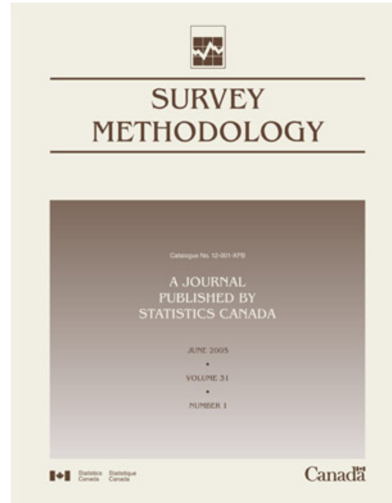
Catalogue no. 12-001-XIE

Survey Methodology

June 2005



Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada

Business Survey Methods Division

Survey Methodology

June 2005

Published by authority of the Minister responsible for Statistics Canada

May 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 31, No. 1, pp. 85-93
Statistics Canada, Catalogue No. 12-001

Towards Nonnegative Regression Weights for Survey Samples

Mingue Park and Wayne A. Fuller [1]

Abstract

Procedures for constructing vectors of nonnegative regression weights are considered. A vector of regression weights in which initial weights are the inverse of the approximate conditional inclusion probabilities is introduced. Through a simulation study, the weighted regression weights, quadratic programming weights, raking ratio weights, weights from logit procedure, and weights of a likelihood-type are compared.

Key Words: Raking ratio; Maximum likelihood; Quadratic programming; Simple Conditionally Weighted (SCW) estimator.

1. Introduction

In survey sampling, information about the population is often available at the analysis stage. One method of using this information is through regression estimation. There are a number of ways to construct a regression estimator of the population mean or total. One regression estimator of the mean is

$$\bar{y}_{\text{reg}} = \sum_{i=1}^{n} w_i\, y_i = \bar{y}_\pi + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\,\tilde{\beta}, \qquad (1)$$

where

$$w_i = \alpha_i + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\left(\sum_{j=1}^{n} \bar{\mathbf{x}}'_j\, \phi_{jj}^{-1}\, \mathbf{x}_j\right)^{-1} \mathbf{x}_i\, \phi_{ii}^{-1}, \qquad (2)$$

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

$$\tilde{\beta} = \left(\sum_{i=1}^{n} \mathbf{x}'_i\, \phi_{ii}^{-1}\, \mathbf{x}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{x}'_i\, \phi_{ii}^{-1}\, y_i,$$

$$\alpha_i = \left(\sum_{j=1}^{n} \pi_j^{-1}\right)^{-1} \pi_i^{-1},$$

$\mathbf{\Phi} = \text{diag}(\phi_{11}, ..., \phi_{nn})$ is a nonsingular diagonal matrix, the π_i's are the selection probabilities and $\bar{\mathbf{x}}_N$ is the population mean of $\mathbf{x}$. A possible choice of ϕ_{ii}^{-1} is α_i. A review of the use of such information in regression estimation for sample surveys is given by Fuller (2002).

It is well known that regression weights that are used to define a regression estimator such as (2) can be very large or (and) can be negative. If the regression weights are to be used to estimate a finite population total in a general purpose survey, it seems reasonable that no individual weight should be less than one. Also, it seems reasonable, on robustness grounds, to avoid very large weights.

There are several ways to construct regression weights with a reduced range of values. Huang and Fuller (1978) defined a procedure to modify the w_i so that there are no negative weights and no large weights. Husain (1969) suggested quadratic programming as a procedure to place bounds on the weights. Quadratic programming and a number of other procedures build on the fact that the weights can be defined as values that optimize some function. Deville and Särndal (1992) considered seven objective functions that can be used to construct weights. They suggested objective functions that can be used to produce weights which fall within a given range. Deville, Särndal and Sautory (1993) introduced the program, CALMAR, written as a SAS macro that can be used to calculate weights corresponding to four different objective functions when auxiliary information in the survey consists of known marginal counts in a frequency table.

Another modification of regression weights is to relax some of the restrictions used in constructing the estimator. Husain (1969) considered modifying weights for a simple random sample from a normal distribution. He derived the weights that minimize the mean square error (MSE) of the resulting estimator. Bardsley and Chambers (1984) considered an estimator based on an objective function and the division of the auxiliary variable into two components. They studied the behavior of the estimator from a model perspective. Rao and Singh (1997) studied an estimator in which tolerances are given for the difference between the final estimator for part of the auxiliary variables vector and the corresponding elements of the population vector.

In this paper, we consider different types of regression weights including a procedure based on Tillé's (1998) conditional selection probabilities. The approximate conditional

1. Mingue Park, University of Nebraska, 103 Miller Hall, Lincoln, NE, 68588-0712, U.S.A.; Wayne A. Fuller, Iowa State University, 221 Snedecor Hall, Ames, IA 50011-1210, U.S.A.

inclusion probabilities are used to compute regression weights that are positive for most samples. These regression weights are compared to raking ratio weights, to quadratic programming weights, weights from logit procedure, and to weights based on a likelihood-type objective function.

