Calibration estimation using empirical likelihood in survey sampling

Jae Kwang Kim

Abstract:

Calibration estimation, which can be roughly described as a method of adjusting the original design weights to incorporate the known population totals of the auxiliary variables, has become very popular in sample surveys. The calibration weights are chosen to minimize a given distance measure while satisfying a set of constraints related with the auxiliary variable information.

Under simple random sampling, Chen and Qin (1993) suggested that the calibration estimator maximizing the constrained empirical likelihood can make efficient use of the auxiliary variables. We extend the result to unequal probability sampling and propose an algorithm to implement the proposed procedure. Asymptotic properties of the proposed calibration estimator are discussed. Results from a limited simulation study are presented.

Key Words: Generalized regression estimator, Nonparametric maximum likelihood estimator, Optimal regression estimator, Weighting procedure.

1. Introduction

In the samples selected from a finite population, auxiliary variables with known population totals are often observed. The known population totals usually come from external sources such as administrative data or census. Calibration estimation, which can be roughly described as a method of adjusting the original design weights to incorporate the known population totals of the auxiliary variables, has become very popular in sample surveys. Generally speaking, the calibration procedure chooses the adjusted weights that minimize a distance between the original weights and the adjusted weights, while satisfying a set of constraints related with the auxiliary variable information. Fuller (2002) provides a comprehensive overview of the calibration procedure in sample surveys.

In a purely mathematical point of view, the calibration estimation problem is a standard optimization problem with constraints and, given the same constraints, the choice of the objective function determines the properties of the resulting estimator. The classical regression estimator described in Cochran (1977) uses a Euclidian distance function. Deville and Särndal (1992) gave conditions for the distance functions to produce the calibration estimators that are asymptotically equivalent to the regression estimator.

In addition to the above interpretation of minimizing a distance function, the calibration estimator can also be viewed as a maximum likelihood estimator in some cases. Anderson (1957) derived the regression estimator as a solution to the maximum likelihood estimation under the bivariate normal distribution assumption. Hartley and Rao (1968) used a multinomial distribution for distinct sample values and proposed a scale-load estimator that can be obtained through a constrained maximum likelihood estimation. The empirical likelihood, termed by Owen (1988), is essentially the likelihood of the multinomial distribution used in Hartley and Rao (1968), where the parameters are the point masses assigned to the distinct sample values. Under simple random sampling, Chen and Qin (1993) proposed a calibration estimator that maximizes the empirical likelihood with constraints. Chen and Sitter (1999) extended the method to unequal probability sampling designs but, as discussed in Section 2, the Chen-Sitter estimator lacks the maximum likelihood interpretation. One advantage of having a maximum likelihood interpretation is that the resulting estimator will be asymptotically optimal.

In this paper, we propose a calibration estimator that preserves the maximum likelihood interpretation under unequal probability sampling. The objective function we consider is different from that of Chen and Sitter (1999) and thus has different asymptotic properties.

The paper is organized as follows. In Section 2, the basic setup is introduced and the proposed method is described. In Section 3, asymptotic properties of the proposed estimator are discussed. Variance estimation is also discussed in Section 4. In section 5, results from a simulation study are presented.

Department of Applied Statistics, Yonsei University, Seoul, 120-749, Korea.

2. Empirical likelihood calibration method

We begin by introducing the notion of empirical likelihood in a simple setup. Let $y_1, y_2, \cdots, y_n$ be the outcomes of the independently and identically distributed random variables from a continuous distribution function $F_0 \in \mathcal{F}$. We consider a class $\mathcal{F}_1 \subset \mathcal{F}$ of distribution functions that have support on the convex hull of $\{y_1, y_2, \cdots, y_n\}$. Thus, the elements in $\mathcal{F}_1$ can be written as

