  $(y_i - y_j)(z_i - z_j)$  on  $(x_i - x_j)^2$ .

#### 4. SOME EMPIRICAL RESULTS

We carried out a limited simulation study to investigate the finite sample performance of the proposed variance estimators using the 1996 Statistics Canada's Family Expenditure (FAMEX) Survey data for the province of Ontario, downloaded from Statistics Canada Databases. The dataset contains observations on  $N = 2396$  sampled households over a variety of variables. For the purpose of illustration, we choose  $x_1$ , number of people in the household, and  $x_2$ , total income after taxes, as auxiliary variables and  $y_1$ , annual expenditures on clothing, and  $y_2$ , total expenditure, as the study variables. Because extra information is not available to us, we treat this dataset itself as the finite population in the simulation study. The population is split into eight strata using the original design weights.

A scatterplot reveals that a linear working model might be appropriate, but the relationship between the  $\mathbf{x}$  variables and the  $\mathbf{y}$  variables is not particularly strong. The finite population correlation coefficients are  $\rho_{x_1, y_1} = .40$ ,  $\rho_{x_1, y_2} = .44$ ,  $\rho_{x_2, y_1} = .60$ , and  $\rho_{x_2, y_2} = .87$ .

A stratified simple random sample of size  $n = 64$ , with 8 units from each stratum, was drawn and the MC, GD, and PEML estimators of Section 3 for the population variances  $S_{y_1}^2$  and  $S_{y_2}^2$  and covariance  $S_{12} = \text{cov}(y_1, y_2)$  were computed. The usual HT estimators were also included for comparison. This process was repeated  $B = 1,000$  times. Note that the sampling fraction here is approximately 2.5%.

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

$$\text{RB} = 100 \times \frac{1}{B} \sum_{b=1}^B \frac{\hat{S}_b^2 - S^2}{S^2} \quad \text{and} \quad \text{RE} = \frac{\text{MSE}_{HT}}{\text{MSE}},$$

Table 1. Percentage RB and RE of Estimators for the Finite Population Variance and Covariance

Parameter	HT	MC	GD	PEML
Percentage RB				
$S_{y_1}^2$	.584	-3.259	-.993	-3.759
$S_{y_2}^2$	.560	4.743	3.739	5.664
$S_{12}$	.591	.625	1.346	.022
RE to HT				
$S_{y_1}^2$	1.000	1.629	1.454	1.648
$S_{y_2}^2$	1.000	1.318	1.525	1.289
$S_{12}$	1.000	1.913	1.852	1.956

where  $\hat{S}_b^2$  is the estimator of  $S^2$  computed from the  $b$ th simulated sample,  $\text{MSE} = B^{-1} \sum_{b=1}^B (\hat{S}_b^2 - S^2)^2$ , and  $\text{MSE}_{HT}$  is the MSE of  $\hat{S}_{HT}^2$ .

Table 1 reports the RB and RE for estimators included in the simulation. The absolute values of RBs are all within reasonable range, with the largest occurring for the PEML estimators at 5.664%. In terms of efficiency, all of MC, GD, and PEML estimators outperform the conventional HT estimators, with MSE reduced by almost half for estimating the covariance  $S_{12}$ . The MC and PEML estimators perform better than the GD estimator when estimating  $S_{y_1}^2$  and  $S_{12}$ , but this is reversed when estimating  $S_{y_2}^2$ . One possible reason for this is that the linear correlation between  $y_2$  and  $(x_1, x_2)$  is much stronger than that of  $y_1$  ( $\rho_{x_2, y_1} = .60, \rho_{x_2, y_2} = .87$ ). The GD estimator usually performs well in this case (see Wu and Sitter 2001 for a discussion).

When we rerun the foregoing simulation with a doubled sample size ( $n = 128$ ), the pattern for RE remains the same, but in this case all of the absolute values of RBs are less than 3%, with the largest at 2.7% for the PEML estimators.

## 5. VARIANCE ESTIMATION FOR THE GENERALIZED REGRESSION ESTIMATOR

One must take care when the auxiliary information has been used in estimating a population quantity and one wishes to again use it in estimating the variance of this estimator. For example, to estimate the population total  $Y$  when the population total  $\mathbf{X}$  is known, the MC and GD estimators under a linear working model both coincide with the widely used GREG,  $\hat{Y}_{GR} = \hat{Y}_{HT} + (\mathbf{X} - \hat{\mathbf{X}}_{HT})' \hat{\boldsymbol{\beta}}$  (Wu and Sitter 2001). Its asymptotic design-based variance is given by

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

where  $e_i = y_i - \mathbf{x}_i' \boldsymbol{\beta}_N$  and  $\boldsymbol{\beta}_N$  is the finite population regression coefficient.

