Calibration estimation in survey sampling

Jae Kwang Kim    Mingue Park

September 8, 2009

Abstract

Calibration estimation, where the sampling weights are adjusted to make certain estimators match known population totals, is commonly used in survey sampling. The generalized regression estimator is an example of a calibration estimator. Given the functional form of the calibration adjustment term, we establish the asymptotic equivalence between the functional-form calibration estimator and an instrumental variable calibration estimator where the instrumental variable is directly determined from the functional form in the calibration equation. Variance estimation based on linearization is discussed and applied to some recently proposed calibration estimators. The results are extended to the estimator that is a solution to the calibrated estimating equation. Results from a limited simulation study are presented.

Key Words: benchmarking estimator, domain estimation, generalized regression estimator, instrumental variable regression estimator, variance estimation.
1 INTRODUCTION

Consider the problem of estimating the population total $Y = \sum_{i=1}^{N} y_i$ for a finite population of size $N$. Let $A$ denote the index set of the sample obtained by a probability sampling scheme and let $y$ be observed in the sample. The Horvitz-Thompson (HT) estimator of the form

$$\hat{Y}_d = \sum_{i \in A} d_i y_i$$

is unbiased for $Y$, where $d_i = 1/\pi_i$ is the inverse of the first order inclusion probability of unit $i$ in the population. The weight $d_i$ is often called the design weight since it is directly obtained from the sampling design.

If, in addition to $y_i$, an auxiliary variable vector $x_i$ is available from the sample and the population total $X = \sum_{i=1}^{N} x_i$ is known, it is possible that

$$\sum_{i \in A} d_i x_i \neq X.$$

The class of calibration estimators, calibrated to $X$, is the class of the estimators of the form

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

where $w_i$ satisfies

$$\sum_{i \in A} w_i x_i = X.$$

Thus, we allow the final weight $w_i$ to be a function of $x_i$ but not of $y_i$.

Note that estimators of type (1) satisfying (2) form a class of estimators. To define a unique estimator, Deville and Särndal (1992) used a distance function $\sum_{i \in A} Q(d_i, w_i)$ in an optimization problem under the constraint (2). In this paper, unlike Deville and Särndal, we directly study the asymptotic
properties of the class of calibration estimators (1) satisfying (2). In particular, we find conditions for design consistency of the calibration estimator. The results in this paper are applicable to calibration estimators obtained from the functional form approach. In the functional form approach, only the functional form of the calibration equation is used. We establish the asymptotic equivalence of the functional-form calibration estimator and an instrumental-variable calibration estimator. Instrumental-variable (IV) calibration estimators form another class of calibration estimator, indexed by an instrumental variable $z_i$, of the form

$$\hat{Y}_{IV} = \hat{Y}_d + \left(X - \hat{X}_d\right)' \hat{B}_z$$

(3)

where $\hat{B}_z = \left(\sum_{i \in A} d_i z_i x_i'\right)^{-1} \sum_{i \in A} d_i z_i y_i$. The IV calibration estimator has been discussed in Estevao and Särndal (2000, 2006) and Kott (2003).

Variance estimation for the calibration estimator is an important problem in survey sampling. Fuller (1975) proposed a variance estimator for the regression estimator in a simple setup. Särndal et al (1989) proposed the weighted residual technique for variance estimation. Singh and Folsom (2000) adopted the estimating function framework of Binder (1983) for variance estimation and applied the technique to variance estimation for calibration estimators. The variance estimation method proposed in this paper is more generally applicable than that of Deville and Särndal (1992) in many situations.

In Section 2, the literature on calibration estimation in survey sampling is reviewed. In Section 3, some asymptotic properties of the calibration estimators are established and variance estimation is discussed. In Section 4, the proposed variance estimation method is applied to some recently proposed
calibration methods. In Section 5, the estimator that is a solution to a set of calibrated estimating equations is discussed. In Section 6, results from a limited simulation study are presented.

2 Review of literature

One well-known type of calibration estimation is regression estimation. In Cochran (1942), the regression estimator was proposed under a regression superpopulation model that postulates a relationship between the study variable $y$ and a single auxiliary variable $x$. Mickey (1959) used the idea of scale and location invariance to define a class of regression estimators. Cassel et al (1976) considered the difference estimator for unequal probability samples and suggested the term generalized regression (GREG) estimator when the regression coefficients are estimated. Särndal et al (1989) defined the so called $g$-weight, as the multiplier for the inverse of the inclusion probability that gives the GREG estimator.

Royall (1970, 1976) adapted a linear prediction theory to finite population estimation and suggested the regression estimator as the best linear unbiased predictor for a finite population total. Isaki and Fuller (1982) discussed the optimality of the regression estimator in the class of design consistent predictors under the regression superpopulation model. Wright (1983) proposed a class of predictors, called QR-predictors, that contains most regression estimators and gave sufficient conditions for the QR-predictors to be asymptotically design unbiased. Montanari (1987) reviewed the result of Fuller and Isaki (1981) and Wright (1983) and provided the coefficient for the QR-predictor that minimizes the design variance. Rao (1994) also discussed
the optimal regression estimator that minimize the design variance.

The regression estimator can be extended to more general models. Firth and Bennett (1998) used non-linear regression models in the calibration estimation. Wu and Sitter (2001) suggested the model-calibration estimator that uses an explicit working model for \( E(y_i \mid x_i) \). Instead of using an explicit parametric model, Breidt and Opsomer (2000) adopted a local polynomial regression model to derive a nonparametric regression estimator. Montanari and Ranalli (2005) used the nonparametric neural network model in calibration estimation. Park and Fuller (2009) considered the regression superpopulation model with random components.

The term *calibration estimation* was introduced by Deville and Särndal (1992) as a procedure of minimizing a distance measure between initial weights and final weights subject to calibration equations. Estevao and Särndal (2000) removed the requirement of minimizing a distance measure in calibration estimation and considered the functional form of the calibration weights,

\[ w_i = d_i (1 + \lambda' z_i) \quad (4) \]

for some instrumental vector \( z_i \), where \( \lambda \) is determined from (2). By the fact that the functional form calibration estimator using (4) can be expressed as the instrumental variable calibration estimator (3), the functional form approach using (4) was later termed as *instrument vector approach* by Estevao and Särndal (2006). Kott (2006) defined calibration weights to satisfy the calibration equations and to give a design consistent estimator. Kott (2006) used such a definition to permit a nonlinear type of calibration weights that depend on the nuisance parameters in modeling the unit nonresponse mech-
anism.