$$F_w(x) = \sum_{i=1}^{n} w_i I(y_i \le x) \tag{1}$$

with $\sum_{i=1}^{n} w_i = 1$ and $w_i \ge 0$, where $I(y_i \le x)$ takes the value one if $y_i \le x$ and takes the value zero otherwise. The parameter w_i is the amount of point mass that unit y_i represents in the population. That is, $w_i = F_0(y_i) - F_0(y_i-)$. Note that $F_w(y)$ is a distribution function, not an estimator, indexed by the set of parameters $w_1, w_2, \cdots, w_n$. For any parameter of the form $\theta = \theta(F)$, the estimator $\hat{F}$ of F_0 can be used to estimate θ by $\hat{\theta} = \theta\left(\hat{F}\right)$. For a parameter θ linear in y in the population, the estimator $\hat{\theta}$ using the class of distributions (1) leads to a linear estimator that is linear in y in the sample. Linear estimation is very popular in sample surveys because it provides internal consistency between estimators for several items.

The empirical distribution function, defined for $w_i = n^{-1}$ in (1), given no ties, is the nonparametric maximum likelihood estimator (NPMLE) of F_0, since it maximizes the following likelihood function,

$$L(w) = \prod_{i=1}^{n} w_i \tag{2}$$

over all w_i's satisfying $\sum_{i=1}^{n} w_i = 1$ and $w_i \ge 0$. Note that if the w_i are known functions of a fixed number of unknown constants then (2) is the usual parametric likelihood function. For any parameter of the form $\theta = \theta(F)$, the NPMLE $\hat{F}$ of F_0 can be used to compute the NPMLE of θ by $\hat{\theta} = \theta\left(\hat{F}\right)$.

If we observe the auxiliary variable x_i in the sample and the population mean of x_i is known, denoted by μ_x, the additional information of μ_x can be used to construct a constrained NPMLE of F_0. Chen and Qin (1993) proposed computing the constrained NPMLE of F_0 by solving

$$\text{maximize} \sum_{i=1}^{n} \log(w_i) \tag{3}$$

subject to

$$\sum_{i=1}^{n} w_i x_i = \mu_x \quad \text{and} \quad \sum_{i=1}^{n} w_i = 1. \tag{4}$$

The constrained NPMLE for the population total of y is

$$\hat{Y}_{cm} = \sum_{i=1}^{n} w_i^* y_i,$$

where w_i^* is the solution to the optimization problem in (3) and (4). Also, the constrained NPMLE for the p-th quantile of the distribution of y is

$$\hat{Q}_{cm,y}(p) = \inf\left\{x; \hat{F}_{cm,y}(x) \ge p\right\}$$

where

$$\hat{F}_{cm,y}(x) = \sum_{i=1}^{n} w_i^* I(y_i \le x).$$

We now consider an extension of the constrained NPMLE to samples selected from a finite population with unequal selection probabilities. Assume that unit i is selected with known probability π_i. Recall that w_i in (1) is a parameter value, so the distribution function (1) indexed by w does not depend on the sampling design. However, under unequal probability sampling, the empirical likelihood will be different from (2) because the amount that the i-th unit represents has been changed by unequal probability sampling. We suggest the empirical likelihood

$$L(w) = \prod_{i=1}^{n} \left(\frac{\pi_i w_i}{\sum_{j=1}^{n} \pi_j w_j} \right), \tag{5}$$

with $\sum_{i=1}^{n} w_i = 1$ and $w_i \ge 0$. The maximum likelihood estimator of w_i using the empirical likelihood (5) is

$$w_i^* = \frac{\pi_i^{-1}}{\sum_{j=1}^{n} \pi_j^{-1}}, \tag{6}$$

which reduces to the Hájek estimator of the population mean.