Thus  $\text{var}(\hat{Y}_{GR})$  is a special case of  $T$ . To apply the estimation strategies of Section 2, using the original linear working model, the fitted values for  $e_i$  would be  $\tilde{e}_i = \hat{y}_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}} = 0$ . The MC, GD, and EL estimators for  $\text{var}(\hat{Y}_{GR})$  all reduce to the usual HT estimator if  $\tilde{e}_i$ 's are used in the construction of these estimators. The fitted values  $\hat{t}_\alpha = (\pi_i \pi_j - \pi_{ij})(\tilde{e}_i / \pi_i - \tilde{e}_j / \pi_j)^2$  form a constant sequence that has zero correlation with the "response variable,"  $t_\alpha = (\pi_i \pi_j - \pi_{ij})(e_i / \pi_i - e_j / \pi_j)^2$ .

Our proposed model calibration approach of Section 2 should now be modified so as to find a calibration variable  $u_{ij} = u(\mathbf{x}_i, \mathbf{x}_j)$  that is correlated with  $\phi(e_i, e_j) = (\pi_i \pi_j - \pi_{ij})(e_i / \pi_i - e_j / \pi_j)^2$ . We can then construct estimators as in Section 2 by calibrating over the  $u(\mathbf{x}_i, \mathbf{x}_j)$ 's. The best choice of  $u(\mathbf{x}_i, \mathbf{x}_j)$  is the one with the highest correlation with  $\phi(e_i, e_j)$ .

Whether the same auxiliary information can be used to improve both the estimation of the population total  $Y$  using  $\hat{Y}_{GR}$  and the estimation of its variance depends on the joint distribution of the residual variable,  $e$ , and the auxiliary variables,  $\mathbf{x}$ . Let  $E_N = (e_1, \dots, e_N)'$ ,  $X_N = (\mathbf{x}_1, \dots, \mathbf{x}_N)'$ , where  $\mathbf{x}_i$  contains 1 as its first component. Under an ordinary least squares fitting, we have  $X_N' E_N = \mathbf{0}$ , and the auxiliary variables  $\mathbf{x}$  are uncorrelated with the residual variable,  $e$ . If the joint distribution of  $e$  and  $\mathbf{x}$  is close to normal, then  $e$  and  $\mathbf{x}$  will be nearly independent. Any choice of  $u(\mathbf{x}_i, \mathbf{x}_j)$  will be nearly uncorrelated with  $\phi(e_i, e_j)$ . The same auxiliary information cannot be used to improve the variance estimation for the GREG.

Some situations exist in which variance estimation for the GREG can be improved. A simple example is when extra auxiliary information not used for the initial model building to estimate the total is available from other sources. Another is where a GREG estimator has been used to estimate the total because of internal consistency requirements to match marginal totals or because it is well known and accepted by practitioners. In this case, if a nonlinear working model or a linear working model with nonhomogeneous variance is more appropriate, then extra modeling work at the variance estimation stage and calibrating over fitted values from a more appropriate model will improve the estimation of  $\text{var}(\hat{Y}_{GR})$ . Note that matching marginal totals at the variance estimation stage is not as important.

We demonstrate the foregoing points by considering the case in which a linear working model with implicitly assumed homogeneous variance is used to estimate the population total (i.e., a GREG) when the homogeneous variance assumption is false. Under a linear working model  $y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i$  with variance of the error term  $\varepsilon_i$  depending on  $\mathbf{x}_i$ , there will be room for improvement in estimation of  $\text{var}(\hat{Y}_{GR})$ . To see this, consider simple random sampling with  $\phi(e_i, e_j) = (e_i - e_j)^2$  and  $u_{ij} = [u(\mathbf{x}_i) - u(\mathbf{x}_j)]^2$  for a certain choice of  $u_i = u(\mathbf{x}_i)$ . It is straightforward to show that  $\text{cov}[\phi(e_i, e_j), u_{ij}] = (N^*)^{-1} \sum_{i=1}^N \sum_{j=i+1}^N \phi(e_i, e_j) u_{ij} - \bar{\phi} \bar{u} \doteq 2\{N^{-1} \sum_{i=1}^N e_i^2 u_i^{*2} - \bar{e}^2 \bar{u}^{*2}\} = 2\text{cov}(e_i^2, u_i^{*2})$ , where  $u_i^* = u_i - \bar{u}$ , and  $\bar{\phi}$ ,  $\bar{e}^2$ , and so on are in obvious notation. If, for instance,  $V_\varepsilon(\varepsilon_i) = u_i^2 \sigma^2$ , then  $\text{cov}(e_i^2, u_i^{*2}) \neq 0$ . Calibrating over  $u_{ij} = [(u(\mathbf{x}_i) - u(\mathbf{x}_j))]^2$  will improve the variance estimation for the GREG.