Calibration weights defined by minimizing a distance measure under calibration equations can be very large or negative. If the weights are to be used to estimate the population total, it seems reasonable that no individual weight should be less than one. Huang and Fuller (1978) first considered a procedure that prevents both negative and also large weights. Théberge (2000) investigated the conditions for the existence of a solution to the calibration equations with weights within a given interval. Another way of obtaining nonextreme calibration weights is to relax the some or all calibration equations. Bardsley and Chambers (1984) and Chambers (1996) considered the ridge type regression estimator in which the calibration equation is added to the objective function minimized with a certain coefficient matrix. Rao and Singh (1997) proposed a method of ridge shrinkage which is an iterative method of adjusting weight to meet a range restriction and to satisfy the calibration equation within given tolerances. Singh and Mohl (1996) compared several nonnegative regression type estimators through numerical examples. Using the idea of the conditional inclusion probabilities introduced by Tillé (1998), Park and Fuller (2005) introduced a set of regression weights that are positive in most samples. A more comprehensive overview of the calibration estimator can be found in Fuller (2002) and Särndal (2007).

3 Main result

To study the properties of the calibration estimator, assume that the final weight \( w_i \) of the (linear) calibration estimator can be expressed as

\[
w_i = d_i g_i(\hat{\lambda})
\]
for some known function \( g_i(\lambda) = g(x_i; \lambda) \) of a vector \( \lambda \), where \( \hat{\lambda} \) is uniquely determined from
\[
\sum_{i \in A} d_i g_i(\lambda) x_i = X. \tag{5}
\]
In such a case, the calibration estimator can be written
\[
\hat{Y}_w = \hat{Y}_w(\hat{\lambda}) = \sum_{i \in A} d_i g_i(\hat{\lambda}) y_i. \tag{6}
\]
The parameter \( \lambda \) is called a nuisance parameter in the sense that we are not directly interested in \( \lambda \) but the information associated with \( \hat{\lambda} \) can improve the estimation of \( Y \). For example, the GREG estimator described in Särndal et al (1992) uses \( g(x_i; \lambda) = 1 + x_i' \lambda / c_i \), where \( \hat{\lambda} \) is determined by the calibration constraints and \( c_i \) is a positive constant. Let \( \lambda_0 \) be the unique solution to the population analogue of (5):
\[
\sum_{i=1}^N g_i(\lambda) x_i = X. \tag{7}
\]

To discuss the asymptotic properties of the estimators, assume a sequence of finite populations and samples as in Isaki and Fuller (1982) and assume:

[C.1] The HT estimator is \( \sqrt{n} \)-consistent:
\[
N^{-1} \sum_{i \in A} d_i (x'_i, y_i) - N^{-1} \sum_{i=1}^N (x'_i, y_i) = O_p \left(n^{-1/2}\right). \tag{8}
\]

[C.2] For each \( i \), \( g_i(\lambda) \) is a continuous function of \( \lambda \) in a closed interval \( B \) containing \( \lambda_0 \) as an interior point, where \( \lambda_0 \) satisfies (7). Also,
\[
N^{-1} \sum_{i \in A} d_i g_i(\lambda) (x'_i, y_i) - N^{-1} \sum_{i=1}^N g_i(\lambda) (x'_i, y_i) = o_p(1) \tag{9}\]
holds uniformly in \( \lambda \in B \).
[C.3] For each $i$, $g_i(\lambda)$ is a differentiable function of $\lambda$. The partial derivatives $h_i(\lambda) = \partial g_i(\lambda) / \partial \lambda$ are continuous in a closed interval $B$ containing $\lambda_0$. Assume that

$$N^{-1} \sum_{i \in A} d_i h_i(\lambda) (x'_i, y_i) = N^{-1} \sum_{i=1}^N h_i(\lambda) (x'_i, y_i) + o_p(1)$$

holds uniformly in $\lambda \in B$ and that $\sum_{i=1}^N h_i(\lambda_0) x'_i$ is nonsingular.

Condition [C.1] is a standard condition for a sequence of finite populations and samples. The convergence in (8) and (9) is uniform convergence. That is, in (8), given $\epsilon > 0$, there exist $n_0 = n_0(\epsilon)$ such that

$$\Pr \left\{ N^{-1} \left| \sum_{i \in A} d_i g_i(\lambda) (x'_i, y_i) - \sum_{i=1}^N g_i(\lambda) (x'_i, y_i) \right| > \epsilon \right\} \leq \epsilon$$

holds for all $n \geq n_0$ and for all $\lambda \in B$. A sufficient condition for the uniform convergence in [C.2] is that $g_i(\lambda) < M$ for some constant $M$ for all $i$ and for all $\lambda \in B$.

The following theorem presents some asymptotic properties of the calibration estimator (6) satisfying (5).

**Theorem 1** Assume a sequence of finite populations and samples satisfying [C.1]. Assume the calibration equation (5) has exactly one solution $\hat{\lambda}$ almost everywhere and $g_i(\lambda)$ in (5) satisfies [C.2]-[C.3]. Then, the calibration estimator (6) with (5) satisfies

$$\hat{Y}_w = \bar{Y}_l + o_p(\sqrt{1/n})$$

where

$$\bar{Y}_l = X'B_0 + \sum_{i \in A} d_i g_i(\lambda_0) (y_i - x'_i B_0),$$

(11)
\[ B_0 = \left\{ \sum_{i=1}^{N} h_i(\lambda_0) x_i' \right\}^{-1} \sum_{i=1}^{N} h_i(\lambda_0) y_i; \]

\[ h_i(\lambda) = \partial g_i(\lambda) / \partial \lambda, \text{ and } \lambda_0 \text{ is the solution to } (7). \]

**Proof.** See Appendix A. 

Theorem 1 states that \( \hat{Y}_w \) is asymptotically equivalent to \( \tilde{Y}_l \) in (11) in the sense that (10) holds. If

\[ g_i(\lambda_0) = 1, \quad (12) \]

then \( \tilde{Y}_l = \hat{Y}_d + (X - \hat{X}_d)'B_0 \) and, by Theorem 1,

\[ \hat{Y}_w = \hat{Y}_d + (X - \hat{X}_d)'B_0 + o_p \left( n^{-1/2}N \right), \quad (13) \]

where \( (\hat{X}_d', \hat{Y}_d) = \sum_{i \in A} d_i (x_i', y_i) \). Thus, under the conditions of Theorem 1, the consistency of \( \hat{Y}_w \) follows if condition (12) holds. In the GREG estimator where \( g_i(\lambda) = 1 + x_i' \lambda / c_i \), condition (12) is satisfied with \( \lambda_0 = 0 \).