2. Maximum Likelihood and Raking Ratio

Consider a two-way table with r rows and c columns. The population cell U_{ij} contains N_{ij} elements; $i = 1, \ldots, r$, $j = 1, \ldots, c$. Assume marginal counts $N_{i\cdot}$, $N_{\cdot j}$ are known. The population characteristics of interest are the N_{ij} or, equivalently, $p_{ij} = N^{-1} N_{ij}$. For a simple random nonreplacement sample of size n, Deming and Stephan (1940) suggested a raking ratio procedure to get the solution for the cell frequencies. See also Stephan (1942). If we assume the sample is a random sample from a multinomial distribution defined by the population entries in a two way table, we can construct an estimator using the maximum likelihood procedure.

Deville and Särndal (1992) defined a class of calibration estimators, $\bar{y}_{\text{cal}}$, of the population mean of y as

$$\bar{y}_{\text{cal}} = \sum_{i=1}^{n} w_i \, y_i, \tag{3}$$

where the w_i's minimize the objective function $\sum_{i=1}^{n} G(w_i, \alpha_i)$ subject to constraints

$$\sum_{i=1}^{n} w_i = 1 \quad \text{and} \quad \sum_{i=1}^{n} w_i \, \mathbf{x}_i = \bar{\mathbf{x}}_N, \tag{4}$$

and $G(w_i, \alpha_i)$ is a measure of distance between an initial weight α_i and a final weight w_i. The raking ratio and maximum likelihood estimators of the population cell fraction, p_{ij}, belong to the class of calibration estimators.

The raking ratio weights for the population cell fraction, with a simple random sample, can be obtained by minimizing

$$\sum_{k=1}^{n} w_k \log\left(\frac{w_k}{n^{-1}}\right) - w_k + n^{-1}, \tag{5}$$

subject to the constraints (4) with

$$\mathbf{x}_k = (\delta_{1\cdot}, \ldots, \delta_{r\cdot}, \delta_{\cdot 1}, \ldots, \delta_{\cdot c}), \tag{6}$$

where $\delta_{i\cdot} = 1$ if k^{th} element belongs to the i^{th} row and $\delta_{i\cdot} = 0$ otherwise, and $\delta_{\cdot j} = 1$ if k^{th} element belongs to the j^{th} column and $\delta_{\cdot j} = 0$ otherwise. The raking ratio estimator for the population cell fraction p_{ij} is the estimator (3) where $y_k = 1$ if the k^{th} element belongs to cell ij and $y_k = 0$ otherwise.

For the maximum likelihood estimator of the population fraction, with a simple random sample, Deville and Särndal (1992) suggested minimizing

$$\sum_{k=1}^{n} - n^{-1} \log\left(\frac{w_k}{n^{-1}}\right) + w_k - n^{-1} \tag{7}$$

subject to (4) with $\mathbf{x}$ defined in (6).

Chen and Sitter (1999) suggested a *pseudo empirical likelihood estimator*. They defined the population likelihood of y_i as

$$\sum_{i=1}^{N} \log w_{i,U}, \tag{8}$$

where $w_{i,U}$ is the density at observation y_i. With a sample of size n, they suggested the pseudo empirical likelihood estimator of the form

$$\bar{y}_{\text{EL}} = \sum_{i=1}^{n} w_i \, y_i, \tag{9}$$

where w_i's are obtained by minimizing the function

$$-\sum_{i=1}^{n} \pi_i^{-1} \log w_i, \tag{10}$$

under the restrictions (4). The resulting w_i are equal to those obtained by minimizing (7) with $= N\pi_i$ under the restrictions (4).

Deville and Särndal (1992) showed that the raking ratio and maximum likelihood estimators are approximately equal to a regression estimator of the form (1), and, hence, have the same limiting distribution as the regression estimator. Weights for the raking ratio and maximum likelihood estimators are nonnegative if the solutions for the weights exist.

3. Weighted Regression Using Conditional Probabilities

Tillé (1998) suggested the use of approximate conditional inclusion probabilities, conditioning on the Horvitz-Thompson estimators of auxiliary variables, to compute an estimator for the population mean of the study variable. His approximation can be extended to produce regression weights that are nonnegative with high probability.

Assume that the vector of population means of auxiliary variables, $\bar{\mathbf{x}}_N$, is known. Consider the Horvitz-Thompson estimator of $\bar{\mathbf{x}}_N$ given by

$$\bar{\mathbf{x}}_{\text{HT}} = \frac{1}{N} \sum_{i=1}^{n} \frac{\mathbf{x}_i}{\pi_i}, \tag{11}$$

where $\mathbf{x}_i = (x_{i1}, ..., x_{ip})$ and π_i is the unconditional inclusion probability. Tillé (1998) introduced the simple conditionally weighted (SCW) estimator,

$$\bar{y}_{p\pi} = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\pi_{i|\bar{x}_{\mathrm{HT}}}}, \tag{12}$$

where $\pi_{i|\bar{x}_{\mathrm{HT}}}$ is the conditional inclusion probability of the i^{th} element conditioning on $\bar{\mathbf{x}}_{\mathrm{HT}}$. To construct the SCW-estimator of $\bar{y}_N$, the conditional inclusion probability $\pi_{i|\bar{x}_{\mathrm{HT}}}$ is required. If $\bar{\mathbf{x}}_{\mathrm{HT}}$ takes the value $\mathbf{t}$, we have

$$\pi_{i|\bar{x}_{\mathrm{HT}}} = \pi_i \frac{P\{\bar{\mathbf{x}}_{\mathrm{HT}} = \mathbf{t} \mid i \in A\}}{P\{\bar{\mathbf{x}}_{\mathrm{HT}} = \mathbf{t}\}}, \tag{13}$$

where A is the set of indices for the sample elements.