Using the likelihood function (5), the empirical likelihood calibration estimator can be derived as a constrained NPMLE for the distribution function of the finite population. The constrained maximization problem can be formulated as maximizing (5) subject to the constraints in (4). Using the Lagrange multiplier method, the objective function to be minimized is

$$\begin{aligned} Q(w) &= \sum_{i=1}^{n} \log(\pi_i w_i) - n \log\left(\sum_{i=1}^{n} \pi_i w_i\right) \\ &\quad + \lambda_1 \left(\sum_{i=1}^{n} w_i - 1\right) + \lambda_2 \left(\sum_{i=1}^{n} w_i x_i - \mu_x\right). \end{aligned}$$

Setting the partial derivative of Q with respect to w_i equal to zero gives

$$\frac{\partial Q}{\partial w_i} = \frac{1}{w_i} - \frac{n\pi_i}{\sum_{i=1}^n \pi_i w_i} + \lambda_1 + \lambda_2 x_i = 0.$$

Using $\sum_{i=1}^n w_i \left(\partial Q / \partial w_i\right) = 0$, we have

$$w_i = \frac{1}{\lambda_1 \pi_i + \lambda_2 \left(x_i - \mu_x\right)}, \tag{7}$$

where the λ_i, $i = 1, 2$, are the solutions to

$$g_1\left(\lambda_1, \lambda_2\right) =: \sum_{i=1}^n \frac{1}{\lambda_1 \pi_i + \lambda_2 \left(x_i - \mu_x\right)} = 1 \tag{8}$$

and

$$g_1\left(\lambda_1, \lambda_2\right) =: \sum_{i=1}^n \frac{x_i - \mu_x}{\lambda_1 \pi_i + \lambda_2 \left(x_i - \mu_x\right)} = 0. \tag{9}$$

The notation $A =: B$ means that B is defined to be equal to A. A modified Newton-Raphson method can be used to solve the nonlinear equations (8) and (9). See Appendix A.

Chen and Sitter (1999) also considered unequal probability sampling and proposed the pseudo empirical likelihood estimator. Instead of maximizing (5), they proposed maximizing

$$L\left(w\right) = \sum_{i=1}^n \frac{1}{\pi_i} \log\left(w_i\right), \tag{10}$$

subject to the same constraints (4). The resulting pseudo empirical maximum likelihood estimator (PEMLE) for the mean of y is $\bar{y}_{PEMLE} = \sum_{i=1}^n w_i y_i$ where

$$w_i = \frac{1}{\pi_i \left(\lambda_1 + \lambda_2 x_i\right)} \tag{11}$$

where λ_1 and λ_2 satisfy (4). Because the Chen and Sitter (1999) method lacks the maximum likelihood interpretation, we expect that our method is more efficient in large samples. Efficiency will be investigated further in the next section.

3. Asymptotic Properties

We now study the asymptotic properties of the calibration NPMLE estimator of the population mean. To discuss the asymptotic properties of the empirical likelihood estimator, assume a sequence of finite populations with finite fourth moments of as defined in Isaki and Fuller (1982).

Assume the sampling mechanism satisfies

$$K_1 < \max_i \left\{n^{-1} N \pi_i\right\} < K_2 \tag{12}$$

for some positive constants K_1 and K_2, where N is the number of elements in the finite population. Let $u_i = x_i - \mu_x$ and assume that

$$\max_i |u_i| = o_p\left(n^{-1/2}\right) \tag{13}$$

and

$$\frac{\sum_{i=1}^n \pi_i^{-1} u_i}{\sum_{i=1}^n \pi_i^{-1} u_i^2} = O_p\left(n^{-1/2}\right). \tag{14}$$

Although assumptions (13) and (14) can be derived from the existence of the second moment and from the consistency of the Horvitz-Thompson estimator, we state them for convenience in the derivation.

Under assumptions similar to (12) - (14), Chen and Sitter (1999) proved that their pseudo empirical likelihood estimator is asymptotically equivalent to the generalized regression (GREG) estimator $\bar{y}_{GREG}$, where

$$\bar{y}_{GREG} = \bar{x}_\pi + \left(\mu_x - \bar{x}_\pi\right) \hat{B} \tag{15}$$

and

$$\left(\bar{x}_\pi, \bar{y}_\pi\right) = \left(\sum_{i=1}^n \pi_i^{-1}\right)^{-1} \sum_{i=1}^n \pi_i^{-1} \left(x_i, y_i\right)$$

$$\hat{B} = \frac{\sum_{i=1}^n \pi_i^{-1} \left(x_i - \bar{x}_\pi\right)\left(y_i - \bar{y}_\pi\right)}{\sum_{i=1}^n \pi_i^{-1} \left(x_i - \bar{x}_\pi\right)^2}.$$

The following theorem states some asymptotic properties of the calibration NPMLE using the weights in (7) with (8).