Under the regularity conditions for \( \hat{B} - B_0 = o_p (1) \), the linearization in (13) can be written

\[ \hat{Y}_w = \hat{Y}_{IV} + o_p \left( n^{-1/2}N \right), \quad (14) \]

where

\[ \hat{Y}_{IV} = \hat{Y}_d + (X - \hat{X}_d)'\hat{B}, \quad (15) \]

and \( h_i(\lambda) = \partial g_i(\lambda) / \partial \lambda \). Note that the estimator (15) takes the form of the instrumental-variable (IV) calibration estimator (3) with the instrumental variable \( z_i = h_i(\lambda) \). Result (14) states that the calibration estimator defined in a functional form is asymptotically equivalent to the IV calibration.
estimator. Furthermore, if
\[
\lim_{\lambda \to \lambda_0} h_i(\lambda) = Kx_i/c_i \tag{16}
\]
holds for some constant \(K\), the IV calibration estimator (15) reduces to the
GREG estimator. Estevao and Särndal (2000) established results similar to (13) for the special case of \(g_i(\lambda) = 1 + \lambda^t z_i\). Théberge (2000) also derived a
linearization of the calibration estimator from an optimization approach.

To discuss variance estimation for the calibration estimators, first assume
that a consistent estimator of the variance of \(\hat{Y}_d\) exists that is a quadratic
function of the sample observations. That is, the variance estimator
\[
\hat{V} = \sum_{i \in A} \sum_{j \in A} \Omega_{ij} y_i y_j \tag{17}
\]
satisfies
\[
\hat{V} / V(\hat{Y}_d) = 1 + o_p(1), \tag{18}
\]
for any \(y\) with bounded fourth moments.

By (10), the calibration estimator is asymptotically equivalent to \(\tilde{Y}_l\) whose
variance is
\[
V(\tilde{Y}_l) = V \left\{ \sum_{i \in A} d_i g_i(\lambda_0) (y_i - x_i^t B_0) \right\}.
\]
To use the variance estimator (17) to estimate the variance of \(\tilde{Y}_l\), we compute
the weighted residual
\[
\hat{g}_i \hat{e}_i = g_i(\hat{\lambda}) \left( y_i - x_i^t \hat{B} \right),
\]
where
\[
\hat{B} = \left\{ \sum_{i \in A} d_i h_i(\hat{\lambda}) x_i^t \right\}^{-1} \sum_{i \in A} d_i h_i(\hat{\lambda}) y_i. \tag{19}
\]
The resulting variance estimator is

\[ \hat{V} = \sum_{i \in A} \sum_{j \in A} \Omega_{ij} \hat{g}_i \hat{g}_j \hat{e}_i \hat{e}_j. \]  

(20)

The variance estimator (20) using (19) was also considered in Estevao and Särndal (2000) in the context of IV calibration estimation. The following theorem provides the consistency of the proposed variance estimator (20).

**Theorem 2** Let the conditions of Theorem 1 hold. Also, assume

\[ nN^{-2} V \left[ \sum_{i \in A} d_i \left( \frac{g_i(\lambda_0)}{h_i(\lambda_0)} \right) (x'_i, y_i) \right] = O(1), \]  

(21)

where \( h_i(\lambda) = \partial g_i(\lambda) / \partial \lambda \). Assume the variance estimator (17) satisfies (18). Then, the plug-in variance estimator (20) satisfies

\[ \hat{V} = V \left( \tilde{Y}_l \right) + o_p \left( n^{-1} N^2 \right), \]

where \( \tilde{Y}_l \) is defined in (10).

**Proof.** See Appendix B. \( \blacksquare \)

Theorem 2 states that the variance estimator (17) is consistent for the variance of \( \tilde{Y}_l \), the linearized version of \( \hat{Y}_w \). The variance estimator is also consistent for the variance of \( \hat{Y}_w \) if

\[ V \left( \hat{Y}_w \right) = V \left( \tilde{Y}_l \right) + o \left( n^{-1} N^2 \right), \]  

(22)

which can be justified under some regularity conditions. A sketched proof of (22) can be obtained from the authors. Note that, usually

\[ V \left[ \sum_{i \in A} d_i g_i(\lambda_0) (x'_i, y_i) \right] = O \left( n^{-1} N^2 \right). \]
Thus, if \( h_i(\lambda_0) \) is uniformly bounded, then (21) will be satisfied. Instead of using (19), Deville and Särndal (1992) proposed a variance estimator of the form (20) with

\[
\hat{B} = \left\{ \sum_{i \in A} d_i \hat{g}_i x_i x_i' \right\}^{-1} \sum_{i \in A} d_i \hat{g}_i x_i y_i. \tag{23}
\]

If \( h_i(\lambda) \) satisfies (16), the term (23) is equivalent to (19). Most of the calibration weights considered in Deville and Särndal (1992) satisfy condition (16).

4 Applications

4.1 Model calibration

Wu and Sitter (2001) considered the following superpopulation model

\[
E_{\zeta} (y_i \mid x_i) = \mu (x_i, \theta)
\]

and proposed the model calibration estimator using the calibration equation

\[
\sum_{i \in A} d_i g_i (\hat{\lambda})(1, \hat{\mu}_i) = \sum_{i=1}^N (1, \hat{\mu}_i), \tag{24}
\]

where \( \hat{\mu}_i = \mu(x_i, \theta) \) and considered a variance estimator for a simple situation.

The variance estimation method proposed in Section 3 can be directly applied to the model calibration estimator, recognizing that there are two types of the nuisance parameters. One is \( \lambda \), which was used to compute \( g_i \), and the other is \( \theta \), which was used to compute \( \hat{\mu}_i \). It is shown in Appendix
C that the model calibration estimator satisfies
\[
\hat{Y}_w = \sum_{i=1}^{N} z_i^\prime B_1 + \sum_{i \in A} d_i g_i (\lambda_0) (y_i - z_i^\prime B_1) + B_1^\prime B_2 \hat{U}_2(\theta) + O_p(n^{-1}N)
\]
(25)

where
\[
B_1 = \left\{ \sum_{i=1}^{N} h_i (\lambda_0) z_i' \right\}^{-1} \sum_{i=1}^{N} h_i (\lambda_0) y_i,
\]
\[
B_2 = \left\{ E \left( \partial \hat{U}_2/\partial \theta \right) \right\}^{-1} E \left( \partial \hat{U}_1/\partial \theta \right),
\]
\[
\hat{U}_1 (\lambda, \theta) = \sum_{i \in A} d_i g_i (\lambda) [1, \mu (x_i, \theta)] - \sum_{i=1}^{N} [1, \mu (x_i, \theta)],
\]
\[
\hat{U}_2 (\theta) \text{ is the estimating equation for the superpopulation parameters and } z_i = (1, \hat{\mu}_i)', \text{ The term } B_1^\prime B_2 \hat{U}_2(\theta) \text{ in (25) represents the effect of estimating } \theta \text{ from } \hat{U}_2(\theta) = 0. \text{ Under the conditions described in Theorem 1 of Wu and Sitter (2001), we have } B_2 = o_p(1) \text{ and the } B_1^\prime B_2 \hat{U}_2(\theta) \text{ term can be safely ignored. We note that } E \left( \partial \hat{U}_1/\partial \theta \right) \text{ in } B_2 \text{ is essentially the bias of the calibration estimator using } y_i = \partial \mu (x_i; \theta)/\partial \theta. \text{ Since the asymptotic bias of the calibration estimator is negligible under (12), we have } B_2 = o_p(1) \text{ and the linearization (25) reduces to }
\]
\[
\hat{Y}_w = \sum_{i=1}^{N} z_i^\prime B_1 + \sum_{i \in A} d_i g_i (\lambda_0) (y_i - z_i^\prime B_1) + O_p(n^{-1}N).
\]