In order to compute the conditional inclusion probabilities, it is necessary to know the probability distribution of $\bar{\mathbf{x}}_{\mathrm{HT}}$ unconditionally and conditionally on the presence of each unit in the sample. Except for some particular cases, this probability distribution is very complex. For this reason, approximation of the conditional inclusion probability is considered.

Under the assumption that $\bar{\mathbf{x}}_{\mathrm{HT}}$ has an approximately normal distribution unconditionally and conditionally on the presence of each unit in the sample, the conditional inclusion probability (13) can be approximated by

$$\hat{\pi}_{i|\bar{x}_{\mathrm{HT}}} = \pi_i \left|\Sigma_{\bar{x}\bar{x}}\right|^{1/2} \left|\Sigma_{\bar{x}\bar{x},(i)}\right|^{-1/2} \exp\{0.5\,(\mathbf{G}_{\bar{x}\bar{x}} - \mathbf{G}_{\bar{x}\bar{x},(i)})\}, \tag{14}$$

where $\Sigma_{\bar{x}\bar{x}} = \mathrm{Var}\{\bar{\mathrm{x}}_{\mathrm{HT}} | F\}$, $\Sigma_{\bar{x}\bar{x},(i)} = \mathrm{Var}\{\bar{\mathbf{x}}_{\mathrm{HT}} | F,\, i \in A\}$,

$$\mathbf{G}_{\bar{x}\bar{x}} = (\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N)\, \Sigma_{\bar{x}\bar{x}}^{-1}\, (\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N)',$$

$$\mathbf{G}_{\bar{x}\bar{x},(i)} = (\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_{N,(i)})\, \Sigma_{\bar{x}\bar{x},(i)}^{-1}\, (\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_{N,(i)})',$$

$$\bar{\mathbf{x}}_{N,(i)} = E\{\bar{\mathbf{x}}_{\mathrm{HT}} \mid F,\, i \in A\} = (N\pi_i)^{-1} \mathbf{x}_i + N^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} (\pi_i\, \pi_j)^{-1} \pi_{ij}\, \mathbf{x}_j,$$

A is the set of indices appearing in the sample and $F = \{y_1, ..., y_N\}$ is the finite population. Tillé (1998) gives an expression for $\Sigma_{\bar{x}\bar{x},(i)}$ for the general case.

Assume the design covariance matrices $\Sigma_{\bar{x}\bar{x}}$ and $\Sigma_{\bar{x}\bar{x},(i)}$ are positive definite and assume the vector of auxiliary variables is normally distributed. Tillé (1999) showed that the SCW-estimator defined in (12) with the approximate conditional inclusion probabilities of (14) satisfies

$$\bar{y}_{p\pi} = \bar{y}_{\mathrm{HT}} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\, \boldsymbol{\beta}_N + O_p\,(n^{-1}) \tag{15}$$

$$= \bar{y}_{\mathrm{reg}} + O_p\,(n^{-1}), \tag{16}$$

where

$$\boldsymbol{\beta}_N = \Sigma_{\bar{x}\bar{x}}^{-1}\, \Sigma_{\bar{x}\bar{y}},$$

$$\bar{y}_{\mathrm{reg}} = \bar{y}_{\mathrm{HT}} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\, \hat{\boldsymbol{\beta}},$$

$$\hat{\boldsymbol{\beta}} = \hat{\Sigma}_{\bar{x}\bar{x}}^{-1}\, \hat{\Sigma}_{\bar{x}\bar{y}} = (\mathbf{X}'\, \boldsymbol{\Phi}^{-1}\, \mathbf{X})^{-1}\, \mathbf{X}'\, \boldsymbol{\Phi}^{-1}\, \mathbf{y},$$

$\mathbf{X} = (\mathbf{x}_1', ..., \mathbf{x}_n')'$, $\mathbf{y} = (y_1, ..., y_n)'$, the ij^{th} element of $\boldsymbol{\Phi}^{-1}$ is $N^{-2}(\pi_{ij}\, \pi_i\, \pi_j)^{-1}(\pi_{ij} - \pi_i\, \pi_j)$, $\Sigma_{\bar{x}\bar{x}}$ is the design variance of $\bar{\mathbf{x}}_{\mathrm{HT}}$, $\Sigma_{\bar{x}\bar{y}}$ is the design covariance of $\bar{\mathbf{x}}_{\mathrm{HT}}$ and $\bar{y}_{\mathrm{HT}}$, $\hat{\Sigma}_{\bar{x}\bar{x}}$ is the Horvitz-Thompson variance estimator of $\bar{\mathbf{x}}_{\mathrm{HT}}$, and $\hat{\Sigma}_{\bar{x}\bar{y}}$ is the Horvitz-Thompson estimator of the covariance of $\bar{\mathbf{x}}_{\mathrm{HT}}$ and $\bar{y}_{\mathrm{HT}}$.