Theorem 1 *Under the assumptions (12)-(14), the NPMLE of the mean of y is asymptotically equivalent to*

$$\bar{y}_{opt} = \bar{y}_\pi + \left(\mu_x - \bar{x}_\pi\right) \hat{B}^* \tag{16}$$

where

$$\hat{B}^* = \frac{\sum_{i=1}^n \pi_i^{-2} \left(x_i - \mu_x\right)\left(y_i - \bar{y}_\pi\right)}{\sum_{i=1}^n \pi_i^{-2} \left(x_i - \mu_x\right)^2}.$$

Here, $\left(\bar{x}_\pi, \bar{y}_\pi\right)$ is defined after (15).

The proof of the Theorem is given in Appendix B.

Roughly speaking, the $\hat{B}$ in the GREG estimator (15) estimates the population slope for the regression of y on x. On the other hand, in many sampling designs, the $\hat{B}^*$ in (16) estimates $\left[Var\left(\bar{x}_\pi\right)\right]^{-1} Cov\left(\bar{x}_\pi, \bar{y}_\pi\right)$, which leads to the ‘optimal’ estimator discussed by Rao (1994). Zhong and Rao (2000) also derived an optimal estimator very similar to (16) under stratified random sampling. The idea of using π_i^{-2} to compute the regression coefficient also appears in Isaki and Fuller (1982).

4. Variance estimation

For variance estimation, we use the linearization method of Binder (1983). When the parameter of interest is the population mean, the NPMLE can be written as the solution to the estimating equation

$$U_1(\theta, \lambda_1, \lambda_2) \equiv \sum_{i=1}^{n} \frac{d_i y_i}{\lambda_1 + \lambda_2 d_i u_i} - \theta = 0 \qquad (17)$$

where $d_i = \pi_i^{-1}$ and (λ_1, λ_2) is the solution to the simultaneous estimating equation :

$$U_2(\theta, \lambda_1, \lambda_2) \equiv \sum_{i=1}^{n} \frac{d_i}{\lambda_1 + \lambda_2 d_i u_i} - 1 = 0 \qquad (18)$$

and

$$U_3(\theta, \lambda_1, \lambda_2) \equiv \sum_{i=1}^{n} \frac{d_i u_i}{\lambda_1 + \lambda_2 d_i u_i} = 0. \qquad (19)$$

Using the linearization method of Binder (1983), the variance-covariance matrix of $\left(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2\right)'$ can be written

$$J^{-1}\Sigma_U(\theta, \lambda_1, \lambda_2)\left(J^{-1}\right)' \qquad (20)$$

where

$$J = \begin{pmatrix} \partial U_1/\partial\theta & \partial U_1/\partial\lambda_1 & \partial U_1/\partial\lambda_2 \\ \partial U_2/\partial\theta & \partial U_2/\partial\lambda_1 & \partial U_2/\partial\lambda_2 \\ \partial U_3/\partial\theta & \partial U_3/\partial\lambda_1 & \partial U_3/\partial\lambda_2 \end{pmatrix}$$

and $\Sigma_U(\theta, \lambda_1, \lambda_2)$ is the variance-covariance matrix of $(U_1, U_2, U_3)'$ treating $\theta, \lambda_1, \lambda_2$ as constants. Using $\partial U_2/\partial\theta = \partial U_3/\partial\theta = 0$, we have

$$J^{-1} = \begin{pmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ \mathbf{0} & A_{22}^{-1} \end{pmatrix}.$$

where $A_{11} = \partial U_1/\partial\theta$, $A_{12} = (\partial U_1/\partial\lambda_1, \partial U_1/\partial\lambda_2)$ and

$$A_{22} = \begin{pmatrix} \frac{\partial U_2}{\partial\lambda_1} & \frac{\partial U_2}{\partial\lambda_2} \\ \frac{\partial U_3}{\partial\lambda_1} & \frac{\partial U_3}{\partial\lambda_2} \end{pmatrix}.$$