To further demonstrate, we performed a limited simulation study. For the FAMEX data discussed in Section 4, scatterplots of fitted residuals against covariates and fitted values demonstrate homogeneity of the error terms for response variables  $y_1$  and  $y_2$ . Several calibration variables  $u_{ij}$  were used, and the resulting estimators for  $\text{var}(\hat{Y}_{GR})$  showed no improvement, as expected. We performed another simulation in which the finite population has the same stratified structure and size as the FAMEX data but is generated from a superpopulation model  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$  with  $V_\varepsilon(\varepsilon_i) = x_{2i}^2 \sigma^2$ . Using

$u_{ij} = (x_{2i} - x_{2j})^2$  as the calibration variable, the simulated RE of the MC, GD, and PEMPL estimators for  $\text{var}(\hat{Y}_{GR})$  were 1.33, 1.04, and 1.34. Similar results were also obtained under the choice of  $u_{ij} = (\hat{y}_i - \hat{y}_j)^2$ , where  $\hat{y}_i$  are the fitted values.

## 6. CONCLUDING REMARKS

Note that the finite population variance can be written as  $S_y^2 = (N-1)^{-1}[\sum_{i=1}^N y_i^2 - (\sum_{i=1}^N y_i)^2/N]$ . The two population totals,  $\sum_{i=1}^N y_i^2$  and  $\sum_{i=1}^N y_i$ , can then be estimated separately using the MC or PEMPL method. The resulting estimators involve no second-order inclusion probabilities; however, the nonnegativity of the estimators obtained from the two separate pieces cannot be guaranteed.

To close, we note that the idea of treating quadratic finite population functions as totals of certain generalized characteristics over a derived finite population was used previously by Hanurav (1966) and Liu and Thompson (1983). This, combined with the MC and PEMPL methods, provides a general way of using auxiliary information in estimating quadratic and other second-order finite population functions. Some attractive features of the proposed methodology are as follows:

1. Estimators can be constructed under a general sampling design.

2. The methods can handle scalar or vector-valued auxiliary variables, as well as linear or nonlinear working models and thus are very flexible.

3. The approach is model-assisted in that the resulting estimators are asymptotically design unbiased regardless of the correctness of the model but have high efficiency if the working model adequately describes the relationship between the response variables and the covariates.

4. The approaches require no additional step for the modeling; the same fitted values are used for any quadratic or other second-order population functions.

5. In the case of linear working models, estimation of population variances and covariances or the variance of a linear estimator requires that only the  $S_x^2$  or other second-order summary statistics of  $\mathbf{x}$  be known at the population level. This is much like the GREG for the population mean where only  $\bar{\mathbf{X}}$  need be known to construct the estimator.

6. After the initial modeling stage (e.g., estimation of  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}$ ), the construction of proposed estimators involves only scalar manipulations, and the resulting estimators have very simple forms.

7. The limited empirical results show that the proposed estimators for population variances and covariances are very efficient compared to the conventional estimators.

8. The model-calibrated PEMPL estimators ensure nonnegative estimation for certain positive quantities, such as population variances.

Estimation of quadratic and other second-order finite population functions requires the knowledge of second-order inclusion probabilities,  $\pi_{ij}$ , and knowledge of the auxiliary information at the second-order level, such as  $S_x^2$  or some other aggregated summary statistics. Knowing  $\bar{\mathbf{X}}$  usually provides little or no information for improving the estimation of quadratic functions. When the true relationship between the response variables and the covariates is seriously nonlinear, complete auxiliary information at unit level and a careful

model-building process are required for the effective use of the auxiliary information and efficient estimation of quadratic functions.

## APPENDIX: REGULARITY CONDITIONS AND PROOFS

The asymptotic framework assumes a sequence of sampling designs and a sequence of finite populations, indexed by  $\nu$ . Both the sample size  $n_\nu$  and the population size  $N_\nu$  approach infinity as  $\nu \rightarrow \infty$ . All limiting processes are understood to be as  $\nu \rightarrow \infty$ , but here the  $\nu$  is suppressed to simplify notation.