Thus, a consistent variance estimator uses (20) with
\[
\hat{e}_i = y_i - z_i^\prime \hat{B}_1,
\]
where
\[
\hat{B}_1 = \left\{ \sum_{i \in A} d_i h_i (\hat{\lambda}) z_i' \right\}^{-1} \sum_{i \in A} d_i h_i (\hat{\lambda}) y_i.
\]
4.2 Calibration using empirical likelihood

Empirical likelihood, investigated by Owen (1988) and first considered by Hartley and Rao (1968), is a likelihood function derived by assuming that the distribution has support only on the observed sample points. The empirical likelihood calibration estimator is proposed in Chen and Qin (1993) under simple random sampling and is extended to unequal probability sampling by Chen and Sitter (1999) and Kim (2009). The adjustment term given by the method of Chen and Sitter (1999) and Kim (2009) can be written

\[ g(x_i, \lambda) = \frac{1}{\lambda_1 + \lambda_2 u_i} \]

with \( u_i = (x_i - \bar{x}_N) \) and \( u_i = d_i (x_i - \bar{x}_N) \), respectively, where \( \lambda_1 \) and \( \lambda_2 \) are computed from

\[
\sum_{i \in A} d_i g(x_i; \lambda) (1, x_i') = \sum_{i=1}^{N} (1, x_i')
\]

and \( \bar{x}_N = N^{-1} \sum_{i=1}^{N} x_i \).

In either case, we can use the variance estimation formula (20), where

\[ \hat{e}_i = y_i - (1, x_i') \hat{B}, \]

\[ \hat{B} = \left( \sum_{i \in A} d_i g(\hat{\lambda})^2 \tilde{u}_i \tilde{u}_i' \right)^{-1} \left( \sum_{i \in A} d_i g(\hat{\lambda})^2 \tilde{u}_i y_i \right), \tag{26} \]

and \( \tilde{u}_i = (1, u_i')' \). Instead of (26), Chen and Sitter (1999) considered using

\[ \hat{B} = \left( \sum_{i \in A} d_i \tilde{u}_i \tilde{u}_i' \right)^{-1} \left( \sum_{i \in A} d_i \tilde{u}_i y_i \right), \]

which is motivated from the fact that their empirical likelihood calibration estimator is asymptotically equivalent to the generalized regression (GREG) estimator.
4.3 Raking ratio estimation

Deming and Stephan (1940) suggested a raking ratio procedure to estimate the cell frequency when the true marginal distributions in a two-way table are known. Deville et al (1993) obtained the raking ratio weights by minimizing the distance function between the sampling weights and the adjusted weights

\[ w_i \log \left( \frac{w_i}{d_i} \right) - w_i + d_i, \]

under the restriction (2). Thus, the adjustment term obtained from the raking method is

\[ g_i(\lambda) = g(u_i, \lambda) = \exp(\mathbf{u}_i' \lambda), \]

where \( \mathbf{u}_i = (1, \mathbf{x}_i')' \) and \( \lambda \) is the solution of

\[ \sum_{i \in A} d_i \left[ \exp(\mathbf{u}_i' \lambda) \right] \mathbf{u}_i = \sum_{i=1}^{N} \mathbf{u}_i. \]

We can use the variance estimator of (20) where

\[ \hat{e}_i = y_i - \mathbf{u}_i' \widehat{\mathbf{B}}, \]

and

\[ \widehat{\mathbf{B}} = \left( \sum_{i \in A} d_i g_i(\hat{\lambda}) \mathbf{u}_i \mathbf{u}_i' \right)^{-1} \left( \sum_{i \in A} d_i g_i(\hat{\lambda}) \mathbf{u}_i y_i \right). \]

(27)

Because condition (16) holds, \( \widehat{\mathbf{B}} \) of (27) is equivalent to the one used by Deville and Särndal (1992).

4.4 Logistic calibration

We now consider a new calibration method that restricts the range of the weights. The adjustment factor \( g_i(\hat{\lambda}) \) is often constructed to satisfy some
moment restriction (5), or (24). In addition to these moment restriction, range restriction is also important in practice. Huang and Fuller (1978), Park and Fuller (2005) and Chen et al (2002) considered the range restriction in deriving a calibration estimator. Suppose that, in addition to the calibration constraint (5), we also add a range restriction of the form, for all \( i \in A \),

\[ g_i(\hat{\lambda}) \in (0, M). \]  \( (28) \)

Restriction (28) is a range restriction with the upper bound \( M \) on \( g_i(\hat{\lambda}) \).

One way to achieve the two restrictions, (5) and (28), in the calibration estimation is to use the following adjustment factor

\[ g_i(\lambda) = M \frac{\exp(x'_i \lambda)}{M - 1 + \exp(x'_i \lambda)}. \]  \( (29) \)

This adjustment factor is a special case of the range restricted calibration estimator considered in Deville and Särndal (1992, Case 6 in p.378). For the calibration restriction, the parameters \( \lambda \) are computed from

\[ \sum_{i \in A} d_i g_i(\lambda) x_i = \sum_{i=1}^{N} x_i, \]  \( (30) \)

which is a nonlinear equation for \( \lambda \). A Newton-type algorithm can be used to solve the nonlinear equation (30). The calibration estimator using (29) is called the logistic calibration estimator.

For variance estimation, we have only to apply the formula (20) using the residual \( \hat{e}_i = y_i - x'_i \hat{B} \) with \( \hat{B} \) in (19). For the logistic regression weighting method using (29), the residual is computed with

\[ \hat{B} = \left\{ \sum_{i \in A} d_i \hat{g}_i (1 - M^{-1} \hat{g}_i) x_i x'_i \right\}^{-1} \sum_{i \in A} d_i \hat{g}_i (1 - M^{-1} \hat{g}_i) x_i y_i. \]

The variance estimation formula is different from that of Deville and Särndal (1992).
5 Calibration for estimating equations

We now consider the case where the parameter of interest, $\theta_N$, is a solution to an estimating equation of the form

$$W(\theta) \equiv \sum_{i=1}^{N} \omega(x_i, y_i; \theta) = 0$$

for some $\omega(x_i, y_i; \theta)$ that is a continuous differentiable function of $\theta$. Binder (1983) and Binder and Patak (1994) discussed the estimating equation approach to parameter estimation in survey sampling. We assume that the population mean of $x_i$ is known and is used in the calibration to define the final weights. The resulting estimator of $\theta_N$ is obtained by solving

$$\hat{W}_\omega(\theta) \equiv \sum_{i \in A} d_i g_i(\hat{\lambda}) \omega(x_i, y_i; \theta) = 0 \quad (31)$$

where $\hat{\lambda}$ is uniquely determined from (5).