Thus, after some algebra, we have

$$V\left(\hat{\theta}\right) = A_{11}^{-1}\left[(1, -B_{12})\Sigma_U(\theta, \lambda_1, \lambda_2)(1, -B_{12})'\right]A_{11}^{-1}. \qquad (21)$$

where $A_{11} = -1$, $B_{12} = A_{12}A_{22}^{-1}$ and

$$A_{12} = -\sum_{i=1}^{n} \frac{d_i}{(\lambda_1 + \lambda_2 d_i u_i)^2}(1, d_i u_i)\, y_i \qquad (22)$$

$$A_{22} = -\sum_{i=1}^{n} \frac{d_i}{(\lambda_1 + \lambda_2 d_i u_i)^2}\begin{pmatrix} 1 & d_i u_i \\ u_i & d_i u_i^2 \end{pmatrix}. \qquad (23)$$

Let $\hat{V} = \sum_{i=1}^{n}\sum_{j=1}^{n}\Omega_{ij}$ be the design unbiased variance estimator of $N^{-1}\sum_{i=1}^{n} d_i y_i$. Using (21), a "plug-in" variance estimator can be derived as

$$\hat{V}\left(\hat{\theta}\right) = \hat{A}_{11}^{-1}\left[\left(I, -\hat{B}_{12}\right)\hat{\Sigma}_U\left(I, -\hat{B}_{12}\right)'\right]\hat{A}_{11}^{-1}, \qquad (24)$$

where $\hat{A}_{11}^{-1} = 1$, $\hat{B}_{12} = \hat{A}_{12}\hat{A}_{22}^{-1}$, and $\hat{A}_{12}$ and $\hat{A}_{22}$ are derived from (22) and (23), respectively, with (λ_1, λ_2) replaced by $\left(\hat{\lambda}_1, \hat{\lambda}_2\right)$. Since the component Σ_U should represent the sampling variance of U terms only, we propose that $\hat{\Sigma}_U$ be computed as

$$\hat{\Sigma}_U = \sum_{i=1}^{n}\sum_{j=1}^{n}\Omega_{ij}\mathbf{h}_i\mathbf{h}_j' \qquad (25)$$

where $\mathbf{h}_i = (y_i, 1, u_i)$. Therefore, combining (24) and (25), we have

$$\hat{V}\left(\hat{\theta}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\Omega_{ij}e_i e_j \qquad (26)$$

where $e_i = y_i - \hat{y}_i$ with $\hat{y}_i = \hat{A}_{12}\hat{A}_{22}^{-1}(1, u_i)'$. In some calibration literature, this method has been called the residual technique because the standard variance estimator is applied to the residuals. The only difference here is that the regression coefficients are computed differently.

5. Simulation Studies

To study the properties of the proposed calibration estimator, we performed a limited simulation study. In the simulation study, four artificial finite populations for (x_i, y_i, z_i) of size $N = 10,000$ are generated. The populations are

[A] $z_i \sim \chi^2(2)$

$$\begin{aligned} x_i &= a_i + 0.5z_i + 2 \\ y_i &= 1 + \sqrt{0.5}(x_i - 3) + e_i \end{aligned}$$

[B] (x_i, z_i) are the same as in population [A] and $y_i = (x_i - 3)^2 + e_i$

[C] $z_i \sim \chi^2(2) + 2$

$$\begin{aligned} x_i &= a_i + 0.5z_i + 1 \\ y_i &= 1 + \sqrt{0.5}(x_i - 3) + e_i \end{aligned}$$

[D] (x_i, z_i) are the same as in population [C] and $y_i = (x_i - 3)^2 + e_i$

where $a_i \sim N(0,1)$, independent of z_i, and $e_i \sim N(0,1)$, independent of (u_i, a_i, z_i), in the four populations.