Given a complex design, a number of the quantities in (14) are difficult to compute. However, approximations giving the same large sample properties for the estimator are relatively easy to compute. We replace $\Sigma_{\bar{x}\bar{x}}$ and $\Sigma_{\bar{x}\bar{x},(i)}$ with estimators, replace $\bar{\mathbf{x}}_{N,(i)}$ with $\bar{\mathbf{x}}_N + \mathbf{d}_{x_i}$, define

$$\hat{\mathbf{M}}_{\bar{x}\bar{y}} = \sum_{i \in A} (N\pi_i)^{-1} + \mathbf{d}_{x_i}'\, y_i, \tag{17}$$

and assume

$$\mathrm{Var}\{n(\hat{\mathbf{M}}_{\bar{x}\bar{y}} - \mathbf{M}_{\bar{x}\bar{y}})\} = O(n^{-1}), \tag{18}$$

$$\mathbf{d}_{x_i} = O_p\,(n^{-1}), \tag{19}$$

where $\mathbf{d}_{x_i}$ is a function of the sample and $\mathbf{M}_{\bar{x}\bar{y}}$ is a population quantity. Often $\mathbf{M}_{\bar{x}\bar{y}}$ is the population covariance matrix $\Sigma_{\bar{x}\bar{y}}$, but this equality is not required in order for the estimator to be well defined. In many cases one can compute $\mathbf{d}_{x_i}$ as a multiple of the jackknife deviate. Also in many situations, an adequate value for the estimator, $\tilde{\Sigma}_{\bar{x}\bar{x},(i)}$, of $\Sigma_{\bar{x}\bar{x},(i)}$ is $n^{-1}(n-1)\hat{\Sigma}_{\bar{x}\bar{x}}$. We write our generalization of (14) as

$$\tilde{\pi}_{i|\bar{x}_{\mathrm{HT}}} = \pi_i \left|\hat{\Sigma}_{\bar{x}\bar{x}}\right|^{1/2} \left|\tilde{\Sigma}_{\bar{x}\bar{x},(i)}\right|^{-1/2} \exp\{0.5\,(\hat{\mathbf{G}}_{\bar{x}\bar{x}} - \tilde{\mathbf{G}}_{\bar{x}\bar{x},(i)})\}, \tag{20}$$

where

$$\hat{\mathbf{G}}_{\bar{x}\bar{x}} = (\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N)\, \hat{\Sigma}_{\bar{x}\bar{x}}^{-1}\, (\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N)',$$

$$\tilde{\mathbf{G}}_{\bar{x}\bar{x},(i)} = (\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N - \mathbf{d}_{x_i})\, \tilde{\Sigma}_{\bar{x}\bar{x},(i)}^{-1}\, (\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N - \mathbf{d}_{x_i})'.$$

Let the estimator (12) constructed with the $\tilde{\pi}_{i\,|\,\bar{x}_{\mathrm{HT}}}$ of (20) be

$$\bar{y}_{p\tilde{\pi}} = N^{-1} \sum_{i=1}^{n} \tilde{\pi}_{i\,|\,\bar{x}_{\mathrm{HT}}}^{-1}\, y_i. \tag{21}$$

An approximate conditional inclusion probability with a simple random sample and a single auxiliary variable is

$$\tilde{\pi}_{i \mid \bar{x}_n} = \frac{n}{N}\left[\frac{\hat{\sigma}_{\bar{x}}}{\tilde{\sigma}_{\bar{x},(i)}}\right] \exp\left\{\frac{1}{2}\left[\frac{(\bar{x}_n - \bar{x}_N)^2}{\hat{\sigma}^2_{\bar{x}}} - \frac{(\bar{x}_n - \bar{x}_N - d_{x_i})^2}{\tilde{\sigma}^2_{\bar{x},(i)}}\right]\right\},$$

where

$$d_{x_i} = [n(N-1)]^{-1}(N-n)(x_i - \bar{x}_N),$$

$$\tilde{\sigma}^2_{\bar{x},(i)} = \frac{(N-n)(n-1)}{n^2(N-2)}\left[s_x^2 - \frac{N(x_i - \bar{x}_n)^2}{(N-1)^2}\right] \approx \frac{n-1}{n}\hat{\sigma}^2_{\bar{x}},$$

$$\hat{\sigma}^2_{\bar{x}} = (n^{-1} - N^{-1})\, s_x^2,$$

and

$$s_x^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2.$$

In this case, $d_{x_i} = \bar{x}_{N,(i)} - \bar{x}_N$ and $M_{\bar{x}\bar{y}} = \mathrm{Cov}(\bar{x}_{\mathrm{HT}}, \bar{y}_{\mathrm{HT}})$.

The SCW-estimator (21) with the approximate conditional inclusion probabilities is not calibrated, that is, the estimator (21) for the mean of the vector of auxiliary variables is not the vector of population means. It is relatively easy to standardize the probabilities so that they sum to one or sum to the stratum fraction for stratified sampling. To construct a calibrated estimator for the general case, we suggest computing the regression estimator with $[\sum_{j=1}^{n}\tilde{\pi}^{-1}_{j \mid \bar{x}_{\mathrm{HT}}}]^{-1}\tilde{\pi}^{-1}_{i \mid \bar{x}_{\mathrm{HT}}}$ as initial weights. The suggested estimator is

$$\bar{y}_{w\mathrm{reg}} = \bar{y}_c + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_c)\,\hat{\beta}_{c,1} = \sum_{i=1}^{n} w_i\, y_i, \qquad (22)$$

where

$$(\bar{y}_c, \bar{\mathbf{x}}_c) = \sum_{i=1}^{n}\alpha_i(y_i, \mathbf{x}_i),$$

$$\left(\hat{\beta}_{c,0}, \hat{\beta}_{c,1}\right)' = \left[\sum_{i=1}^{n}\alpha_i\,\mathbf{z}_i'\,\mathbf{z}_i\right]^{-1}\left[\sum_{i=1}^{n}\alpha_i\,\mathbf{z}_i'\,y_i\right],$$