The following theorem presents some asymptotic properties of the estimator that is a solution to (31).

**Theorem 3** Let $\omega(x_i, y_i; \theta)$ be a uniformly continuous function of $\theta$ for each $i$. Under the conditions of Theorem 1, the unique solution $\hat{\theta}_\omega$ to (31) satisfies

$$\sqrt{n} \left( \hat{\theta}_\omega - \tilde{\theta}_i \right) = o_p(1). \quad (32)$$

where $\tilde{\theta}_i$ is the unique solution to $\hat{W}_i(\theta) = 0$ where

$$\hat{W}_i(\theta) \equiv \sum_{i \in A} d_i g_i(\lambda_0) \omega(x_i, y_i; \theta)$$

$$+ \left( X - \sum_{i \in A} d_i g_i(\lambda_0) x_i \right) \left\{ \sum_{i=1}^{N} h_i(\lambda_0) x_i' \right\}^{-1} \sum_{i=1}^{N} h_i(\lambda_0) \omega(x_i, y_i; \theta),$$

and $h_i(\lambda) = \partial g_i(\lambda)/\partial \lambda$ and $\lambda_0$ is the solution to (7).
The proof of Theorem 3 can be found in Appendix D. By (32), under the regularity conditions discussed in Binder (1983),

\[ \hat{\theta}_l - \theta_N = - \left\{ \sum_{i=1}^{N} \frac{\partial \omega (x_i, y_i; \theta_N)}{\partial \theta} \right\}^{-1} \hat{W}_l (\theta_N) + o_p \left( n^{-1/2} \right) \]

and the asymptotic variance is

\[ V (\hat{\theta}_l) = \left\{ \sum_{i=1}^{N} \frac{\partial \omega (x_i, y_i; \theta_N)}{\partial \theta} \right\}^{-2} \left[ \sum_{i \in A} d_i g_i (\lambda_0) \{ \omega_i (\theta_N) - x_i^l B (\lambda_0, \theta_N) \} \right] \]

(33)

where \( \omega_i (\theta) = \omega (x_i, y_i; \theta) \) and

\[ B (\lambda_0, \theta) = \left\{ \sum_{i=1}^{N} h_i (\lambda_0) x_i^l \right\}^{-1} \sum_{i=1}^{N} h_i (\lambda_0) \omega_i (\theta). \]

The variance formula (33) can be easily extended to the vector \( \theta \) case. To estimate the variance (33), a plug-in estimator can be used. The plug-in estimator is

\[ \hat{V} = \left\{ \sum_{i \in A} d_i \hat{g}_i \partial \omega \left( x_i, y_i; \hat{\theta}_\omega \right) / \partial \theta \right\}^{-2} \sum_{i \in A} \sum_{j \in A} \Omega_{ij} \hat{g}_i \hat{g}_j \hat{e}_i \hat{e}_j \]

(34)

where \( \Omega_{ij} \) is defined in (17), \( \hat{g}_i = g_i (\hat{\lambda}), \hat{e}_i = \hat{\omega}_i - x_i^l \hat{B} (\lambda, \hat{\theta}_\omega) \) and \( \hat{u}_i = u \left( x_i, y_i; \hat{\theta}_\omega \right) \) with

\[ \hat{B} (\hat{\lambda}, \hat{\theta}_\omega) = \left\{ \sum_{i \in A} d_i h_i (\hat{\lambda}) x_i^l \right\}^{-1} \sum_{i \in A} d_i h_i (\hat{\lambda}) \hat{\omega}_i. \]

**Example 1** Let \( \theta_D \) be the population mean of \( y \) in domain \( D \), defined by

\[ \theta_D = \left( \sum_{i=1}^{N} \delta_i \right)^{-1} \sum_{i=1}^{N} \delta_i y_i, \]
where $\delta_i = 1$ if $i \in D$ and $\delta_i = 0$ otherwise. Note that $\theta_D$ can be written as a solution to $W(\theta_D) \equiv \sum_{i=1}^{N} \delta_i (y_i - \theta_D) = 0$. The calibration estimator of $\theta_D$ can be written

$$
\hat{\theta}_D = \left( \sum_{i \in A} d_i \hat{g}_i \delta_i \right)^{-1} \sum_{i \in A} d_i \hat{g}_i \delta_i y_i, \tag{35}
$$

where $\hat{g}_i$ satisfies the calibration condition (7). The plug-in variance estimator (34) applied to $\hat{\theta}_D$ in (35) can be written

$$
\hat{V}(\hat{\theta}_D) = \left\{ \sum_{i \in A} d_i \hat{g}_i \delta_i \right\}^{-2} \sum_{i \in A} \sum_{j \in A} \Omega_{ij} \hat{g}_i \hat{g}_j \hat{e}_i \hat{e}_j, \tag{36}
$$

where

$$
\hat{e}_i = \delta_i \left( y_i - \hat{\theta}_D \right) - x_i' \left\{ \sum_{i \in A} d_i \mathbf{h}_i(\hat{\lambda}) x_i \right\}^{-1} \sum_{i \in A} d_i \mathbf{h}_i(\hat{\lambda}) \delta_i \left( y_i - \hat{\theta}_D \right).
$$

6 Simulation Study

To compare the performances of proposed estimators, we performed a limited simulation study. Two types of artificial finite populations of size $N = 10,000$ were generated. In population A, the population values were generated from $X_i \sim \text{exponential}(1)$ and $Y_i \mid X_i = 2 + X_i + e_i$, where $e_i$ are independently generated by the standard normal distribution. In population B, we used the same $X_i$ values in population B but the $Y_i$ values are generated from $Y_i \mid X_i = 2 + \sqrt{X_i} + \sqrt{X_i} \cdot e_i$ and $e_i$ are independently generated by the standard normal distribution.

From the finite populations generated above, $B = 10,000$ Monte Carlo samples of size $n = 100$ and $n = 500$ were independently selected, respectively. We considered two parameters: $\theta_1 = E(y)$ and $\theta_2 = E(y \mid x > 2)$,
where $\theta_1$ is the grand mean and $\theta_2$ is a domain mean. From each Monte Carlo sample, five estimators were computed. The estimators are

1. Horvitz-Thompson (HT) estimator
2. Regression estimator
3. Empirical likelihood estimator
4. Logistic calibration estimator with $M = 3$
5. Logistic calibration estimator with $M = 2$

where $M$ is the upper bound of the ratio of the final weight to the original design weight. The last four estimators are calibration estimators. The control used in the four calibration estimators is the population mean of the $X$ variable. Variance estimators were computed using the weighted residual formula in (20). For variance estimation of domain mean estimators, the variance estimation formula (36) was used.