From each of the finite population generated above, Probability Proportional to Size (PPS) samples of size $n = 200$ and $n = 500$ are generated where the probability of selecting a single element p_i is proportional to z_i. We assume that the population mean of x_i is known and is used for the calibration. From each sample, four estimators of the population mean of y are computed. The estimators are the Hansen-Hurwitz (HH) estimator for the PPS sampling using weight $d_i = N^{-1}n^{-1}p_i^{-1}$, where $p_i = z_i / \left(\sum_{i=1}^{N} z_i\right)$, the GREG estimator defined in (15), the pseudo empirical likelihood estimator (PEMLE) of Chen and Sitter (1999) defined in (11), and the proposed NPMLE defined in (7).

Variance estimators are also computed for the last two point estimators. The variance estimator for the PEMLE estimator, derived using the same arguments in Section 4, is computed by the residual method with the residual

$$e_i = y_i - \hat{y}_i$$

where $\hat{y}_i = \hat{A}_{12}\hat{A}_{22}^{-1}(1, u_i)'$

$$\hat{A}_{12} = -\sum_{i=1}^{n} \frac{d_i}{\left(\hat{\lambda}_1 + \hat{\lambda}_2 u_i\right)^2} (1, u_i)\, y_i$$

$$\hat{A}_{22} = -\sum_{i=1}^{n} \frac{d_i}{\left(\hat{\lambda}_1 + \hat{\lambda}_2 u_i\right)^2} \begin{pmatrix} 1 & u_i \\ u_i & u_i^2 \end{pmatrix}.$$

The variance estimator for the NPMLE is computed using the residual method described in (26).

Table 1 reports the simulation results of the four point estimators. Table 2 reports the relative biases and the t-statistics of the two variance estimators. The relative bias is the Monte Carlo bias divided by the Monte Carlo variance of the point estimator. The t-statistic is the statistic used to test the significance of the Monte Carlo bias of the variance estimator.

From the results in Table 1 and in Table 2, we have the following conclusions.

1. The HH estimators in population A and B have bigger variances than those in population C and D. Since the z-variables are highly variable in population A and B, the resulting sampling weights for the HH estimator are also highly variable and increase the variances of the resulting HH estimators.

2. In population C, the three calibration estimators show similar performances because the weights are relatively homogeneous. The ratio of the variance of the calibration estimator to the variance of the HH estimator is about 0.5, which is consistent with the theory because the population correlation between x and y is equal to $\sqrt{0.5}$.

3. In population A and B, the NPMLE shows better performance than the other calibration estimators. Note that the two empirical likelihood estimators can be written

$$\hat{\theta}_{PEMLE} = \sum_{i=1}^{n} \frac{d_i y_i}{\lambda_1 + \lambda_2 u_i}$$
$$\hat{\theta}_{NPMLE} = \sum_{i=1}^{n} \frac{d_i y_i}{\lambda_1 + \lambda_2 d_i u_i}.$$

Thus, the PEMLE will be efficient if $y_i \propto d_i^{-1} x_i$, while the NPMLE will be efficient if $y_i \propto x_i$. If the design weights d_i are highly, the PEMLE can be very inefficient. Therefore, the NPMLE will be less sensitive to extreme design weights.

4. In population D, where the design weights are relatively homogeneous and the linear relationship between y and x does not hold, the calibration estimators do not improve the efficiency of the HH estimator. In population B, the NPMLE is more efficient than the HH estimator because the efficiency of the HH estimator is mitigated by the extreme weights.

5. The variance estimator for the NPMLE shows good performances in terms of the relative biases. The variance estimator for PEMLE shows significant biases for population B. Roughly speaking, when some of the design weights are extremely large, the condition (14) does not hold and the second order term in the Taylor linearization is no longer negligible unless linear relationship between the study variable and the control variable holds.