$$\mathbf{z}_i = (1,\ \mathbf{x}_i - \bar{\mathbf{x}}_c),$$

$$\alpha_i = \left[\sum_{j=1}^{n}\tilde{\pi}^{-1}_{j \mid \bar{x}_{\mathrm{HT}}}\right]^{-1}\tilde{\pi}^{-1}_{i \mid \bar{x}_{\mathrm{HT}}},$$

$$w_i = \alpha_i + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_c)\left[\sum_{j=1}^{n}\alpha_j(\mathbf{x}_j - \bar{\mathbf{x}}_c)'(\mathbf{x}_j - \bar{\mathbf{x}}_c)\right]^{-1}\alpha_i(\mathbf{x}_i - \bar{\mathbf{x}}_c)',$$

and $\tilde{\pi}_{i \mid \bar{x}_{\mathrm{HT}}}$ is the approximate conditional inclusion probability of (20). We assume the vector of auxiliary variables contains one so that the estimator is location invariant.

The estimator (21) is approximately equal to a regression estimator and estimator (22) is also approximately equal to the same regression estimator.

Theorem: Let a sequence of populations and samples, $\{F_N, A_N\}$, satisfy

$$(\bar{y}_{\mathrm{HT}}, \bar{\mathbf{x}}_{\mathrm{HT}}) - (\bar{y}_N, \bar{\mathbf{x}}_N) = O_p(n^{-1/2}). \qquad (23)$$

Assume that the sequences of estimated covariance matrices, $\hat{\Sigma}_{\bar{x}\bar{x}}$ and $\tilde{\Sigma}_{\bar{x}\bar{x},(i)}$, satisfy

$$\left[\mathbf{D}^{-1/2}\tilde{\Sigma}_{\bar{x}\bar{x},(i)}\mathbf{D}^{-1/2}\right]^{-1} - \left[\mathbf{D}^{-1/2}\hat{\Sigma}_{\bar{x}\bar{x}}\,\mathbf{D}^{-1/2}\right]^{-1} = O_p(n^{-1}), \qquad (24)$$

where $\mathbf{D}$ denotes a diagonal matrix having the elements of the diagonal of $\hat{\Sigma}_{\bar{x}\bar{x}}$ on its diagonal. Let $\mathbf{d}_{x_i}$ be a function of the sample satisfying (19) and assume (18) holds. Assume the sequence of Horvitz-Thompson variance estimators satisfies

$$\mathrm{Var}\left\{n\left[\mathrm{Vech}\left(\hat{\Sigma}_{\bar{z}\bar{z},\mathrm{HT}} - \Sigma_{\bar{z}\bar{z}}\right)\right]\right\} = O(n^{-1}), \qquad (25)$$

where $\mathbf{z}_i = (\mathbf{x}_i, y_i)$ and $\Sigma_{\bar{z}\bar{z}}$ is positive definite. Assume $E\{\tilde{\pi}^{-2}_{i \mid \bar{x}_{\mathrm{HT}}}\}$ is bounded, where $\tilde{\pi}_{i \mid \bar{x}_{\mathrm{HT}}}$ is defined in (20). Then, the SCW– estimator $\bar{y}_{p\tilde{\pi}}$ of (21) satisfies

$$\bar{y}_{p\tilde{\pi}} = \bar{y}_{\mathrm{HT}} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\,\boldsymbol{\theta}_N + O_p(n^{-1}) = \bar{y}_{\mathrm{HT}} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\,\hat{\boldsymbol{\theta}} + O_p(n^{-1}),$$

where $\hat{\boldsymbol{\theta}} = \hat{\Sigma}^{-1}_{\bar{x}\bar{x}}\,\hat{\mathbf{M}}_{\bar{x}\bar{y}}$ and $\boldsymbol{\theta}_N = \Sigma^{-1}_{\bar{x}\bar{x}}\,\mathbf{M}_{\bar{x}\bar{y}}$.

If $\mathrm{Var}\{\sum_{i=1}^{n}\pi_i^{-1}\} > 0$, assume $\mathbf{x}_i$ contains one as an element. Assume $\mathbf{M}_{\bar{x}\bar{y}} = \Sigma_{\bar{x}\bar{y}}$. Then the weighted regression estimator of (22) satisfies

$$\bar{y}_{w\mathrm{reg}} = \bar{y}_{\mathrm{HT}} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\,\hat{\boldsymbol{\theta}} + O_p(n^{-1}).$$

For proof, see the appendix.

To illustrate the nature of the different types of regression weights, we selected a simple random sample of size 40 from a normal population with mean zero and variance one. The sample mean is –0.614 and the population mean is zero. The weight for the regression estimator is given by (2) with $\alpha_i = \phi_{ii}^{-1} = n^{-1}$. The weights for the raking ratio and MLE are obtained by minimizing the objective functions (5) and (7), respectively, under the restriction (4). Weights for the SCW-weighted regression estimator are given in (22). The weights are plotted against the sample x values in Figure 1. Five of the simple regression weights are less than zero because of the large discrepancy between the sample and the population means. All weights for the SCW-weighted regression estimator, MLE and raking ratio are nonnegative. Figure 1 shows that the behaviors of raking ratio and SCW-weighted regression weights are similar and that MLE has an extremely large weight in this sample.