In Table 1, the Monte Carlo mean squared error of the point estimators are presented. All the estimators are nearly unbiased. The calibration estimators are more efficient than the HT estimator and the variance reduction is about 50\% in population A, which is consistent with the theory because the population correlation between $X_i$ and $Y_i$ is $\sqrt{0.5}$. When $n = 100$, the logistic weighting method shows similar performances to that of the regression estimator for $\theta_1$. The empirical likelihood estimator is less efficient than the regression estimator. When $n = 500$, the four calibration estimators perform similarly. For $\theta_2$, the performance of logistic weighting with $M = 2$ is slightly
better because the domain estimator is more sensitive to the existence of the extreme weights.

< Table 1 around here. >

In Table 2, the Monte Carlo relative biases of the variance estimators are presented. The variance estimators for the grand mean are nearly unbiased when \( n = 500 \), which is consistent with the theory of Section 3 and Section 5. The variance estimators for the domain mean show slight biases. The bias comes from two sources: one is from the use of Taylor linearization in obtaining the variance of the estimating equation and the other is the ratio bias in the denominator of (36). The first source is negligible for a large sample size, but the second source is negligible only for a large domain sample size. To correct the bias of the second type, bias-correction methods using degrees-of-freedom adjustment can be considered, as in Fuller (2009). The relative biases are negligible for \( n = 500 \).

< Table 2 around here. >

**Acknowledgement**

The work of the first author was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2007-313-C00123). The authors wish to thank Emily Berg, Wayne Fuller, the referees and the editor for their very constructive comments.
Appendix

A. Proof of Theorem 1

We first state two lemmas. The first one is about consistency of \( \hat{\lambda} \) for \( \lambda_0 \). The second one is about the \( \sqrt{n} \)-consistency of \( \hat{\lambda} \) for \( \lambda_0 \).

**Lemma 1** Assume that

[A1] \( \hat{U}(\lambda) \) converges in probability to \( U(\lambda) \) uniformly in \( \lambda \in \mathcal{B} \).

[A2] \( \hat{U}(\lambda) = 0 \) has exactly one solution \( \hat{\lambda} \).

[A3] \( U(\lambda) \) is continuous and \( U(\lambda) = 0 \) has a unique solution at \( \lambda_0 \).

Then, \( \hat{\lambda} \) converges in probability to \( \lambda_0 \).

Lemma 1 is stated without proof because it is a special case of Lemma 3 in Appendix D.

**Lemma 2** Let \( \hat{H}(\lambda) = \partial \hat{U}/\partial \lambda \) and \( H(\lambda) = \partial U/\partial \lambda \). Assume that

[B1] \( \sqrt{n}\hat{U}(\lambda_0) \) and \( \sqrt{n}\hat{H}(\lambda_0) \) are bounded in probability.

[B2] \( \hat{U}(\lambda) \) is differentiable with derivative \( \hat{H}(\lambda) \) that is continuous in \( \lambda \) in a closed interval \( \mathcal{B} \) containing \( \lambda_0 \). Also, \( U(\lambda) \) is differentiable with continuous partial derivatives in \( \mathcal{B} \).

[B3] \( \hat{H}(\lambda) \) is also uniformly consistent. That is,

\[
\hat{H}(\lambda) = H(\lambda) + o_p(1)
\]

for all \( \lambda \) uniformly in \( \mathcal{B} \). Also, \( H(\lambda) \) is nonsingular at \( \lambda_0 \).
The (unique) solution \( \hat{\lambda} \) of \( \hat{U}(\lambda) = 0 \) converges to \( \lambda_0 \) which is the unique solution to \( U(\lambda) = 0 \).

Then,

\[
\hat{\lambda} - \lambda_0 = - \{ H(\lambda_0) \}^{-1} \hat{U}(\lambda_0) + o_p(n^{-1/2}).
\] (A.1)

**Proof.** By [B2], we can apply the mean value theorem,

\[
\hat{U}(\hat{\lambda}) - \hat{U}(\lambda_0) = \hat{H}(\lambda^*) (\hat{\lambda} - \lambda_0)
\] (A.2)

where \( \lambda^* \) is a point between \( \hat{\lambda} \) and \( \lambda_0 \). For given \( \epsilon > 0 \),

\[
Pr \left\{ \left| \hat{H}(\lambda^*) - H(\lambda_0) \right| > \epsilon \right\}
\] (A.3)

\[
\leq Pr \left\{ \left| \hat{H}(\lambda^*) - H(\lambda) \right| > \epsilon/2 \right\} + Pr \left\{ \left| H(\lambda^*) - H(\lambda_0) \right| > \epsilon/2 \right\}
\]

\[
\leq Pr \left\{ \sup_{\lambda \in B} \left| \hat{H}(\lambda) - H(\lambda) \right| > \epsilon/2 \right\} + Pr \left\{ \left| H(\lambda^*) - H(\lambda_0) \right| > \epsilon/2 \right\}.
\]

By [B3], we can find \( n_0 = n_0(\epsilon) \) such that

\[
Pr \left\{ \sup_{\lambda \in B} \left| \hat{H}(\lambda) - H(\lambda) \right| > \epsilon/2 \right\} \leq \epsilon/2
\]

holds for all \( n \geq n_0 \). For the second term in (A.3), because \( \hat{\lambda} - \lambda_0 = o_p(1) \) by [B4], we have \( \lambda^* - \lambda_0 = o_p(1) \). By the continuity of \( H(\lambda) \) at \( \lambda_0 \), we can find \( n_1 = n_1(\lambda_0, \epsilon) \) and \( \delta = \delta(\lambda_0, \epsilon) \) such that

\[
Pr \left\{ \left| H(\lambda^*) - H(\lambda_0) \right| > \epsilon/2 \right\} \leq Pr \left\{ |\lambda^* - \lambda| > \delta \right\} \leq \epsilon/2,
\]

holds for \( n \geq n_1 \). Thus, (A.3) term is less than \( \epsilon \) for \( n \geq \max \{ n_0, n_1 \} \) so that

\[
\hat{H}(\lambda^*) = H(\lambda_0) + o_p(1)
\] (A.4)
holds and (A.2) reduces to, by [B4] and \( \hat{U}(\hat{\lambda}) = 0 \),
\[
-\sqrt{n} \hat{U}(\lambda_0) = \sqrt{n} H(\lambda_0) \left( \hat{\lambda} - \lambda_0 \right) + o_p \left( \sqrt{n} \| \hat{\lambda} - \lambda_0 \| \right).
\] (A.5)

By the Cauchy-Schwarz inequality,
\[
\sqrt{n} \| \hat{\lambda} - \lambda_0 \| \leq \| H^{-1}(\lambda_0) \| \| \sqrt{n} H(\lambda_0) \| \hat{\lambda} - \lambda_0 \|
= O_p(1) + o_p \left( \sqrt{n} \| \hat{\lambda} - \lambda_0 \| \right)
\]
which implies \( \sqrt{n} \)-consistency of \( \hat{\lambda} \). Thus, \( o_p \left( \sqrt{n} \| \hat{\lambda} - \lambda_0 \| \right) = o_p(1) \) and (A.5) reduces to
\[
-\sqrt{n} \hat{U}(\lambda_0) = \sqrt{n} H(\lambda_0) \left( \hat{\lambda} - \lambda_0 \right) + o_p(1)
\]
and (A.1) follows.