Appendix

A. Algorithm

We propose solving the nonlinear equations (8) - (9) by a modified Newton-Raphson method as follows:

[Step 1] Set $\lambda_1^{(0)} = \sum_{i=1}^{n} \pi_i^{-1}$ and $\lambda_2^{(0)} = 0$. Also, set $\gamma = 1$.

[Step 2] Compute the updated values of λ's iteratively as

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} + \gamma \left[\Delta^{(k)}\right]^{-1} \left(\mathbf{g}^* - \mathbf{g}^{(k)}\right)$$

where $\boldsymbol{\lambda}^{(k)} = \left(\lambda_1^{(k)}, \lambda_2^{(k)}\right)'$, $\Delta^{(k)}$ is a 2×2 matrix of $\partial g_i / \partial \lambda_j$ evaluated at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(k)}$, $\mathbf{g}^* = (1,0)'$, and $\mathbf{g}^{(k)} = \left[g_1\left(\boldsymbol{\lambda}^{(k)}\right), g_2\left(\boldsymbol{\lambda}^{(k)}\right)\right]'$.

[Step 3] If $\lambda_1^{(k+1)}\pi_i + \lambda_2^{(k+1)}(x_i - \mu_x) < 0$ for some $i = 1, 2, \cdots, n$, then set $\gamma = \gamma/2$ and go to Step 2.

[Step 4] If $\max_i \left|\lambda_i^{(k+1)} - \lambda_i^{(k)}\right| < \epsilon$ for sufficiently small $\epsilon > 0$, stop and compute the final weights from (7) using $\boldsymbol{\lambda}^{(k+1)}$. Otherwise, set $k = k+1$ and $\gamma = 1$, and go to Step 2.

Step 2 essentially describes a Newton-Raphson solution to the nonlinear equations (8) through (9). Step 3 guarantees that the resulting calibration weights be always positive. If we further want to restrict the weights to be $w_i \in [L_w, U_w]$ for given L_w and U_w values, it is enough to check $U_w^{-1} \leq \lambda_1^{(k+1)}\pi_i + \lambda_2^{(k+1)}(x_i - \mu_x) \leq L_w^{-1}$ in Step 3. See also Chen et al (2002).

B. Proof of Theorem 1

First note that the two constraints, (8) and (9), can be written

$$\sum_{i=1}^{n} w_i = 1 \quad \text{(A.1)}$$

$$\sum_{i=1}^{n} w_i u_i = 0. \quad \text{(A.2)}$$

The NPMLE of the population mean of y can be written

$$\bar{y}_{NPMLE} = \frac{N^{-1}\sum_{i=1}^{n}\left(1+\delta\pi_i^{-1}u_i\right)^{-1}\pi_i^{-1}y_i}{N^{-1}\sum_{i=1}^{n}\left(1+\delta\,\pi_i^{-1}u_i\right)^{-1}\pi_i^{-1}} \quad \text{(A.3)}$$

where $\delta = \lambda_2/\lambda_1$ and N is the size of the finite population. By (A.2), the δ satisfies

$$\sum_{i=1}^{n} \frac{\pi_i^{-1}u_i}{1+\delta\pi_i^{-1}u_i} = 0. \quad \text{(A.4)}$$

Using the argument of Owen (1990, p 100-101), it can be shown that $\delta = O_p\left(n^{-1/2}\right)$ and

$$\delta = \left(\sum_{i=1}^{n} \pi_i^{-2}u_i^2\right)^{-1} \sum_{i=1}^{n} \pi_i^{-1}u_i + o_p\left(n^{-1/2}\right) \quad \text{(A.5)}$$

Let $\gamma_i = \delta\pi_i^{-1}u_i$. The numerator part of (A.3) can be written

$$\begin{aligned} & N^{-1}\sum_{i=1}^{n}\frac{y_i}{\pi_i}\left(1-\gamma_i+\frac{\gamma_i^2}{1+\gamma_i}\right) \\ &= N^{-1}\sum_{i=1}^{n}\frac{y_i}{\pi_i}(1-\gamma_i) + o_p\left(n^{-1/2}\right). \end{aligned}$$