Table 1 contains selected weights for the smallest x values, x values close to the sample mean, x values close to the population mean, and the largest x values. For the x–values farthest from the population mean MLE gives the largest weights. For x–values near the sample mean the ordinary least squares weights are close to n^{-1} while the other weights are less than n^{-1}. The MLE weights are close to n^{-1} for x–values close to the population mean while the other weights are larger.

Table 1
Selected Regression Weights for Illustrated Example

x	Weights multiplied by $n = 40$			
	Reg	W. Reg	Raking	MLE
– 2.103	– 0.56	0.12	0.16	0.40
– 1.941	– 0.40	0.12	0.20	0.40
– 1.727	– 0.16	0.20	0.24	0.44
– 0.710	0.88	0.68	0.68	0.68
– 0.670	0.96	0.72	0.68	0.68
– 0.468	1.16	0.88	0.84	0.76
– 0.103	1.52	1.28	1.24	0.92
0.021	1.68	1.44	1.40	1.00
0.097	1.76	1.56	1.52	1.08
0.628	2.32	2.60	2.60	1.84
0.662	2.36	2.68	2.72	1.92
1.237	2.96	4.60	4.88	9.12

Simulation Study

To compare the alternative methods of constructing regression weights we conducted a simulation study. A total of 30,000 simple random samples of size 32 were selected from a χ^2 distribution with two degrees of freedom. The parameters being estimated are those of the infinite generating mechanism. Let x_i be the value for the i^{th} sampled element. Six estimation procedures were considered.

1. Ordinary least squares regression (OLS)
2. Quadratic programming with upper and lower bounds (QP)
3. Weighted regression with SCW weights (SCW reg)
4. Maximum likelihood objective function (MLE)
5. Raking objective function (Raking reg)
6. Logit procedure with upper and lower bounds (Logit)

The weights for the OLS estimator were calculated by (2) with $\alpha_i = n^{-1}$. The quadratic programming weights minimize $\sum_{i=1}^{n} w_i^2$ subject to the constraint $0 \le w_i \le 0.065$ for all i and subject to constraints (4). The quadratic programming procedure is equivalent to the truncated linear method of case 7 of Deville and Särndal (1992). Weights for the SCW weighted regression were calculated by minimizing $\sum_{i=1}^{n} \alpha_i^{-1} w_i^2$ subject to constraints (4), where α_i is defined in (22). The weights for raking and maximum likelihood were obtained by minimizing the objective functions (5) and (7), respectively, under the restriction (4). Weights calculated by the logit procedure minimize the function $\sum_{i=1}^{n} G(nw_i)$ subject to constraints (4),where

$$G(nw_i) = a^{-1}\left[(nw_i)\ln(nw_i) + (u - nw_i)\ln\left(\frac{u - nw_i}{u-1}\right)\right],$$

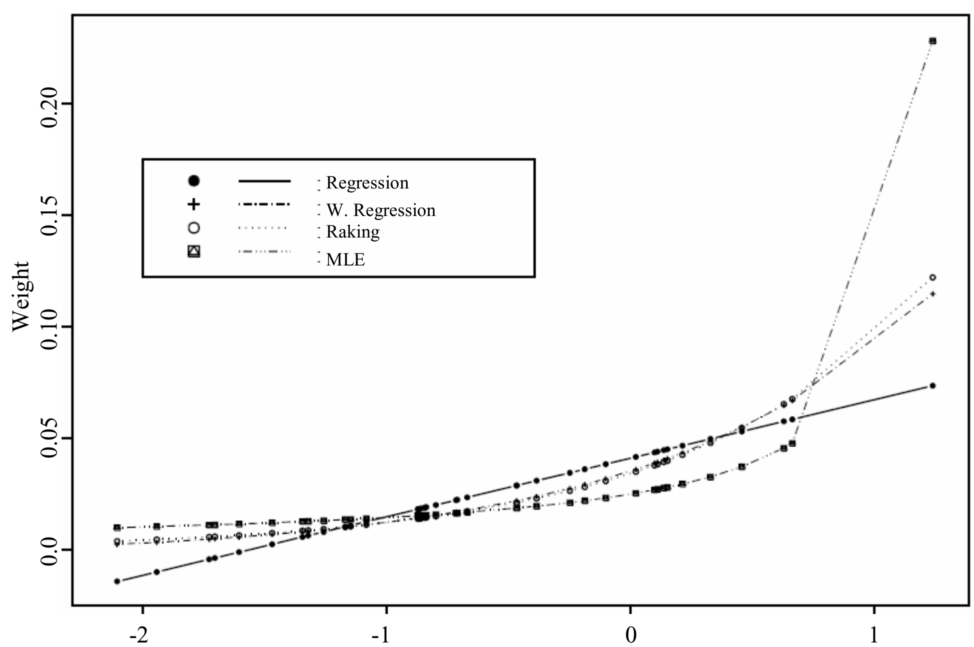


Figure 1. Comparison of four sets of weights.

if $0 < nw_i < u$ and ∞ elsewhere, $a = u(u-1)^{-1}$, and $u = 2.08$. Note that the solution for the logit procedure, if it exists, satisfies the bound restrictions $0 \le w_i \le 0.065$ for all i. The logit procedure was introduced as a case 6 in Deville and Särndal (1992). As the upper bound for the weight, 0.065 was used so that 3,026 samples (approximately 10%) have at least one raking regression weight greater than 0.065. In 99 samples among 30,000, no solution for the quadratic programming and logit procedure is possible because no feasible point satisfies (4) and the bound restriction. For those 99 samples, the maximum of the OLS regression weights was used as the upper bound for the quadratic programming and logit procedures.