Now, we use Lemma 3 to prove the theorem. Note that Lemma 1 is used to prove Lemma 2 and Lemma 2 is used to prove Lemma 3. Write
\[
\hat{U}(\lambda) = N^{-1} \left\{ \sum_{i \in A} d_i g_i(\lambda) x_i - \sum_{i=1}^N x_i \right\}
\]
and
\[
U(\lambda) = N^{-1} \left\{ \sum_{i=1}^N g_i(\lambda) x_i - \sum_{i=1}^N x_i \right\}.
\]
Then, conditions [A1]-[A3] and [B1]-[B3] are satisfied and the \( \sqrt{n} \)-consistency of \( \hat{\lambda} \) follows.

Now, by the mean value theorem,
\[
\hat{Y}_w(\hat{\lambda}) = \hat{Y}_w(\lambda_0) + \left[ \sum_{i \in A} d_i h_i(\lambda^*)' y_i \right] \left( \hat{\lambda} - \lambda_0 \right), \] (A.6)
where \( \lambda^* \) is a point between \( \hat{\lambda} \) and \( \lambda_0 \). Using the same argument for (A.4),
\[
N^{-1} \sum_{i \in A} d_i h_i(\lambda^*) y_i = N^{-1} \sum_{i=1}^N h_i(\lambda_0) y_i + o_p(1). \] (A.7)
Thus, inserting (A.7) into (A.6), we have

\[ \hat{Y}_w(\hat{\lambda}) = \hat{Y}_w(\lambda_0) + \left[ \sum_{i=1}^{N} h_i(\lambda_0)' y_i \right] (\hat{\lambda} - \lambda_0) + o_p\left((n^{-1/2}N)\right). \]  

(A.8)

Inserting (A.1) into (A.8), we have (10).

**B. Proof of Theorem 2**

Define

\[ \hat{V}_{yy}(\lambda) = \sum_{i \in A} \sum_{j \in A} \Omega_{ij} g_i(\lambda) g_j(\lambda) y_i y_j \]

\[ \hat{V}_{xy}(\lambda) = \sum_{i \in A} \sum_{j \in A} \Omega_{ij} g_i(\lambda) g_j(\lambda) x_i y_j \]

\[ \hat{V}_{xx}(\lambda) = \sum_{i \in A} \sum_{j \in A} \Omega_{ij} g_i(\lambda) g_j(\lambda) x_i x_j'. \]

Then, the proposed variance estimator \( \hat{V} \) in (20) can be written

\[ \hat{V} = \hat{V}_{yy}(\hat{\lambda}) - \hat{V}_{xy}(\hat{\lambda})' \hat{B} - \hat{B}' \hat{V}_{xy}(\hat{\lambda}) + \hat{B}' \hat{V}_{xx}(\hat{\lambda}) \hat{B}. \]

Now, by the mean value theorem,

\[ \hat{V}_{yy}(\hat{\lambda}) = \hat{V}_{yy}(\lambda_0) + \left[ \hat{R}_{yy}(\lambda^*) \right]' (\hat{\lambda} - \lambda_0), \]

(B.1)

where

\[ \hat{R}_{yy}(\lambda) = \partial \hat{V}_{yy}(\lambda) / \partial \lambda = 2 \sum_{i \in A} \sum_{j \in A} \Omega_{ij} g_i(\lambda) h_j(\lambda) y_i y_j \]

and \( \lambda^* \) is a point between \( \hat{\lambda} \) and \( \lambda_0 \). By the uniform continuity of \( g_i(\lambda) h_j(\lambda) \) around \( \lambda_0 \), we have, by (21),

\[ \hat{R}_{yy}(\lambda^*) = \hat{R}_{yy}(\lambda_0) + o_p\left((n^{-1}N^2)\right). \]  

(B.2)
Thus, inserting (B.2) into (B.1),

$$
\hat{V}_{yy}(\hat{\lambda}) = \hat{V}_{yy}(\lambda_0) + o_p\left(n^{-1}N^2\right)
$$

holds. Similarly, we have

$$
\hat{V}_{xy}(\hat{\lambda}) = \hat{V}_{xy}(\lambda_0) + o_p\left(n^{-1}N^2\right)
$$

and

$$
\hat{V}_{xx}(\hat{\lambda}) = \hat{V}_{xx}(\lambda_0) + o_p\left(n^{-1}N^2\right).
$$

Therefore, using $\hat{B} - B_0 = o_p(1)$, the result follows.

### C. Proof of (25)

Writing $\eta = (\lambda', \theta')'$, the estimating equation for $\eta$ is

$$
U(\eta) = \begin{pmatrix} U_1(\lambda, \theta) \\ U_2(\theta) \end{pmatrix} = 0,
$$

where

$$
U_1(\lambda, \theta) = \sum_{i \in A} d_i g_i(\lambda) \left[ 1, \mu(x_i, \theta) \right] - \sum_{i=1}^N \left[ 1, \mu(x_i, \theta) \right]
$$

and $U_2(\theta)$ is the estimating equation for the superpopulation parameters.

Thus, using the same argument for deriving (10), we have

$$
\hat{Y}_w = \sum_{i \in A} d_i g_i(\lambda_0) y_i - \left\{ E \left[ \frac{\partial \hat{Y}_w(\lambda_0)}{\partial \eta'} \right] \right\} \left\{ E \left[ \frac{\partial \hat{U}(\eta_0)}{\partial \eta} \right] \right\}^{-1} \hat{U}(\eta_0) + O_p(n^{-1}N).
$$

(C.1)

Now, define $z_i = (1, \hat{\mu}_i)'$ and using

$$
\frac{\partial \hat{Y}_w(\lambda_0)}{\partial \eta'} = \left( \frac{\partial \hat{Y}_w(\lambda_0)}{\partial \lambda'}, 0' \right)
$$

$$
\frac{\partial \hat{U}(\eta_0)}{\partial \eta} = \begin{pmatrix} \sum_{i \in A} d_i h_i(\lambda_0) z_i' & \partial \hat{U}_1/\partial \theta \\ 0 & \partial \hat{U}_2/\partial \theta \end{pmatrix},
$$

(C.1) reduces to (25).
D. Proof of Theorem 3

For each fixed $\theta$, we can apply Theorem 1 to get

\[
\hat{W}_\omega(\theta) - \hat{W}_l(\theta) = o_p(n^{-1}N). \tag{D.1}
\]

Here, the convergence in (D.1) is the point-wise convergence. Uniform convergence also follows in the neighborhood of $\theta_N$ since $\omega(x_i, y_i; \theta)$ is a uniformly continuous function of $\theta$. Now the following Lemma can be applied to get $\hat{\theta}_w - \hat{\theta}_l \to 0$ in probability.

