Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Calibration estimation using exponential tilting in sample surveys

by Jae Kwang Kim

December 2010

Statistics Canada Statistique Canada

Canada

Survey Methodology, December 2010
Vol. 36, No. 2, pp. 145-155
Statistics Canada, Catalogue No. 12-001-X

Calibration estimation using exponential tilting in sample surveys

Jae Kwang Kim [1]

Abstract

We consider the problem of parameter estimation with auxiliary information, where the auxiliary information takes the form of known moments. Calibration estimation is a typical example of using the moment conditions in sample surveys. Given the parametric form of the original distribution of the sample observations, we use the estimated importance sampling of Henmi, Yoshida and Eguchi (2007) to obtain an improved estimator. If we use the normal density to compute the importance weights, the resulting estimator takes the form of the one-step exponential tilting estimator. The proposed exponential tilting estimator is shown to be asymptotically equivalent to the regression estimator, but it avoids extreme weights and has some computational advantages over the empirical likelihood estimator. Variance estimation is also discussed and results from a limited simulation study are presented.

Key Words: Benchmarking estimator; Empirical likelihood; Instrumental variable calibration; Importance sampling; Regression estimator.

1. Introduction

Consider the problem of estimating $Y = \sum_{i=1}^{N} y_i$ for a finite population of size N. Let A denote the index set of the sample obtained by a probability sampling scheme. In addition to y_i, suppose that we also observe a p-dimensional auxiliary vector $\mathbf{x}_i$ in the sample such that $\mathbf{X} = \sum_{i=1}^{N} \mathbf{x}_i$ is known from an external source. We are interested in estimating Y using the auxiliary information $\mathbf{X}$.

The Horvitz-Thompson (HT) estimator of the form

$$\hat{Y}_d = \sum_{i \in A} d_i \, y_i, \qquad (1)$$

where $d_i = 1/\pi_i$ is the design weight and π_i is the first order inclusion probability, is unbiased for Y. But, it does not make use of the information given by $\mathbf{X}$. According to Kott (2006), a calibration estimator can be defined as the estimator of the form

$$\hat{Y}_w = \sum_{i \in A} w_i \, y_i$$

where the weights w_i satisfy

$$\sum_{i \in A} w_i \, \mathbf{x}_i = \mathbf{X} \qquad (2)$$

and $\hat{Y}_w$ is asymptotically design unbiased (ADU). Calibration estimation has become very popular in survey sampling because it provides consistency across different surveys and often improves the efficiency. (Särndal 2007).

The regression estimator, using the weights

$$w_i = d_i + (\mathbf{X} - \hat{\mathbf{X}}_d)' \left(\sum_{j \in A} d_j \, \mathbf{x}_j \, \mathbf{x}_j' \right)^{-1} d_i \, \mathbf{x}_i, \qquad (3)$$

obtained by minimizing

$$\sum_{i \in A} (w_i - d_i)^2 / d_i$$

subject to constraint (2), is asymptotically design unbiased. Note that if an intercept term is included in the column space of X matrix then (2) implies that the population size N is known. If N is unknown, one can require that the sum of the final weights are equal to the sum of the design weights. Thus,

$$\sum_{i \in A} w_i = \hat{N}, \qquad (4)$$

where

$$\hat{N} = \begin{cases} N & \text{if } N \text{ is known} \\ \sum_{i \in A} d_i & \text{otherwise,} \end{cases}$$

can be imposed as a constraint in addition to (2), which yields the weights

$$w_i = \frac{\hat{N}}{\hat{N}_d} d_i + \left(\mathbf{X} - \frac{\hat{N}}{\hat{N}_d} \hat{\mathbf{X}}_d \right)' \left\{ \sum_{j \in A} d_j (\mathbf{x}_j - \bar{\mathbf{X}}_d)(\mathbf{x}_j - \bar{\mathbf{X}}_d)' \right\}^{-1} d_i (\mathbf{x}_i - \bar{\mathbf{X}}_d), \qquad (5)$$

where $\hat{\mathbf{X}}_d = \sum_{i \in A} d_i \, \mathbf{x}_i$, $\hat{N}_d = \sum_{i \in A} d_i$, and $\bar{\mathbf{X}}_d = \hat{\mathbf{X}}_d / \hat{N}_d$. We define the regression estimator to be $\hat{Y}_{\text{reg}} = \sum_{i \in A} w_i \, y_i$ using the weights (5). The regression estimator can be efficient if y_i is linearly related with $\mathbf{x}_i$ (Isaki and Fuller 1982; Fuller 2002), but the weights in the regression estimator can take negative or extremely large values.

1. Jae Kwang Kim, Department of Statistics, Iowa State University, Ames, Iowa, 50011, U.S.A. E-mail: jkim@iastate.edu.

The empirical likelihood (EL) calibration estimator, discussed by Chen and Qin (1993), Chen and Sitter (1999), Wu and Rao (2006), and Kim (2009), is obtained by maximizing the pseudo empirical likelihood

$$\sum_{i \in A} d_i \ln(w_i)$$

subject to constraints (2) and (4). The solution to the optimization problem can be written as

$$w_i = d_i \frac{1}{\lambda_0 + \boldsymbol{\lambda}_1' (\mathbf{x}_i - \mathbf{X}/\hat{N})}, \tag{6}$$

where λ_0 and $\boldsymbol{\lambda}_1$ satisfy constraints (2), (4), and $w_i > 0$ for all i. The EL calibration estimator is asymptotically equivalent to the regression estimator using weights (5) and avoids negative weights if a solution exists, but can result in extremely large weights.

Because the empirical likelihood method requires solving nonlinear equations, the computation can be cumbersome. Furthermore, in some extreme cases, $\bar{\mathbf{X}} = N^{-1} \sum_{i=1}^N \mathbf{x}_i$ does not belong to the convex hull of the sample $\mathbf{x}_i$'s and the solution does not exist. In this extreme situation, the constraint (2) can be relaxed.

Rao and Singh (1997) solved a similar problem by allowing

$$\left| \sum_{i \in A} w_i x_{ij} - X_j \right| \le \delta_j X_j, \quad j = 1, 2, \ldots, p,$$

for some small tolerance level $\delta_j > 0$ where $X_j = \sum_{i=1}^N x_{ij}$. Note that the choice of $\delta_j = 0$ leads to the exact calibration condition (2). Rao and Singh (1997) chose the tolerance level δ_j using a shrinkage factor in the ridge regression but their approach does not directly apply to the empirical likelihood method and the choice of δ_j is somewhat unclear. Chambers (1996) and Beaumont and Bocci (2008) also discussed a ridge regression estimation in the context of avoiding extreme weights. Breidt, Claeskens and Opsomer (2005) used penalized spline approach to obtain the ridge calibration. Recently, Park and Fuller (2009) developed a method of obtaining the shrinkage factor δ_j using a regression superpopulation model with random components.

Chen, Variyath and Abraham (2008) tackled a similar problem in the context of the empirical likelihood method and proposed a solution by adding an artificial point such that $\bar{\mathbf{X}} = N^{-1} \sum_{i=1}^N \mathbf{x}_i$ would belong to the convex hull of the augmented $\mathbf{x}_i$'s. The proposed estimator in Chen *et al.* (2008) only satisfies the calibration property approximately in the sense that

$$\sum_{i \in A} w_i \mathbf{x}_i - \mathbf{X} = o_p(n^{-1/2} N). \tag{7}$$

This approximate calibration property is attractive because it allows more generality in the choice of weights. In particular, when the dimension of the auxiliary variable $\mathbf{x}$ is large the calibration constraint (2) can be quite restrictive. As can be seen in Section 2, an estimator satisfying the asymptotic calibration property (7) enjoys most of the desirable properties of the empirical likelihood calibration estimator and is computationally efficient.

In this paper, we consider a class of empirical-likelihood-type estimators that satisfy the approximate calibration property (7). In Section 2, the idea of estimated importance sampling of Henmi *et al.* (2007) is discussed and a new estimator using this methodology is proposed. In Section 3, a weight trimming technique to avoid extreme calibration weights is proposed. In Section 4, variance estimation of the proposed estimator is discussed. In Section 5, results from a simulation study are presented. Concluding remarks are made in Section 6.