Table 2 shows the average of the sum of squares for the six weights. The average weight is 1/32 = 0.03125 for every estimator. The least squares procedures have the smallest sum of squares of the weights because this is the objective function for those procedures. The least squares procedures also have a slightly smaller range in the sum of squares. One percent of the least squares samples have a normalized mean of squares greater than 1.401 while one percent of the mean of squares for raking are greater than 1.441.

Table 2
Monte Carlo Average of the Sum of Squares of the Weights

	OLS	QP	SCW Reg	MLE	Raking Reg	Logit
Average of $\mathbf{w}'\mathbf{w}$ (×32)	1.043	1.044	1.045	1.053	1.045	1.045

Table 3 contains properties for the minimum of the weights. Maximum likelihood has the largest average minimum weight while the least squares procedures have a smaller average for the minimum weight. The variance of the minimum weight is largest for the ordinary least squares procedures. Note that QP permits weights that equal the lower bound of zero.

Table 3
Monte Carlo Mean, Variance and Quantiles of the Minimum Weight

			Quantiles (×32)				
Procedure	Mean (×10²)	Variance (×10⁵)	0.01	0.10	0.50	0.90	0.99
OLS	2.22	6.46	−0.10	0.34	0.79	0.96	1.00
QP	2.21	6.32	0.00	0.32	0.79	0.96	1.00
SCW Reg	2.44	3.58	0.22	0.49	0.84	0.97	0.99
MLE	2.45	2.79	0.33	0.52	0.83	0.97	1.00
Raking Reg	2.36	3.81	0.20	0.45	0.81	0.97	1.00
Logit	2.25	5.23	0.09	0.36	0.78	0.96	1.00

Among the procedures without bound restrictions on the weights, the ordinary least squares procedure has smaller maximum weight on average and much smaller variance for the maximum. See Table 4. The SCW-weighted regression has a smaller fraction of very large weights than MLE or raking ratio but a higher fraction of large weights than the ordinary least squares procedure. The bounded QP and Logit procedures have smaller mean and variance for the maximum weight than the procedures with no upper bound restrictions.

Table 4
Monte Carlo Mean, Variance and Quantiles of the Maximum Weight

			Quantiles (×32)				
Procedure	Mean (×10²)	Variance (×10⁵)	0.01	0.10	0.50	0.90	0.99
OLS	4.25	17.35	1.00	1.03	1.20	1.92	2.93
QP	4.17	11.91	1.00	1.03	1.20	1.92	2.08
SCW Reg	4.56	26.42	1.03	1.07	1.27	2.12	3.47
MLE	4.75	56.13	1.00	1.04	1.25	2.31	4.72
Raking Reg	4.46	30.25	1.00	1.03	1.23	2.09	3.63
Logit	4.13	10.23	1.00	1.03	1.21	1.82	2.08

To evaluate the performance of the procedures when the linear model does not hold, we considered estimation of the percentiles of the distribution function of x. Table 5 contains the Monte Carlo bias of the percentile estimators where the table entries are

$$[\min\{P, (1-P)\}]^{-1}[\hat{E}\{\hat{P}\} - P]\times 100,$$

and P is the percentile. For example, the Monte Carlo estimated relative bias in the ordinary least squares estimator of the 0.01 percentile is −7.75%. The ordinary least squares estimator has the largest biases in estimating the population percentiles, among the procedures without bound restrictions. The MLE has the smallest bias for all percentiles except the 75th, 95th and 99th, where the SCW-weighted regression estimator has the smallest bias. For samples of size 32, many samples contain no observation greater than the 99th percentile. The QP and Logit procedures have larger bias than other procedures except for the 75th percentile. The biases of the QP and Logit procedures are relatively large for the lower percentiles.

Table 5
Monte Carlo Standardized Bias in Percentile Estimators

Percentile	Procedure					
	OLS	QP	SCW Reg	MLE	Raking Reg	Logit
0.01	−7.75	−8.43	−2.88	−2.13	−4.70	−8.30
0.05	−7.27	−7.95	−2.58	−1.82	−4.30	−7.85
0.10	−6.66	−7.31	−2.27	−1.57	−3.91	−7.26
0.25	−5.25	−5.82	−1.79	−1.25	−3.13	−5.89
0.50	−3.21	−3.46	−1.37	−1.16	−2.18	−3.53
0.75	−2.30	−2.07	−1.60	−2.21	−2.25	−1.78
0.90	4.60	5.31	1.27	0.22	2.62	5.68
0.95	12.75	13.33	6.01	6.41	9.52	13.15
0.99	32.94	32.36	19.03	22.66	26.65	30.03

Table 6 contains the relative MSE of the percentile estimators where the table entries are

$$[\min\{P, (1-P)\}]^{-2}[\hat{E}\{\hat{P} - P\}^2]\times 100.$$

Thus the relative mean square error of the OLS estimator of the 0.01 percentile is 283.27%. Although the OLS estimator of the 0.01 percentile had the largest bias OLS has the smallest mean square error for the 0.01 percentile among the procedures without bound restrictions. The QP, OLS and Logit procedures are superior for the extreme percentiles while the other procedures perform better for the middle percentiles.