The required regularity conditions for Theorem 1 are as follows:

- $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_N + O_p(n^{-1/2})$  and  $\boldsymbol{\theta}_N \rightarrow \boldsymbol{\theta}$ .
- For each  $(\mathbf{x}_i, \mathbf{x}_j)$ ,  $\partial\phi[\mu(\mathbf{x}_i, \mathbf{t}), \mu(\mathbf{x}_j, \mathbf{t})]/\partial\mathbf{t}$  is continuous in  $\mathbf{t}$  and

$$|\partial\phi[\mu(\mathbf{x}_i, \mathbf{t}), \mu(\mathbf{x}_j, \mathbf{t})]/\partial\mathbf{t}| \leq h(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta})$$

for  $\mathbf{t}$  in a neighborhood of  $\boldsymbol{\theta}$ , where  $|\cdot|$  is the  $L_1$  norm and  $\sum_{i=1}^N \sum_{j=i+1}^N h(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta})/N^* = O(1)$ .

- The second-order inclusion probabilities  $\pi_{ij}$  ensure the asymptotic normality of the HT estimator  $\sum \sum_{(ij) \in s^*} d_{ij} c_{ij}/N^*$  for a certain sequence  $c_{ij}$ .
- For each  $(\mathbf{x}_i, \mathbf{x}_j)$ ,  $\partial^2\phi[\mu(\mathbf{x}_i, \mathbf{t}), \mu(\mathbf{x}_j, \mathbf{t})]/\partial\mathbf{t}\partial\mathbf{t}'$  is continuous in  $\mathbf{t}$  and

$$|\partial^2\phi[\mu(\mathbf{x}_i, \mathbf{t}), \mu(\mathbf{x}_j, \mathbf{t})]/\partial\mathbf{t}\partial\mathbf{t}'| \leq k(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta})$$

for  $\mathbf{t}$  in a neighborhood of  $\boldsymbol{\theta}$ , where  $|\cdot|$  denotes the maximum absolute value of the elements in the matrix and  $\sum_{i=1}^N \sum_{j=i+1}^N k(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta})/N^* = O(1)$ .

### Proof of Theorem 1

Under the regularity conditions (a)–(d), following a similar argument as in the proof of theorem 1 of Wu and Sitter (2001), it can be shown that

$$\begin{aligned} \hat{T}_{MC} &= \hat{T}_{HT} + \left\{ \sum_{i=1}^N \sum_{j=i+1}^N \phi(y_i^*, y_j^*) - \sum_{i=1}^n \sum_{j=i+1}^n d_{ij} \phi(y_i^*, y_j^*) \right\} B_N \\ &\quad + o_p\left(\frac{N^*}{\sqrt{n^*}}\right) \\ &= \sum_{i=1}^n \sum_{j=i+1}^n d_{ij} E_{ij} + \sum_{i=1}^N \sum_{j=i+1}^N \phi(y_i^*, y_j^*) + o_p\left(\frac{N^*}{\sqrt{n^*}}\right), \end{aligned}$$

where  $B_N$  is the finite population quantity estimated by  $\hat{B}$ . It is easy to show that  $\sum \sum_{i < j} \pi_{ij} = n(n-1)/2 = n^*$  and  $\sum \sum \pi_{ijlm} = \pi_{ij}(n^* - 1)$ , where the double summation on the second equality is over all  $l < m$  such that  $(lm) \neq (ij)$ . Hence the YG-type variance formula applies here. The results of Theorem 1 then follow by noting that  $B_N = O(1)$  and  $\sum_{i=1}^N \sum_{j=i+1}^N \phi(y_i^*, y_j^*) - \sum_{i=1}^n \sum_{j=i+1}^n d_{ij} \phi(y_i^*, y_j^*) = O_p(N^*/\sqrt{n^*})$ , by condition c.

The required regularity conditions for Theorem 2 are the foregoing a–d, along with e–g as follows:

- $u^* = \max_{(ij) \in s^*} |u_{ij}^*| = o_p(\sqrt{n^*})$ .
- $\sum \sum_{(ij) \in s^*} d_{ij} u_{ij}^* / \sum \sum_{(ij) \in s^*} d_{ij} u_{ij}^{*2} = O_p(1/\sqrt{n^*})$ .
- $h^* = \max_{(ij) \in s^*} |h(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta}_N)| = o_p(n^*)$ ,

where

$$\begin{aligned} u_{ij}^* &= \phi[\mu(\mathbf{x}_i, \boldsymbol{\theta}_N), \mu(\mathbf{x}_j, \boldsymbol{\theta}_N)] \\ &\quad - (N^*)^{-1} \sum_{i=1}^N \sum_{j=i+1}^N \phi[\mu(\mathbf{x}_i, \boldsymbol{\theta}_N), \mu(\mathbf{x}_j, \boldsymbol{\theta}_N)]. \end{aligned}$$