2. Proposed method

To introduce the proposed method, we first discuss estimated importance sampling introduced by Henmi *et al.* (2007). Suppose that $\mathbf{x}_i$ is observed throughout the population but y_i is observed only in the sample. We assume a superpopulation model for $\mathbf{x}_i$ with density $f(\mathbf{x}; \boldsymbol{\eta})$ known up to a parameter $\boldsymbol{\eta} \in \Omega$. The superpopulation model characterized by the density $f(\mathbf{x}; \boldsymbol{\eta})$ is a working model in the sense that the model is used to derive a model-assisted estimator (Särndal, Swenson and Wretman 1992).

Let $\hat{\boldsymbol{\eta}}$ be the pseudo maximum likelihood estimator of $\boldsymbol{\eta}$ computed from the sample

$$\hat{\boldsymbol{\eta}} = \arg \max_{\Omega} \sum_{i \in A} d_i \ln \{ f(\mathbf{x}_i; \boldsymbol{\eta}) \}$$

and let $\boldsymbol{\eta}_{0,N}$ be the maximum likelihood estimator of $\boldsymbol{\eta}$ computed from the population

$$\boldsymbol{\eta}_{0,N} = \arg \max_{\Omega} \sum_{i=1}^{N} \ln \{ f(\mathbf{x}_i; \boldsymbol{\eta}) \}.$$

Following Henmi *et al.* (2007), we can construct the following estimated importance weight

$$w_i = d_i \frac{f(\mathbf{x}_i; \boldsymbol{\eta}_{0N})}{f(\mathbf{x}_i; \hat{\boldsymbol{\eta}})}. \tag{8}$$

To discuss the asymptotic properties of the estimator using the weights in (8), assume a sequence of the finite populations and the samples, as in Isaki and Fuller (1982), such that

$$\sum_{i \in A} d_i (\mathbf{x}_i', y_i)' (\mathbf{x}_i', y_i) - \sum_{i=1}^{N} (\mathbf{x}_i', y_i)' (\mathbf{x}_i', y_i) = O_p (n^{-1/2} N)$$

for all possible A and for each N. The following theorem presents some asymptotic properties of the estimator with the estimated importance weights in (8).

Theorem 1. *Under the regularity conditions given in Appendix A, the estimator* $\hat{Y}_w = \sum_{i \in A} w_i y_i$, *with the* w_i *defined by* (8), *satisfies*

$$\sqrt{n} N^{-1} (\hat{Y}_w - \hat{Y}_l) = o_p(1), \tag{9}$$

where

$$\hat{Y}_l = \hat{Y}_d - \hat{\boldsymbol{\Sigma}}_{sy}' \hat{\boldsymbol{\Sigma}}_{ss}^{-1} \hat{\mathbf{S}}_{0d}, \tag{10}$$

$\hat{Y}_d$ *is defined in* (1), $\hat{\mathbf{S}}_{0d} = \sum_{i \in A} d_i \mathbf{s}_{i0}$, $\hat{\boldsymbol{\Sigma}}_{sy} = N^{-1} \sum_{i \in A} d_i \mathbf{s}_{i0} y_i$, *and* $\hat{\boldsymbol{\Sigma}}_{ss} = N^{-1} \sum_{i \in A} d_i \mathbf{s}_{i0}^{\otimes 2}$. *Here,* $\mathbf{s}_{i0} = \partial \ln f(\mathbf{x}_i; \boldsymbol{\eta}) / \partial \boldsymbol{\eta} |_{\boldsymbol{\eta} = \boldsymbol{\eta}_{0,N}}$ *and the notation* $B^{\otimes 2}$ *denotes* BB'.

The proof of Theorem 1 is presented in Appendix A. Because $\mathbf{S}_{0N} \equiv \sum_{i=1}^{N} \mathbf{s}_{i0} = \mathbf{0}$, we can write (10) as

$$\hat{Y}_l = \hat{Y}_d + \hat{\boldsymbol{\Sigma}}_{sy}' \hat{\boldsymbol{\Sigma}}_{ss}^{-1} (\mathbf{S}_{0N} - \hat{\mathbf{S}}_{0d}),$$

which is a regression estimator of Y using $\mathbf{s}_i(\boldsymbol{\eta}_{0N})$ as the auxiliary variable. Therefore, under regularity conditions, the proposed estimator using estimated importance sampling is asymptotically unbiased and has asymptotic variance no greater than that of the direct estimator $\hat{Y}_d$. Note that the validity of Theorem 1 does not require that the working model $f(\mathbf{x}; \boldsymbol{\eta})$ be true.

If the density of $\mathbf{x}_i$ is a multivariate normal density, then the weights in (8) become

$$w_i = d_i \frac{\phi(\mathbf{x}_i; \bar{\mathbf{X}}_N, \boldsymbol{\Sigma}_{xx,N})}{\phi(\mathbf{x}_i; \bar{\mathbf{X}}_d, \hat{\boldsymbol{\Sigma}}_{xx,d})}, \tag{11}$$

where $\bar{\mathbf{X}}_d$ is defined after (5), $\hat{\boldsymbol{\Sigma}}_{xx,d} = \sum_{i \in A} d_i (\mathbf{x}_i - \bar{\mathbf{X}}_d)^{\otimes 2} / \hat{N}_d$, $\boldsymbol{\Sigma}_{xx,N} = \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{X}}_N)^{\otimes 2} / N$, and $\phi(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of the multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{\Sigma}_{xx,N}$ is unknown and only $\bar{\mathbf{X}}_N$ is available, then we can use

$$w_i = d_i \frac{\phi(\mathbf{x}_i; \bar{\mathbf{X}}_N, \hat{\boldsymbol{\Sigma}}_{xx,d})}{\phi(\mathbf{x}_i; \bar{\mathbf{X}}_d, \hat{\boldsymbol{\Sigma}}_{xx,d})}. \tag{12}$$

Tillé (1998) derived weights similar to those in (12) in the context of conditional inclusion probabilities.

In general, the parametric model for $\mathbf{x}_i$ is unknown. Thus, we consider an approximation for the importance weights in (8) using the Kullback-Leibler information criterion for distance. Let $f(\mathbf{x})$ be a given density for $\mathbf{x}$ and let P_0 be the set of densities that satisfy the calibration constraint. That is,

$$P_0 = \left\{ f_0(\mathbf{x}); \int f_0(\mathbf{x}) \, d\mathbf{x} = 1, \int \mathbf{x} f_0(\mathbf{x}) \, d\mathbf{x} = \bar{\mathbf{X}}_N \right\}.$$

The optimization problem using Kullback-Leibler distance can be expressed as

$$\min_{f_0 \in P_0} \int f_0(\mathbf{x}) \ln \left\{ \frac{f_0(\mathbf{x})}{f(\mathbf{x})} \right\} d\mathbf{x}. \tag{13}$$

The solution to (13) is

$$f_0(\mathbf{x}) = f(\mathbf{x}) \frac{\exp(\hat{\boldsymbol{\lambda}}' \mathbf{x})}{E\{\exp(\hat{\boldsymbol{\lambda}}' \mathbf{x})\}} \tag{14}$$

where $\hat{\boldsymbol{\lambda}}$ satisfies $\int \mathbf{x} f_0(\mathbf{x}) \, d\mathbf{x} = \bar{\mathbf{X}}_N$. Thus, the estimated importance weights in (8) using the optimal density in (14) can be written

$$w_i = d_i \frac{f_0(\mathbf{x}_i)}{f(\mathbf{x}_i)} = d_i \exp(\hat{\lambda}_0 + \hat{\boldsymbol{\lambda}}_1' \mathbf{x}_i) \tag{15}$$

where $\hat{\lambda}_0$ and $\hat{\boldsymbol{\lambda}}_1$ satisfy constraint (2) and (4). The shift from $f(\mathbf{x})$ to $f_0(\mathbf{x})$ in (14) is called exponential tilting. Thus, an estimator using the weight (15) satisfying the calibration constraints (2) and (4) can be called an exponential tilting (ET) calibration estimator. That is, we define the ET calibration estimator as

$$\hat{Y}_{\text{ET}} = \sum_{i \in A} d_i \exp(\hat{\lambda}_0 + \hat{\boldsymbol{\lambda}}_1' \mathbf{x}_i) \, y_i, \tag{16}$$

where $\hat{\lambda}_0$ and $\hat{\boldsymbol{\lambda}}_1$ satisfy constraint (2) and (4). Estimators based on exponential tilting have been used in various contexts. For examples, see Efron (1981), Kitamura and Stutzer (1997), and Imbens (2002). When N is known, Folsom (1991) and Deville, Särndal and Sautory (1993) developed the estimator (16) using a very different approach.