Similarly, the denominator part can be written

$$\begin{aligned} & N^{-1}\sum_{i=1}^{n}\frac{1}{\pi_i}\left(1-\gamma_i+\frac{\gamma_i^2}{1+\gamma_i}\right) \\ =\ & N^{-1}\sum_{i=1}^{n}\frac{1}{\pi_i}(1-\gamma_i) + o_p\left(n^{-1/2}\right). \end{aligned}$$

Hence, the NPMLE is asymptotically equivalent to

$$\frac{\sum_{i=1}^{n} y_i\pi_i^{-1}(1-\gamma_i)}{\sum_{i=1}^{n}\pi_i^{-1}(1-\gamma_i)}$$

and is also asymptotically equivalent to (16).

References

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, **52**, 200-203.

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 estimator for surveys. *Biometrika*, **89** 230-237.

Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, **80**, 107-116.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. New York: Wiley.

Deville, J. C. and Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, **87**, 376-382.

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

Hartley, H. O. and Rao, J.N.K. (1968). A new estimation theory for sample surveys, *Biometrika*, **55**, 547-55.

Fuller, W. A. (2002). Regression estimation for sample surveys. *Survey Methodology*, **28**, 5-23.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237-249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, **18**, 90-120.

Rao, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of Official Statistics*, **10**, 153-165.

Wu, C. and Sitter, R.R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, **96**, 185-193.

Zhong, B. and Rao, J.N.K. (2000). Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika*, **87**, 929-938.

Table 1: Monte Carlo Biases, Variances, and Mean squared errors of the point estimators for the sample, based on 5,000 samples.

n	Pop'n	Estimator	Bias	MSE
200	A	HH	0.00	0.0522
		GREG	0.00	0.0183
		PEMLE	0.00	0.0188
		NPMLE	0.00	0.0179
	B	HH	0.01	0.2877
		GREG	-0.08	0.1872
		PEMLE	0.05	0.3860
		NPMLE	-0.01	0.0972
	C	HH	0.00	0.00926
		GREG	0.00	0.00612
		PEMLE	0.00	0.00612
		NPMLE	0.00	0.00615
	D	HH	0.00	0.0414
		GREG	-0.02	0.0537
		PEMLE	0.01	0.0546
		NPMLE	0.00	0.0459
500	A	HH	0.00	0.01905
		GREG	0.00	0.00840
		PEMLE	0.00	0.00850
		NPMLE	0.00	0.00804
	B	HH	0.00	0.4911
		GREG	-0.04	0.0873
		PEMLE	0.03	0.1718
		NPMLE	-0.01	0.0427
	C	HH	0.00	0.00359
		GREG	0.00	0.00236
		PEMLE	0.00	0.00236
		NPMLE	0.00	0.00236
	D	HH	0.00	0.0161
		GREG	-0.01	0.0211
		PEMLE	0.01	0.0212
		NPMLE	0.00	0.0179

Table 2: Monte Carlo relative bias and the t-statistics of the variance estimators for the sample, based on 5,000 samples.

n	Pop'n	Estimator	Rel. Bias	t-statistic
200	A	PEMLE	0.012	0.41
		NPMLE	-0.005	-0.08
	B	PEMLE	1.561	8.55
		NPMLE	-0.054	-1.49
	C	PEMLE	-0.030	-1.51
		NPMLE	-0.035	-1.79
	D	PEMLE	0.046	2.26
		NPMLE	-0.014	-0.68
500	A	PEMLE	0.035	1.62
		NPMLE	-0.016	-0.42
	B	PEMLE	1.693	7.12
		NPMLE	-0.009	-0.24
	C	PEMLE	0.010	0.52
		NPMLE	0.008	0.38
	D	PEMLE	0.029	1.48
		NPMLE	0.007	0.34