### Proof of Theorem 2

Following the lines of the proof of theorem 2 of Wu and Sitter (2001), if we switch all of the single summations to double summations, replacing  $d_i$  with  $d_{ij}$  and  $u_i$  with  $u_{ij}^*$  defined earlier, then we can show that

$$\hat{T}_{EL} = \frac{N^*}{\hat{N}^*} \hat{T}_{HT} + \left\{ \sum_{i=1}^N \sum_{j=i+1}^N \phi(y_i^*, y_j^*) - \frac{N^*}{\hat{N}^*} \sum_{i=1}^n \sum_{j=i+1}^n d_{ij} \phi(y_i^*, y_j^*) \right\} B_N + o_p\left(\frac{N^*}{\sqrt{n^*}}\right),$$

where  $\hat{N}^* = \sum \sum_{(ij) \in s^*} d_{ij}$ . For most commonly used designs, such as simple random sampling or stratified random sampling,  $\hat{N}^* = N^*$ , it follows that  $\hat{T}_{EL} = \hat{T}_{MC} + o_p(N^*/\sqrt{n^*})$ . Under unequal probability sampling where  $N^*/\hat{N}^* - 1 = O_p(1/\sqrt{n^*})$ , the asymptotic variance of  $\hat{T}_{EL}$  will have to be modified from a ratio estimator in the same fashion as in modifying  $\text{var}(\sum_{i \in s} d_i y_i / N)$  for  $\text{var}(\sum_{i \in s} d_i y_i / \sum_{i \in s} d_i)$ ; the details are omitted here.

[Received September 2000. Revised July 2001.]

### REFERENCES

- 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 Use 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.
- Chen, J., Sitter, R. R., and Wu, C. (2002), "Using Empirical Likelihood Methods to Obtain Range Restricted Weights in Regression Estimators for Surveys," *Biometrika*, 89, to appear.
- Datta, G. S., and Ghosh, M. (1993), "Bayesian Estimation of Finite Population Variances With Auxiliary Information," *Sankhyā*, Ser. B, 55, 156–170.
- Das, A. K., and Tripathi, T. P. (1978), "Use of Auxiliary Information in Estimating the Finite Population Variance," *Sankhyā*, Ser. C, 40, 139–148.
- Deng, L. Y., and Wu, C. F. J. (1987), "Estimation of Variance of the Regression Estimator," *J. Amer. Statist. Assoc.*, 82, 568–576.
- Deville, J. C., and Särndal, C. E. (1992), "Calibration Estimators in Survey Sampling," *J. Amer. Statist. Assoc.*, 87, 376–382.
- Hanurav, T. V. (1966), "Some Aspects of Unified Sampling Theory," *Sankhyā*, Ser. A, 28, 175–204.
- Isaki, C. T. (1983), "Variance Estimation Using Auxiliary Information," *J. Amer. Statist. Assoc.*, 78, 117–123.
- Liu, T. P., and Thompson, M. E. (1983), "Properties of Estimators of Quadratic Finite Population Functions: The Batch Approach," *The Annals of Statistics*, 11, 275–285.
- Särndal, C. E. (1980), "On  $\pi$ -Inverse Weighting Versus Best Linear Unbiased Weighting in Probability Sampling," *Biometrika*, 67, 639–650.
- Shah, D. N., and Patel, P. A. (1996), "Asymptotic Properties of a Generalized Regression-type Predictor of a Finite Population Variance in Probability Sampling," *Canadian Journal of Statistics*, 24, 373–384.
- Singh, S., Horn, S., and Yu, F. (1998), "Estimation of Variance of General Regression Estimator: High Level Calibration Approach," *Survey Methodology*, 24, 41–50.
- Théberge, A. (1999), "Extensions of Calibration Estimators in Survey Sampling," *J. Amer. Statist. Assoc.*, 94, 635–644.
- Wu, C., and Sitter, R. R. (2001), "A Model-Calibration Approach to Using Complete Auxiliary Information From Survey Data," *J. Amer. Statist. Assoc.*, 96, 185–193.
- Zhang, B. (1996), "Estimating a Population Variance With Known Mean," *International Statistical Review*, 64, 215–229.
- Zhong, C. X. B., and Rao, J. N. K. (2000), "Empirical Likelihood Inference Under Stratified Random Sampling Using Auxiliary Information," *Biometrika*, 87, 929–938.