To compute λ_0 and $\boldsymbol{\lambda}_1$ in (16), because of the calibration constraints (2) and (4), we need to solve the following estimating equations:

$$\hat{U}_0(\boldsymbol{\lambda}) \equiv \sum_{i \in A} d_i \exp(\lambda_0 + \boldsymbol{\lambda}_1' \mathbf{x}_i) - \hat{N} = 0 \tag{17}$$

$$\hat{\mathbf{U}}_1(\boldsymbol{\lambda}) \equiv \sum_{i \in A} d_i \exp(\lambda_0 + \boldsymbol{\lambda}_1' \mathbf{x}_i) \mathbf{x}_i - \mathbf{X} = \mathbf{0}, \tag{18}$$

where $\boldsymbol{\lambda}' = (\lambda_0, \boldsymbol{\lambda}_1')$. Writing $\hat{\mathbf{U}}' = (\hat{U}_0, \hat{\mathbf{U}}_1')$, we can use the Newton-type algorithm of the form

$$\hat{\boldsymbol{\lambda}}_{(t+1)} = \hat{\boldsymbol{\lambda}}_{(t)} - \left\{ \frac{\partial}{\partial \boldsymbol{\lambda}'} \hat{\mathbf{U}}(\hat{\boldsymbol{\lambda}}_{(t)}) \right\}^{-1} \hat{\mathbf{U}}(\hat{\boldsymbol{\lambda}}_{(t)})$$

and the solution can be written

$$\hat{\boldsymbol{\lambda}}_{1(t+1)} = \hat{\boldsymbol{\lambda}}_{1(t)} + \left\{ \sum_{i \in A} w_{i(t)} (\mathbf{x}_i - \bar{\mathbf{X}}_{w(t)})^{\otimes 2} \right\}^{-1} \left(\mathbf{X} - \sum_{i \in A} w_{i(t)} \mathbf{x}_i \right), \tag{19}$$

where $w_{i(t)} = d_i \exp(\hat{\lambda}_{0(t)} + \hat{\boldsymbol{\lambda}}'_{1(t)}\mathbf{x}_i)$ and $\bar{\mathbf{X}}_{w(t)} = \sum_{i\in A} w_{i(t)}\mathbf{x}_i / \sum_{i\in A} w_{i(t)}$, with the initial values $\hat{\boldsymbol{\lambda}}_{1(0)} = \mathbf{0}$. Once $\hat{\boldsymbol{\lambda}}_{1(t)}$ is computed by (19), $\hat{\lambda}_{0(t)}$ is computed by

$$\exp(\hat{\lambda}_{0(t)}) = \frac{\hat{N}}{\sum_{i\in A} d_i \exp(\hat{\boldsymbol{\lambda}}'_{1(t)}\mathbf{x}_i)}. \tag{20}$$

Note that, $w_{i(0)} = d_i \hat{N}/\hat{N}_d$ since $\hat{\boldsymbol{\lambda}}_{1(0)} = \mathbf{0}$. Because $\hat{\mathbf{U}}(\boldsymbol{\lambda})$ is twice continuously differentiable and convex in $\boldsymbol{\lambda}$, the sequence $\hat{\boldsymbol{\lambda}}_{(t)}$ always converges if the solution to $\hat{\mathbf{U}}(\boldsymbol{\lambda}) = \mathbf{0}$ exists (Givens and Hoeting 2005). The convergence rate is quadratic in the sense that

$$|\hat{\boldsymbol{\lambda}}_{1(t+1)} - \hat{\boldsymbol{\lambda}}_1| \leq C|\hat{\boldsymbol{\lambda}}_{1(t)} - \hat{\boldsymbol{\lambda}}_1|^2$$

for some constant C, where $\hat{\boldsymbol{\lambda}}_1 = \lim_{t\to\infty} \hat{\boldsymbol{\lambda}}_{1(t)}$.

By construction, the t-step exponential tilting (ET) estimator, defined by

$$\hat{Y}_{\mathrm{ET}(t)} = \sum_{i\in A} d_i \exp(\hat{\lambda}_{0(t)} + \hat{\boldsymbol{\lambda}}'_{1(t)}\mathbf{x}_i)\, y_i \tag{21}$$

where $\hat{\lambda}_{0(t)}$ and $\hat{\boldsymbol{\lambda}}_{1(t)}$ are computed by (19) and (20), satisfies the calibration constraint (2) for sufficiently large t. By the recursive form in (19) with $\hat{\boldsymbol{\lambda}}_{1(0)} = \mathbf{0}$, we can write

$$\hat{\boldsymbol{\lambda}}_{1(t)} = \sum_{j=0}^{t-1} (\mathbf{S}_{xx,w(j)})^{-1}(\tilde{\mathbf{X}}_N - \bar{\mathbf{X}}_{w(j)}), \tag{22}$$

where $\tilde{\mathbf{X}}_N = \mathbf{X}/\hat{N}$ and $\mathbf{S}_{xx,\,w(j)} = \sum_{i\in A} w_{i(t)}(\mathbf{x}_i - \bar{\mathbf{X}}_{w(t)})^{\otimes 2}/\hat{N}$. Thus, the t-step ET estimator (21) can be written as

$$\hat{Y}_{\mathrm{ET}(t)} = \hat{N}\frac{\sum_{i\in A} d_i g_{i(t)} y_i}{\sum_{i\in A} d_i g_{i(t)}},$$

where

$$g_{i(t)} = \prod_{j=0}^{t-1} \frac{\phi\left(\mathbf{x}_i; \tilde{\mathbf{X}}_N, \mathbf{S}_{xx,\,w(j)}\right)}{\phi\left(\mathbf{x}_i; \bar{\mathbf{X}}_{w(j)}, \mathbf{S}_{xx,\,w(j)}\right)}.$$

The following theorem presents some asymptotic properties of the exponential tilting estimator.

Theorem 2. *The t-step ET estimator* (21) *based on equations* (19) *and* (20) *satisfies*

$$\sqrt{n}N^{-1}(\hat{Y}_{\mathrm{ET}(t)} - \hat{Y}_{\mathrm{reg}}) = o_p(1), \tag{23}$$

for each $t = 1, 2, \ldots,$ where $\hat{Y}_{\mathrm{reg}}$ is the regression estimator using the regression weight in (5).

The proof of Theorem 2 is presented in Appendix B. Theorem 2 presents the asymptotic equivalence between the t-step ET estimator and the regression estimator. Unlike the regression estimator, the weights of the ET estimator are always positive. For sufficiently large t, the t-step ET estimator satisfies the calibration constraint (2). Deville and Särndal (1992) proved the result (23) for the special case of $t \to \infty$.

Remark 1. *The one-step ET estimator, defined by $\hat{Y}_{\mathrm{ET}(1)}$, has a closed-form tilting parameter*

$$\hat{\boldsymbol{\lambda}}_{1(1)} = \left\{\sum_{i\in A} d_i(\mathbf{x}_i - \bar{\mathbf{X}}_d)^{\otimes 2} \Big/ \hat{N}_d\right\}^{-1}(\tilde{\mathbf{X}}_N - \bar{\mathbf{X}}_d), \tag{24}$$

where $\tilde{\mathbf{X}}_N = \mathbf{X}/\hat{N}$ and $\bar{\mathbf{X}}_d = \sum_{i\in A} d_i\mathbf{x}_i / \sum_{i\in A} d_i$. By Theorem 2, *the one-step ET estimator is asymptotically equivalent to the regression estimator, but the calibration constraint* (2) *is not necessarily satisfied. Using Theorem* 2 *applied to $\mathbf{x}_i$ instead of y_i, the one-step ET estimator can be shown to satisfy the approximate calibration constraint described in* (7).