**Lemma 3** Assume that

[C1] $\hat{U}_1(\theta) - \hat{U}_2(\theta) \to 0$ converges in probability uniformly in $\theta \in B$.

[C2] $\hat{U}_1(\theta)$ is continuous and $\hat{U}_1(\theta) = 0$ has exactly one solution $\hat{\theta}_1 \in B$.

[C3] $\hat{U}_2(\theta)$ is continuous and $\hat{U}_2(\theta) = 0$ has exactly one solution $\hat{\theta}_2 \in B$.

Then, $\hat{\theta}_1 - \hat{\theta}_2 \to 0$ in probability.

**Proof.** Since $\hat{U}_1(\theta) = 0$ has a unique solution at $\theta = \hat{\theta}_1$, given $\epsilon > 0$, there exists $\delta_1 > 0$ such that $|\hat{\theta}_1 - \hat{\theta}_2| > \epsilon$ implies $|\hat{U}_1(\hat{\theta}_1) - \hat{U}_1(\hat{\theta}_2)| = |0 - \hat{U}_1(\hat{\theta}_2)| > \delta_1$. Similarly, since $\hat{U}_2(\theta) = 0$ has a unique solution at $\theta = \hat{\theta}_2$, we can find $\delta_2 > 0$ such that $|\hat{\theta}_1 - \hat{\theta}_2| > \epsilon$ implies $|\hat{U}_2(\hat{\theta}_1) - \hat{U}_2(\hat{\theta}_2)| = |\hat{U}_2(\hat{\theta}_1) - 0| > \delta_2$. Thus, using $\hat{U}_1(\hat{\theta}_1) = \hat{U}_2(\hat{\theta}_2) = 0$ and letting $\delta = \min(\delta_1, \delta_2) > 0$, we have

\[
Pr\left(|\hat{\theta}_1 - \hat{\theta}_2| > \epsilon\right) \leq Pr\left\{|\hat{U}_1(\hat{\theta}_1) - \hat{U}_2(\hat{\theta}_1)| > \delta\right\} \\
+ Pr\left\{|\hat{U}_1(\hat{\theta}_2) - \hat{U}_2(\hat{\theta}_2)| > \delta\right\} \\
\leq 2Pr\left\{\sup_{\theta \in B}|\hat{U}_1(\theta) - \hat{U}_2(\theta)| > \delta\right\},
\]
which is less than $\epsilon$ for $n > n_0$ for some $n_0$ by [C1]. Therefore, the result follows.

Let $\theta_0 = p\lim \hat{\theta}_1 = p\lim \hat{\theta}_2$. By Lemma 2, we have $\sqrt{n} \left( \hat{\theta}_1 - \theta_0 \right) = O_p(1)$ and $\sqrt{n} \left( \hat{\theta}_2 - \theta_0 \right) = O_p(1)$. Therefore, $\sqrt{n} \left( \hat{\theta}_1 - \hat{\theta}_2 \right) = O_p(1)$ follows, which proves (32) using $\hat{U}_1(\theta) = \hat{W}_{\omega}(\theta)$ and $\hat{U}_2(\theta) = \hat{W}_{l}(\theta)$.
References


**Resume**

L'estimation par calage, pour laquelle les poids de sondage sont ajustés de manière à ce que certains estimateurs coïncident avec des totaux connus dans la population, est fréquemment utilisée en échantillonnage. L'estimateur par la régression généralisée est un exemple d'un estimateur de calage. Dans le cas où les facteurs d'ajustement sont exprimés selon une forme fonctionnelle, nous établissons l'équivalence asymptotique entre l'estimateur de calage avec variable instrumentale, où la variable instrumentale est directement déterminée à partir de la forme fonctionnelle dans l'équation de calage. L'estimation de la variance par linéarisation est traitée et appliquée à certains estimateurs de calage proposés récemment. Les résultats sont généralisés à l'estimateur solution de l'équation estimante calée. Les résultats d'une étude par simulation limitée sont présentés.
Table 1: Monte Carlo Mean squared errors of the point estimators, based on 10,000 Monte Carlo samples.

<table>
<thead>
<tr>
<th>Pop’n</th>
<th>Parameter</th>
<th>Estimator</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>n = 100</td>
</tr>
<tr>
<td>A</td>
<td>$\theta_1$</td>
<td>HT</td>
<td>0.02112</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Regression</td>
<td>0.01035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Empirical Likelihood</td>
<td>0.01064</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=3)</td>
<td>0.01036</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=2)</td>
<td>0.01035</td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>HT</td>
<td>0.1620</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Regression</td>
<td>0.1534</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Empirical Likelihood</td>
<td>0.1733</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=3)</td>
<td>0.1531</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=2)</td>
<td>0.1524</td>
</tr>
<tr>
<td>B</td>
<td>$\theta_1$</td>
<td>HT</td>
<td>0.01261</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Regression</td>
<td>0.01047</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Empirical Likelihood</td>
<td>0.01155</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=3)</td>
<td>0.01046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=2)</td>
<td>0.01045</td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>HT</td>
<td>0.2391</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Regression</td>
<td>0.2409</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Empirical Likelihood</td>
<td>0.2969</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=3)</td>
<td>0.2411</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=2)</td>
<td>0.2394</td>
</tr>
</tbody>
</table>
Table 2: Monte Carlo relative biases of the variance estimators, based on 10,000 Monte Carlo samples.

<table>
<thead>
<tr>
<th>Pop’n</th>
<th>Parameter</th>
<th>Estimator</th>
<th>Rel. Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$n = 100$</td>
</tr>
<tr>
<td>A</td>
<td>$\theta_1$</td>
<td>HT</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Regression</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Empirical Likelihood</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=3)</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=2)</td>
<td>-0.04</td>
</tr>
<tr>
<td>A</td>
<td>$\theta_2$</td>
<td>HT</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Regression</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Empirical Likelihood</td>
<td>-0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=3)</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=2)</td>
<td>-0.13</td>
</tr>
<tr>
<td>B</td>
<td>$\theta_1$</td>
<td>HT</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Regression</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Empirical Likelihood</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=3)</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=2)</td>
<td>-0.05</td>
</tr>
<tr>
<td>B</td>
<td>$\theta_2$</td>
<td>HT</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Regression</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Empirical Likelihood</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=3)</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic (M=2)</td>
<td>-0.10</td>
</tr>
</tbody>
</table>