Remark 2. *The ET estimator can also be derived by finding the weights that minimize*

$$Q(w) = \sum_{i\in A} w_i \ln\left(\frac{w_i}{d_i}\right) \tag{25}$$

subject to constraints (2) *and* (4). *The objective function* (25) *is often called the minimum discrimination function. The minimum value of $Q(w)$ is zero if* (4) *is the only calibration constraint and is monotonically increasing if additional calibration constraints are imposed.*

3. Instrumental-variable calibration

We consider some extension of the proposed method in Section 2 to a more general class of ET calibration estimator using instrumental-variables. Use of instrumental-variable in the calibration estimation has been discussed in Estevao and Särndal (2000) and Kott (2003) in some limited simulations. Let $\mathbf{z}_i = \mathbf{z}(\mathbf{x}_i)$ be an instrumental-variable derived from $\mathbf{x}_i$, where the function $\mathbf{z}(\cdot)$ is to be determined. The instrumental-variable exponential tilting (IVET) estimator using the instrumental variable $\mathbf{z}_i$ can be defined as

$$\hat{Y}_{\mathrm{IVET}} = \sum_{i\in A} w_i y_i = \sum_{i\in A} d_i \exp(\hat{\lambda}_0 + \hat{\boldsymbol{\lambda}}'_1\mathbf{z}_i)\, y_i, \tag{26}$$

where $\hat{\lambda}_0$ and $\hat{\boldsymbol{\lambda}}_1$ are computed from (2) and (4). Note that the IVET estimator (26) is a class of estimators indexed by $\mathbf{z}_i$. The instrumental-variable approach defined in (26) provides more flexibility in creating the ET estimator. The choice of $\mathbf{z}_i = \mathbf{x}_i$ leads to the standard ET estimator in (16) but some transformation $\mathbf{z}_i = \mathbf{z}(\mathbf{x}_i)$ can make the resulting ET estimator in (26) more attractive in practice. The solution to the calibration equations can be obtained iteratively by

$$\hat{\boldsymbol{\lambda}}_{1(t+1)} = \hat{\boldsymbol{\lambda}}_{1(t)} + \left\{\sum_{i\in A} w_{i(t)}(\mathbf{x}_i - \bar{\mathbf{X}}_{w(t)})(\mathbf{z}_i - \bar{\mathbf{Z}}_{w(t)})'\right\}^{-1} \left(\mathbf{X} - \sum_{i\in A} w_{i(t)}\mathbf{x}_i\right), \tag{27}$$

where $w_{i(t)} = d_i \exp(\hat{\lambda}_{0(t)} + \hat{\boldsymbol{\lambda}}'_{1(t)}\mathbf{z}_i)$ and $\bar{\mathbf{Z}}_{w(t)} = \sum_{i\in A} w_{i(t)}\mathbf{z}_i / \sum_{i\in A} w_{i(t)}$, with equation (20) unchanged and $\hat{\boldsymbol{\lambda}}_{1(0)} = \mathbf{0}$.

The IVET estimator (26) is useful in creating the final weights that have less extreme values. Since the final weight in (26) is a function of $\mathbf{z}_i$, we can make $g_i = w_i/d_i$ bounded by making $\mathbf{z}_i$ bounded. To create bounded $\mathbf{z}_i$, we can use a trimmed version of $\mathbf{x}_i$, noted by $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{ip})$, where

$$z_{ij} = \begin{cases} x_{ij} & \text{if } |x_{ij} - \bar{x}_j| \le C_j S_j \\ \bar{x}_j + C_j S_j & \text{if } x_{ij} > \bar{x}_j + C_j S_j \\ \bar{x}_j - C_j S_j & \text{if } x_{ij} < \bar{x}_j - C_j S_j, \end{cases} \tag{28}$$

$\bar{x}_j = N^{-1}\sum_{i\in A} d_i x_{ij}$, $S_j^2 = N^{-1}\sum_{i\in A} d_i (x_{ij} - \bar{x}_j)^2$, and C_j is a threshold for detecting outliers, for example, $C_j = 3$. Thus, the IVET estimator using the instrumental-variable obtained by trimming $\mathbf{x}_i$ can be used as an alternative approach to weight trimming.

Instead of using the trimmed instrumental variable $\mathbf{z}_i$ in (28), we can consider the following instrumental variable

$$\mathbf{z}_i = \mathbf{x}_i \Phi_i$$

for some symmetric matrix Φ_i such that $\mathbf{z}_i$ is bounded. Some suitable choice of Φ_i can also improve the efficiency of the resulting IVET estimator. To see this, using the same argument from Theorem 2, the instrumental-variable ET estimator (26) using equations (20) and (27) is asymptotically equivalent to

$$\hat{Y}_{\text{IV, reg}} = \tilde{Y}_d + (\mathbf{X} - \tilde{\mathbf{X}}_d)' \hat{\mathbf{B}}_z \tag{29}$$

where

$$(\tilde{\mathbf{X}}_d', \tilde{Y}_d) = \left(\frac{\hat{N}}{\hat{N}_d}\right)(\hat{\mathbf{X}}_d', \hat{Y}_d)$$

and

$$\hat{\mathbf{B}}_z = \left\{\sum_{i\in A} d_i (\mathbf{z}_i - \bar{\mathbf{Z}}_d)(\mathbf{x}_i - \bar{\mathbf{X}}_d)'\right\}^{-1} \sum_{i\in A} d_i (\mathbf{z}_i - \bar{\mathbf{Z}}_d) y_i. \tag{30}$$

The estimator (29) takes the form of a regression estimator and is called the instrumental-variable regression estimator. Thus, under the choice of $\mathbf{z}_i = \Phi_i \mathbf{x}_i$, the instrumental-variable regression estimator can be written as (29) with

$$\hat{\mathbf{B}}_z = \left\{\sum_{i\in A} d_i (\mathbf{x}_i - \bar{\mathbf{X}}_d)\Phi_i (\mathbf{x}_i - \bar{\mathbf{X}}_d)'\right\}^{-1} \sum_{i\in A} d_i (\mathbf{x}_i - \bar{\mathbf{X}}_d)\Phi_i\, y_i$$

and its variance is minimized for $\Phi_i = V_i^{-1}$ where V_i is the model-variance of y_i given $\mathbf{x}_i$ (Fuller 2009). The model-variance is the variance under the working superpopulation model for the regression of y_i on $\mathbf{x}_i$. Thus, instrumental-variable can be used to improve the efficiency of the resulting calibration estimator, in addition to avoid extreme final weights. Furthermore, the optimal instrumental-variable can be trimmed as in (28) to make the final weightsbounded. Further investigation of the optimal choice of Φ is beyond the scope of this paper and will be a topic of future research.

Remark 3. *Deville and Särndal* (1992) *also considered range-restricted calibration weights of the form*

$$w_i = d_i g_i(\hat{\boldsymbol{\lambda}}) = d_i \frac{L(U-1) + U(1-L)\exp(K\hat{\boldsymbol{\lambda}}'\mathbf{x}_i)}{(U-1) + (1-L)\exp(K\hat{\boldsymbol{\lambda}}'\mathbf{x}_i)}, \tag{31}$$

where $K = (U-L)/\{(1-L)(U-1)\}$, *for some* L *and* U *such that* $0 < L < 1 < U$. *If calibration constraints* (2) *and* (4) *are to be satisfied, then we can use* $\hat{\lambda}_0 + \hat{\boldsymbol{\lambda}}_1'\mathbf{x}_i$ *instead of* $\hat{\boldsymbol{\lambda}}'\mathbf{x}_i$ *in* (31). *The resulting calibration estimator is asymptotically equivalent to the regression estimator using the weights in* (5) *while the IVET estimator is asymptotically equivalent to the instrumental-variable regression estimator* (29). *Computation for obtaining* $\hat{\boldsymbol{\lambda}}$ *is somewhat complicated because* $\partial g_i(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}$ *is not easy to evaluate in* (31). *In the IVET estimator, the computation, given by* (27), *is straightforward.*

To compare the proposed weight with existing methods, we consider an artificial example of a simple random sample with size $n = 5$ where $x_k = k$, $k = 1, 2, \ldots, 5$. Calculations are for three population means of x; $\bar{X}_N = 3$, $\bar{X}_N = 4.5$, and $\bar{X}_N = 6$. Table 1 presents the resulting weights for the regression estimator, the empirical likelihood (EL) estimator, the t-step ET estimator (16) with $t = 1$ and $t = 10$, and the t-step instrumental variable exponential tilting (IVET) estimator (26) with $t = 1$ and $t = 10$. For the IVET estimator, the instrumental variable z_i is created by

$$z_i = \begin{cases} 1.5 & \text{if } x_i \le 1.5 \\ x_i & \text{if } x_i \in (1.5,\ 4.5) \\ 4.5 & \text{if } x_i \ge 4.5. \end{cases}$$

The last column of Table 1 presents the estimated mean of X using the respective calibration weights. All the weights are equal to $1/n = 0.2$ for $\bar{X}_N = 3$. The regression estimator is linearly increasing in x_i but has negative weights for the population with $\bar{X}_N = 4.5$ and $\bar{X}_N = 6$. For the population where $\bar{X}_N = 6$, the weights could not be computed for the EL method because $\bar{X}_N$ is outside the range of the sample x_i's. In this extreme case of $\bar{X}_N = 6$, the ET method provides nonnegative weights by sacrificing the calibration constraint and the EL estimator has more extreme weights than the ET estimator or IVET estimator in the sense that the weight for $k = 5$ is the largest among the estimators considered. The weight for the one-step ET estimator is close to that of the regression estimator for large x_i but it is close to that of EL estimator for small x_i. The 10-step ET estimators has better calibration properties in the sense of smaller value of squared error, $(\sum_{k=1}^{5} w_k x_k - \bar{X}_N)^2$, than the one-step ET estimator. The ET estimator and the IVET estimator provide almost the same estimates of $\bar{X}_N$ for both t, but the IVET estimator produces less extreme weights than the ET estimator.

Table 1
An example of calibration weights with a sample of size $n = 5$

Method	$\bar{X}_N$	x_i: 1	2	3	4	5	$\hat{X}_N$
Reg.	3.0	0.200	0.200	0.200	0.200	0.200	3.0
	4.5	-0.100	0.050	0.200	0.035	0.500	4.5
	6.0	-0.400	-0.100	0.200	0.500	0.800	6.0
EL	3.0	0.200	0.200	0.200	0.200	0.200	3.0
	4.5	0.033	0.043	0.063	0.115	0.746	4.5
	6.0	N/A	N/A	N/A	N/A	N/A	N/A
ET $(t = 1)$	3.0	0.200	0.200	0.200	0.200	0.200	3.0
	4.5	0.027	0.057	0.100	0.255	0.540	4.2
	6.0	0.002	0.009	0.039	0.173	0.777	4.7
ET $(t = 10)$	3.0	0.200	0.200	0.200	0.200	0.200	3.0
	4.5	0.009	0.027	0.078	0.227	0.659	4.5
	6.0	0.000	0.000	0.000	0.001	0.999	5.0
IVET $(t = 1)$	3.0	0.200	0.200	0.200	0.200	0.200	3.0
	4.5	0.030	0.047	0.121	0.309	0.493	4.2
	6.0	0.003	0.006	0.041	0.267	0.683	4.6
IVET $(t = 10)$	3.0	0.200	0.200	0.200	0.200	0.200	3.0
	4.5	0.007	0.015	0.066	0.294	0.618	4.5
	6.0	0.000	0.000	0.000	0.087	0.913	4.9

Reg., Regression estimator; EL, empirical likelihood; ET, exponential tilting; IVET, instrumental variable exponential tilting; N/A, Not applicable.

4. Variance estimation

We now discuss variance estimation of the ET calibration estimators of Sections 2 and 3. Because the estimated parameter $(\hat{\lambda}_0, \hat{\boldsymbol{\lambda}}_1')$ in the ET calibration estimator (16) has some sampling variability, variance estimation method should take into account of this sampling variability of these estimated parameters. In this case, variance estimation can be often obtained by a linearization method or by a replication method (Wolter 2007). For the discussion of the linearization method, let the variance of the HT estimator (1) be consistently estimated by

$$\hat{V}(\hat{Y}_d) = \sum_{i \in A} \sum_{j \in A} \Omega_{ij} y_i y_j. \tag{32}$$

The linearization variance estimator for the ET estimator can be obtained by the linearization variance formula for the regression estimator, as in Deville and Särndal (1992), using the asymptotic equivalence between the ET calibration estimator andthe regression estimator, as shown in Theorem 2. Specifically, if the population size N is known, a linearization variance estimator of the IVET estimator in (26) can be written as

$$\hat{V}(\hat{Y}_{\text{IVET}}) = \sum_{i \in A} \sum_{j \in A} \Omega_{ij}\, g_i\, g_j\, \hat{e}_i\, \hat{e}_j \tag{33}$$

where Ω_{ij} are the coefficients of the variance estimator in (32), $g_i = w_i/d_i$ is the weight adjustment factor, and $\hat{e}_i = y_i - \bar{Y}_d - (\mathbf{x}_i - \bar{\mathbf{X}}_d)'\hat{\mathbf{B}}_z$, where $\hat{\mathbf{B}}_z$ is defined in (30). The choice of $\mathbf{z}_i = \mathbf{x}_i$ in (33) gives the linearized variance estimator for the ET estimator in (16). Consistency of the variance estimator (33) can be found in Kim and Park (2010).

For the one-step ET estimator, a replication method can be easily implemented. Let the replication variance estimator be of the form

$$\hat{V}_{\text{rep}} = \sum_{k=1}^{L} c_k (\hat{Y}_d^{(k)} - \hat{Y}_d)^2, \tag{34}$$

where L is the number of replication, c_k is the replication factor associated with replicate k, $\hat{Y}_d^{(k)} = \sum_{i \in A} d_i^{(k)} y_i$, and $d_i^{(k)}$ is the k^{th} replicate of the design weight d_i. For example, the replication variance estimator (34) includes the jackknife and the bootstrap (see Rust and Rao 1996). Assume that the replication variance estimator (34) is a consistent estimator for the variance of $\hat{Y}_d$. The k^{th} replicate of the one-step ET estimator can be computed by

$$\hat{Y}_{\text{ET}(1)}^{(k)} = \sum_{i \in A} d_i^{(k)} \exp(\hat{\lambda}_{0(1)}^{(k)} + \hat{\boldsymbol{\lambda}}_{1(1)}^{(k)\prime} \mathbf{z}_i)\, y_i \tag{35}$$

where

$$\hat{\boldsymbol{\lambda}}_{1(1)}^{(k)} = \left\{ \sum_{i \in A} d_i^{(k)} (\mathbf{x}_i - \bar{\mathbf{X}}_d^{(k)})\, (\mathbf{z}_i - \bar{\mathbf{Z}}_d^{(k)})' \Big/ \hat{N}_d^{(k)} \right\}^{-1} (\mathbf{X}/\hat{N}^{(k)} - \bar{\mathbf{X}}_d^{(k)}),$$

$$\hat{N}^{(k)} = \begin{cases} N & \text{if } \hat{N} = N \\ \hat{N}_d^{(k)} = \sum_{i \in A} d_i^{(k)} & \text{if } \hat{N} = \hat{N}_d, \end{cases}$$

$$(\bar{\mathbf{X}}_d^{(k)}, \bar{\mathbf{Z}}_d^{(k)}) = \frac{\sum_{i \in A} d_i^{(k)} (\mathbf{x}_i, \mathbf{z}_i)}{\sum_{i \in A} d_i^{(k)}},$$

and

$$\exp(\hat{\lambda}_{0(1)}^{(k)}) = \frac{\hat{N}}{\sum_{i \in A} d_i^{(k)} \exp(\mathbf{z}'\hat{\boldsymbol{\lambda}}_{1(1)}^{(k)})}.$$

The replication variance estimator defined by

$$\hat{V}_{\text{rep}} = \sum_{k=1}^{L} c_k (\hat{Y}_{\text{ET}}^{(k)} - \hat{Y}_{\text{ET}})^2, \quad (36)$$

where $\hat{Y}_{\text{ET}}^{(k)}$ is defined in (35), can be used to estimate the variance of the ET calibration estimator in (26).

5. Simulation study

To study the finite sample performance of the proposed estimators, we performed a limited simulation study. In the simulation, two finite populations of size $N = 10{,}000$ were independently generated. In population A, the finite population is generated from an infinite population specified by $x_i \sim \exp(1) + 1$; $y_i = 3 + x_i + x_i e_i$, $e_i \mid x_i \sim N(0, 1)$; $z_i \mid (x_i, y_i) \sim \chi^2(1) + |y_i|$. In population B, (x_i, e_i, z_i) are the same as in population A but $y_i = (5 - 1/\sqrt{8}) + 1/\sqrt{8}(x_i - 2)^2 + e_i$. The auxiliary variable, x_i, is used for calibration and z_i is the measure of size used for unequal probability sampling. From both of the finite populations generated, $M = 10{,}000$ Monte Carlo samples of size n were independently generated under two sampling schemes described below. The parameter of interest is the population mean of y and we assume that the population size N is known.

The simulation setup can be described as a $2 \times 2 \times 8 \times 2$ factorial design with four factors. The factors are (a) two types of finite populations, (b) Sampling mechanism: simple random sampling and probability proportional to size (z_i) sampling with replacement, (c) Calibration method: no calibration, the regression estimator, the EL method in (6) with $t = 1$ and $t = 10$, the t-step ET method in (21) with $t = 1$ and $t = 10$, and the IVET method (26) with $t = 1$ and $t = 10$, (d) sample size: $n = 100$ and $n = 200$. Since N is assumed to be known, the calibration estimators are computed to satisfy $\sum_{i=1}^{n} w_i (1, x_i) = (1, \bar{X}_N)$ in both populations. For the IVET method (26), the instrumental variable z_i is created using the definitions in (28) with threshold $C = 3$.

Using the Monte Carlo samples generated as above, the biases and the mean squared errors of the eight estimators of the population mean of y, the variable of interest, were computed and are presented in Table 2. The calibration estimators are biased but the bias is small if the regression model holds or the sample size is large. In population A, the linear regression model holds and the regression estimator is efficient in terms of mean squared errors. However, the regression estimator is not efficient in population B because the model used for the regression estimator is not a good fit. The seven calibration estimators show similar performances for the larger sample size. The 10-step IVET estimator performs as well as the regression estimator in population A, and it shows slightly better performance than the other six calibration estimators. In population B, the 10-step IVET estimator performs the best among the calibration estimators considered.

In addition to point estimation, variance estimation was also considered. We considered only the variance estimation for the t-step ET estimators and IVET estimators. The linearization variance estimator in (33) and the replication variance estimator in (36) were computed for each estimator in each sample. In the replication method, the jackknife method was used by deleting one element for each replication. The relative biases of the variance estimators were computed by dividing the Monte Carlo bias of the variance estimator by the Monte Carlo variance. The Monte Carlo relative biases of the linearization variance estimators and the replication variance estimators are presented in Table 3. The theoretical relative bias of the variance estimators is of order $o(1)$, which is consistent with the simulation results in Table 3. The linearization variance estimator slightly underestimates the true variance because it ignores the second order term in the Taylor linearization. The replication variance estimator shows slight positive bias in the simulation. The biases of the variance estimators are generally smaller in absolute values in population A because the linear model holds. In population B, variance estimators for the IVET estimator are less biased than those for the ET estimator because of less extreme weights used by the IVET estimator.

6. Concluding remarks

We have considered the problem of estimating Y with auxiliary information of the form $E\{U(\mathbf{X})\} = 0$ with some known function $U(\cdot)$. The class of the linear estimators of the form $\hat{Y} = \sum_{i \in A} w_i y_i$ with $\sum_{i \in A} w_i \{1, U(\mathbf{x}_i)\} = (\hat{N}, 0)$ and $w_i > 0$ is considered. If the density $f(\mathbf{x}; \boldsymbol{\eta})$ of X is known up to $\boldsymbol{\eta} \in \Omega$, then an efficient estimation can be implemented using the estimated importance weight

$$w_i \propto d_i \frac{f(x_i;\ \boldsymbol{\eta}_{0,N})}{f(x_i;\ \hat{\boldsymbol{\eta}})},$$

where d_i are the initial weights and where $\boldsymbol{\eta}_{0,N}$ and $\hat{\boldsymbol{\eta}}$ are the maximum likelihood estimators of $\boldsymbol{\eta}$ based on the population and the sample, respectively. If the parametric form of $f(\mathbf{x}; \boldsymbol{\eta})$ is unknown, thenthe exponential tilting weights of the form

$$w_{i(\lambda)} \propto \exp\{\boldsymbol{\lambda}' U(\mathbf{x}_i)\}$$

can be used, where $\boldsymbol{\lambda}$ is determined to satisfy

$$\sum_{i \in A} w_{i(\lambda)} U(\mathbf{x}_i) = 0. \quad (37)$$

Table 2
Monte Carlo Biases and Monte Carlo Mean squared errors of the point estimators for the mean of y, based on 10,000 Monte Carlo samples

Population	Sample Size	Estimator	SRS		PPS	
			Bias	MSE	Bias	MSE
A	100	No Calibration	0.00	0.02398	0.00	0.02023
		Regression estimator	0.00	0.01261	0.00	0.01289
		EL estimator ($t=1$)	0.01	0.01369	0.01	0.01353
		EL estimator ($t=10$)	0.00	0.01285	0.00	0.01289
		ET estimator ($t=1$)	0.01	0.01334	0.01	0.01353
		ET estimator ($t=10$)	0.00	0.01269	0.00	0.01289
		IVET estimator ($t=1$)	0.01	0.01309	0.01	0.01330
		IVET estimator ($t=10$)	0.00	0.01263	0.00	0.01289
	200	No Calibration	0.00	0.01069	0.00	0.00925
		Regression estimator	0.00	0.00595	0.00	0.00568
		EL estimator ($t=1$)	0.01	0.00632	0.01	0.00604
		EL estimator ($t=10$)	0.00	0.00597	0.00	0.00568
		ET estimator ($t=1$)	0.00	0.00616	0.01	0.00578
		ET estimator ($t=10$)	0.00	0.00596	0.00	0.00568
		IVET estimator ($t=1$)	0.00	0.00605	0.01	0.00574
		IVET estimator ($t=10$)	0.00	0.00591	0.00	0.00567
B	100	No Calibration	0.00	0.02044	0.00	0.01692
		Regression estimator	-0.01	0.01473	0.00	0.01461
		EL estimator ($t=1$)	0.01	0.01652	0.01	0.01516
		EL estimator ($t=10$)	0.00	0.01490	0.01	0.01472
		ET estimator ($t=1$)	0.00	0.01516	0.01	0.01483
		ET estimator ($t=10$)	0.00	0.01470	0.00	0.01459
		IVET estimator ($t=1$)	0.00	0.01497	0.00	0.01458
		IVET estimator ($t=10$)	0.00	0.01472	0.00	0.01453
	200	No Calibration	0.00	0.00888	0.00	0.00823
		Regression estimator	-0.01	0.00705	0.00	0.00735
		EL estimator ($t=1$)	0.01	0.00769	0.01	0.00764
		EL estimator ($t=10$)	0.00	0.00715	0.01	0.00745
		ET estimator ($t=1$)	0.00	0.00723	0.01	0.00749
		ET estimator ($t=10$)	0.00	0.00706	0.01	0.00734
		IVET estimator ($t=1$)	0.00	0.00704	0.00	0.00728
		IVET estimator ($t=10$)	0.00	0.00699	0.00	0.00725

SRS, simple random sampling; PPS, probability proportional to size sampling; MSE, mean squared error; EL, empirical likelihood; ET, exponential tilting; IVET, instrumental-variable exponential tilting.

Table 3
Monte Carlo Relative Biases of the variance estimators, based on 10,000 Monte Carlo samples

Population	Sample size	Estimator	Linearization		Replication	
			SRS	PPS	SRS	PPS
A	100	ET ($t=1$)	-7.02	-2.66	10.65	4.11
		ET ($t=10$)	-4.91	-0.80	5.60	0.67
		IVET ($t=1$)	-5.28	-3.63	7.67	2.25
		IVET ($t=10$)	-4.11	-0.87	4.96	0.41
	200	ET ($t=1$)	-3.97	-0.19	3.65	0.57
		ET ($t=10$)	-2.93	0.87	2.23	-0.35
		IVET ($t=1$)	-3.35	-0.10	2.34	0.02
		IVET ($t=10$)	-2.72	0.78	1.62	-0.53
B	100	ET ($t=1$)	-7.64	-3.01	10.72	4.50
		ET ($t=10$)	-5.98	-0.98	7.21	0.74
		IVET ($t=1$)	-5.77	-2.31	4.53	-0.10
		IVET ($t=10$)	-5.44	-1.86	5.17	-0.51
	200	ET ($t=1$)	-2.41	-1.01	5.76	2.53
		ET ($t=10$)	-1.29	0.18	4.30	1.91
		IVET ($t=1$)	-1.39	-0.35	2.09	1.04
		IVET ($t=10$)	-1.15	-0.06	2.04	0.99

SRS, simple random sampling; PPS, probability proportional to size sampling; ET, exponential tilting; IVET, instrumental-variable exponential tilting.

If a solution to (37) exists, it can be expressed as the limit of the form

$$w_{i(t)} \propto \prod_{s=0}^{t-1} \exp\{-\hat{U}'_{(s)}\hat{\Sigma}^{-1}_{aa(s)}U(\mathbf{x}_i)\} \qquad (38)$$

where $\hat{U}_{(s)} = \sum_{i\in A} w_{i(s)}U(\mathbf{x}_i)$, $\hat{\Sigma}_{aa(t)} = \sum_{i\in A} w_{i(t)}\{U(\mathbf{x}_i) - \bar{U}_{(t)}\}^{\otimes 2}$, $\bar{U}_{(t)} = \sum_{i\in A} w_{i(t)}U(\mathbf{x}_i)/\sum_{i\in A} w_{i(t)}$ with the initial weight $w_{i(0)} = d_i(\hat{N}/\hat{N}_d)$. If the solution to condition (37) does not exist, we can still use the weights in (38), but the equality must be relaxed. Instead, approximate equality will be satisfied in (37) in the sense that $\sum_{i\in A} w_{i(t)}U(\mathbf{x}_i)$ converges to zero much faster than $\sum_{i\in A} w_{i(0)}U(\mathbf{x}_i)$ for $t \geq 1$. Approximate equality in (37) is called the approximate calibration condition.

The estimators $\hat{Y}_{(t)} = \sum_{i\in A} w_{i(t)}y_i$ that use the t-step ET weights in (38), including the one-step estimator $\hat{Y}_{(1)}$, are asymptotically equivalent to the regression estimator of the form

$$\hat{Y}_{\text{reg}} = \hat{Y}_{(0)} - \hat{U}'_{(0)}\hat{\Sigma}^{-1}_{aa(0)}\hat{\Sigma}_{ay(0)},$$

where $\hat{Y}_{(0)} = \sum_{i\in A} w_{i(0)}y_i$ and $\hat{\Sigma}_{ay(0)} = \sum_{i\in A} w_{i(0)}\{U(\mathbf{x}_i) - \bar{U}_{(0)}\}y_i$. Unlike the regression estimator, the weights of the proposed method are always nonnegative. Furthermore, using the instrumental variable technique in Section 3, the weights are bounded above. Suitable choice of the instrumental variable also improves the efficiency of the resulting calibration estimator.

The exponential tilting calibration method is asymptotically equivalent to the empirical likelihood calibration method but it is more attractive computationally in the sense that the partial derivatives are not required in the iterative computation. Because the computation is simple, the variance of the proposed estimator can be easily estimated using a replication method, as discussed in Section 4. Further investigation in this direction, including interval estimation, can be a topic of future research.

Acknowledgements

The author wishes to thank Minsun Kim for computational support and two anonymous referees and the associated editor for very helpful comments that greatly improved the quality of the paper. This research was partially supported by a Cooperative Agreement NRCS 68-3A75-4-122 between the US Department of Agriculture Natural Resources Conservation Service and Iowa State University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the USDA Natural Resources Conservation Service.

Appendix

A. Assumptions and proof of Theorem 1

We first assume the following regularity conditions:

[A-1] The density $f(\mathbf{x}; \boldsymbol{\eta})$ is twice differentiable with respect to $\boldsymbol{\eta}$ for every $\mathbf{x}$ and satisfy

$$\left|\frac{\partial^2 f(\mathbf{x}; \boldsymbol{\eta})}{\partial \eta_i \partial \eta'_j}\right| \leq K(\mathbf{x})$$

for function $K(\mathbf{x})$ such that $E\{K(\mathbf{x})\} < \infty$, in a neighborhood of $\boldsymbol{\eta}_{0,N}$.

[A-2] The pseudo maximum likelihood estimator $\hat{\boldsymbol{\eta}}$ satisfies $\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_{0,N}) = O_p(1)$.

[A-3] The matrix $E\{\mathbf{s}(\boldsymbol{\eta}_{0,N})^{\otimes 2}\}$ exists and is nonsingular, where $\mathbf{s}(\boldsymbol{\eta}_{0,N}) = \partial \ln f(\mathbf{x}_i; \boldsymbol{\eta})/\partial \boldsymbol{\eta}\,|_{\eta=\eta_{0,N}}$.

To prove Theorem 1, write

$$g_i(\boldsymbol{\eta}) = \frac{f(\mathbf{x}_i; \boldsymbol{\eta}_{0,N})}{f(\mathbf{x}_i; \boldsymbol{\eta})},$$

and $w_i(\boldsymbol{\eta}) = d_i g_i(\boldsymbol{\eta})$. The estimated importance weight in (8) can be written $w_i = w_i(\hat{\boldsymbol{\eta}})$. Taking a Taylor expansion of $N^{-1}\sum_{i\in A} d_i\mathbf{s}_i(\hat{\boldsymbol{\eta}}) = 0$ around $\boldsymbol{\eta}_{0,N}$ leads to

$$\mathbf{0} = \frac{1}{N}\sum_{i\in A} d_i\mathbf{s}_i(\boldsymbol{\eta}_{0,N}) + \left\{\frac{\partial}{\partial \boldsymbol{\eta}'}\frac{1}{N}\sum_{i\in A} d_i\mathbf{s}_i(\boldsymbol{\eta}_{0,N})\right\}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_{0,N}) + o_p(|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_{0,N}|).$$

Note that the first term on the right side of

$$\frac{1}{N}\frac{\partial}{\partial \eta}\sum_{i\in A} d_i\mathbf{s}_i(\boldsymbol{\eta}) = \frac{1}{N}\sum_{i\in A} d_i\frac{\partial^2 f(\mathbf{x}_i; \boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}{f(\mathbf{x}_i; \boldsymbol{\eta})} - \frac{1}{N}\sum_{i\in A} d_i\left\{\frac{\partial f(\mathbf{x}_i; \boldsymbol{\eta})/\partial\boldsymbol{\eta}}{f(\mathbf{x}_i; \boldsymbol{\eta})}\right\}^{\otimes 2}. \qquad (A1)$$

converges to $\int\{\partial^2 f(\mathbf{x}; \boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'\}d\mathbf{x}$ which equals to zero by the dominated convergence theorem with [A1]. The second term converges to $E\{\mathbf{s}(\boldsymbol{\eta}_{0,N})^{\otimes 2}\}$. Thus, by [A-2],

$$\bar{\mathbf{S}}_{0d} \equiv \frac{1}{N}\sum_{i\in A} d_i\,\mathbf{s}_i(\boldsymbol{\eta}_{0,N}) = O_p(n^{-1/2}) \qquad (A2)$$

and

$$\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_{0N} = \hat{\boldsymbol{\Sigma}}^{-1}_{ss}\,\bar{\mathbf{S}}_{0d} + o_p(n^{-1/2}). \qquad (A3)$$

Now, taking a Taylor expansion of $N^{-1}\hat{Y}_w = N^{-1}\sum_{i\in A} w_i(\hat{\boldsymbol{\eta}})y_i$ around $\boldsymbol{\eta} = \boldsymbol{\eta}_{0,N}$ leads to

$$\frac{\hat{Y}_w}{N} = \frac{\hat{Y}_d}{N} + \left\{\frac{\partial}{\partial \boldsymbol{\eta}}\frac{1}{N}\sum_{i\in A} w_i(\boldsymbol{\eta}_{0,N})y_i\right\}'(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_{0,N}) + o_p(|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_{0,N}|) \quad (A4)$$

by the uniform continuity of $\partial\{\sum_{i\in A} w_i(\boldsymbol{\eta})y_i\}/\partial\boldsymbol{\eta}$ around $\boldsymbol{\eta}_{0,N}$. Now, using

$$\frac{\partial}{\partial \boldsymbol{\eta}} g_i(\boldsymbol{\eta}) = -\frac{f(\mathbf{x}_i;\boldsymbol{\eta})}{f(\mathbf{x}_i;\boldsymbol{\eta})} \times \frac{\partial f(\mathbf{x}_i;\boldsymbol{\eta})/\partial\boldsymbol{\eta}}{f(\mathbf{x}_i;\boldsymbol{\eta})} = -g_i(\boldsymbol{\eta}) \times s_i(\boldsymbol{\eta}),$$

where $\mathbf{s}_i(\boldsymbol{\eta}) = \partial \ln f(\mathbf{x}_i;\boldsymbol{\eta})/\partial\boldsymbol{\eta}$, we have

$$\frac{\partial}{\partial \boldsymbol{\eta}}\sum_{i\in A} w_i(\boldsymbol{\eta})y_i = -\sum_{i\in A} w_i(\boldsymbol{\eta})\mathbf{s}_i(\boldsymbol{\eta})y_i.$$

Using $w_i(\boldsymbol{\eta}_{0,N}) = d_i$ and writing $\mathbf{s}_i(\boldsymbol{\eta}_{0,N}) = \mathbf{s}_{i0}$, we have, by (A2),

$$\frac{\partial}{\partial \boldsymbol{\eta}}\frac{1}{N}\sum_{i\in A} w_i(\boldsymbol{\eta}_{0,N})y_i = -\frac{1}{N}\sum_{i\in A} d_i\mathbf{s}_{i0}y_i = -\hat{\boldsymbol{\Sigma}}_{sy} + O_p(n^{-1/2}). \quad (A5)$$

Using (A5) and (A3) in (A4), result (9) is obtained.

B. Proof of Theorem 2

Write

$$\hat{\theta}(\boldsymbol{\lambda}_1) = \frac{\sum_{i\in A} d_i m_i(\boldsymbol{\lambda}_1)y_i}{\sum_{i\in A} d_i m_i(\boldsymbol{\lambda}_1)},$$

where $m_i(\boldsymbol{\lambda}_1) = \exp(\boldsymbol{\lambda}_1'\mathbf{x}_i)$. Note that $\hat{Y}_{\text{ET}(t)} = \hat{N}\hat{\theta}(\hat{\boldsymbol{\lambda}}_{1(t)})$ and $\hat{\boldsymbol{\lambda}}_{1(t)}$ is defined in (19). By a Taylor expansion of $\hat{\theta}(\hat{\boldsymbol{\lambda}}_{1(t)}) = \hat{N}^{-1}\hat{Y}_{\text{ET}(t)}$ around $\boldsymbol{\lambda}_1 = \mathbf{0}$ and by the continuity of the partial derivatives of $\hat{\theta}(\boldsymbol{\lambda}_1)$, we have

$$\hat{\theta}(\hat{\boldsymbol{\lambda}}_{1(t)}) = \hat{\theta}(\mathbf{0}) + \dot{\theta}(\mathbf{0})'(\hat{\boldsymbol{\lambda}}_{1(t)} - \mathbf{0}) + o_p(|\hat{\boldsymbol{\lambda}}_{1(t)} - \mathbf{0}|), \quad (B1)$$

where $\dot{\theta}(\boldsymbol{\lambda}) = \partial\hat{\theta}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}$. Because $\hat{\boldsymbol{\lambda}}_{1(t)}$ converges in quadratic order and the one-step estimator satisfies $\hat{\boldsymbol{\lambda}}_{1(1)} = O_p(n^{-1/2})$, equation (22) can be written as

$$\hat{\boldsymbol{\lambda}}_{1(t)} = \left\{\hat{N}_d^{-1}\sum_{i\in A} d_i(\mathbf{x}_i - \bar{\mathbf{X}}_d)^{\otimes 2}\right\}^{-1}(\hat{N}^{-1}\mathbf{X} - \bar{\mathbf{X}}_d) + o_p(n^{-1/2}). \quad (B2)$$

Note that

$$\dot{\theta}(\boldsymbol{\lambda}_1) = \left\{\sum_{i\in A} d_i m_i(\boldsymbol{\lambda}_1)\right\}^{-1}\sum_{i\in A} d_i\dot{m}_i(\boldsymbol{\lambda}_1)\{y_i - \hat{\theta}(\boldsymbol{\lambda}_1)\}$$

where $\dot{m}_i(\boldsymbol{\lambda}_1) = \partial m_i(\boldsymbol{\lambda}_1)/\partial\boldsymbol{\lambda}_1$. Using $m_i(\mathbf{0}) = 1$ and $\dot{m}_i(\mathbf{0}) = \mathbf{x}_i$, we have $\hat{\theta}(\mathbf{0}) = \hat{Y}_d/\hat{N}_d$ and

$$\dot{\theta}(\mathbf{0}) = \hat{N}_d^{-1}\sum_{i\in A} d_i(\mathbf{x}_i - \bar{\mathbf{X}}_d)y_i. \quad (B3)$$

Therefore, inserting (B2) and (B3) into (B1), we have

$$\hat{\theta}(\hat{\boldsymbol{\lambda}}_{1(t)}) = \frac{\hat{Y}_d}{\hat{N}_d} + \left(\frac{\mathbf{X}}{\hat{N}} - \bar{\mathbf{X}}_d\right)'\left\{\sum_{i\in A} d_i(\mathbf{x}_i - \bar{\mathbf{X}}_d)^{\otimes 2}\right\}^{-1}\sum_{i\in A} d_i(\mathbf{x}_i - \bar{\mathbf{X}}_d)y_i + o_p(n^{-1/2}),$$

which proves (23).

References

Beaumont, J.-F., and Bocci, C. (2008). Another look at ridge calibration. *Metron*, LXVI, 5-20.

Breidt, F.J., Claeskens, G. and Opsomer, J.D. (2005). Model-assisted estimation for complex surveys using penalised splines. *Biometrika*, 92, 831-846.

Chambers, R.L. (1996). Robust case-weighting for multipurpose establishment surveys. *Journal of Official Statistics*, 12, 3-32.

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

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

Chen, J., Variyath, A.M. and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, 17, 426-443.

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

Deville, J.-C., Särndal, C.-E. and Sautory, O. (1993). Generalized raking procedure in survey sampling, *Journal of the American Statistical Association*, 88, 1013-1020.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, 9, 139-172.

Estevao, V.M., and Särndal, C.-E. (2000). A functional approach to calibration. *Journal of Official Statistics*, 16, 379-399.

Folsom, R.E. (1991). Exponential and logistic weight adjustment for sampling and nonresponse error reduction. In *Proceedings of the Section on Social Statistics*, American Statistical Association, 197-202.

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

Fuller, W.A. (2009). *Sampling Statistics*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Givens, G.H., and Hoeting, J.A. (2005). *Computational Statistics*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Henmi, M., Yoshida, R. and Eguchi, S. (2007). Importance sampling via the estimated sampler. *Biometrika*, 94, 985-991.

Imbens, G.W. (2002). Generalized method of moments and empirical likelihood. *Journal of Business and Economic Statististics*, 20, 493-506.

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

Kim, J.K. (2009). Calibration estimation using empirical likelihood in survey sampling. *Statistica Sinica*, 19, 145-157.

Kim, J.K., and Park, M. (2010). Calibration estimation in survey sampling. *International Statistical Review*, In press.

Kott, P.S. (2003). A practical use for instrumental-variable calibration. *Journal of Official Statistics*, 19, 265-272.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, 32, 133-142.

Kitamura, Y., and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65, 861-874.

Park, M., and Fuller, W.A. (2009). The mixed model for survey regression estimation. *Journal of Statistical Planning and Inference*, 139, 1320-1331.

Rao, J.N.K., and Singh, A. (1997). A ridge shrinkage method for range restricted weight calibration in survey sampling. In *Proceedings of the Section on Survey Research Methods*, American Statist Association, 57-64.

Rust, K.F., and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5, 283-310.

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33, 99-119.

Särndal, C.-E., Swenson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Tillé, Y. (1998). Estimation in surveys using conditional probabilities: Simple random sampling. *International Statistical Review*, 66, 303-322.

Wolter, K.M. (2007). *Introduction to Variance Estimation*. 2nd Ed. New York: Springer-Verlag.

Wu, C., and Rao, J.N.K. (2006). Pseudo empirical likelihood ratio confidence intervals for complex surveys. *Canadian Journal of Statistics*, 34, 359-375.