Table 6
Monte Carlo Relative MSE of Percentile Estimators

Percentile	Procedure					
	OLS	QP	SCW Reg	MLE	Raking Reg	Logit
0.01	283.27	282.50	309.23	311.58	296.37	282.76
0.05	53.91	54.23	57.41	57.07	54.97	54.06
0.10	25.50	25.97	26.40	25.79	25.26	25.80
0.25	8.00	8.41	7.77	7.23	7.42	8.41
0.50	1.99	2.07	1.88	1.71	1.83	2.12
0.75	3.65	3.68	3.62	3.66	3.63	3.67
0.90	14.50	14.60	14.25	14.57	14.36	14.56
0.95	39.40	38.65	40.99	41.66	39.93	37.94
0.99	200.17	196.24	235.71	216.22	205.85	194.33

In 562 of 30,000 samples at least one of the OLS regression weights is negative. In 17 of the samples at least one of the original SCW regression weights was negative. The use of quadratic programming with the OLS objective function (QP) to produce weights greater than or equal to zero and less than 0.065 increases the average sum of squares by less than one percent. See Table 7. Using quadratic programming to bound the SCW regression weights (SCW (QPL)) by zero increases the average sum of squares very little because there are so few weights that are changed.

Table 7
Monte Carlo Average of the Sum of Squares of the Weights for Samples with at Least One Negative OLS Weight

	OLS	QP	SCW – Reg	SCW (QPL)	MLE	Raking – Reg
Average of $\mathbf{w}'\mathbf{w}$ (×32)	1.208	1.217	1.226	1.227	1.342	1.242

Table 8 gives the Monte Carlo MSE for the 562 samples with negative ordinary least squares weights. The quadratic programming procedure is superior to other nonnegative weight procedures for the 0.01 percentile and is inferior for the 0.99 percentile. Of the 562 samples, 497 had a sample mean greater than the population mean. Recall that the study population has an exponential distribution. Because the weight on the largest observation is zero in the 497 samples there is a 100 percent error in the quadratic programming estimator of the 0.99 percentile for most of the 497 samples with a sample mean greater than the population mean. In sampling from a finite population the bound on the weights would be greater than or equal to N^{-1} and the MSE of the quadratic programming procedure for the 0.99 percentile would be reduced.

Quadratic programming is superior to the other calibrated procedures for the 0.01 percentile in samples with negative OLS weights. Raking regression and SCW-weighted regression are superior to MLE for the 0.01 and 0.05 percentiles. This is because MLE often has the largest maximum weight.

Table 8
Monte Carlo Relative MSE of Percentile Estimators for Samples with at Least One Negative OLS Weight

Percentile	Procedure				
	OLS	QP	SCW (QPL)	MLE	Raking Reg
0.01	287.52	291.11	350.58	461.80	344.06
0.05	76.04	70.58	75.80	88.71	72.50
0.10	44.80	40.74	39.31	38.84	36.05
0.25	20.24	19.14	14.72	9.91	12.56
0.50	5.03	5.31	3.65	2.26	3.35
0.75	5.02	4.53	3.36	4.24	3.45
0.90	23.77	23.69	20.04	18.80	20.49
0.95	51.54	46.04	30.79	28.28	32.54
0.99	206.33	90.08	39.40	57.54	43.49

In 3,026 of 30,000 samples, at least one of the raking regression weights is greater than 0.065. In 2,152 samples, at least one of the OLS regression weights is greater than 0.065, and in 3,209 samples at least one of the SCW regression weights is greater than 0.065. The use of quadratic programming with the OLS objective function to produce weights in (0.000, 0.065) increases the average sum of squares by 1.5 percent. Using quadratic programming to bound the SCW regression weights by 0.000 and 0.065 increases the average sum of squares 0.8 percent. See the column for SCW (QP) of Table 9.

Table 9
Monte Carlo Average of the Sum of Squares of the Weights for Samples with at Least One Raking Reg Weight Greater than 0.065

	OLS	QP	SCW – Reg	SCW (QP)	Raking – Reg	Logit	MLE
Average of $\mathbf{w}'\mathbf{w}$ (×32)	1.210	1.228	1.221	1.231	1.228	1.232	1.290

Table 10 gives the Monte Carlo relative MSE for the 3,026 samples with raking regression weights greater than 0.065. The quadratic programming is superior to SCW (QP) and Logit for the 0.01, 0.95 and 0.99 percentile and the Logit procedure is superior to quadratic programming for other percentiles.

Table 10
Monte Carlo Relative MSE of Percentile Estimators for Samples with at Least One Raking Reg Weight Greater than 0.065

Percentile	Procedure						
	OLS	QP	SCW – Reg	SCW (QP)	Raking – Reg	Logit	MLE
0.01	139.96	130.53	173.86	146.40	124.02	173.65	206.65
0.05	39.83	42.88	39.35	41.69	39.87	37.14	40.83
0.10	26.31	30.92	22.40	28.10	28.88	20.21	19.98
0.25	13.56	17.72	10.13	15.69	17.71	8.65	7.01
0.50	3.95	4.87	3.32	4.75	5.37	3.03	2.28
0.75	4.84	5.35	4.89	5.58	5.37	5.05	5.48
0.90	27.98	29.04	28.70	29.34	29.32	28.79	32.07
0.95	74.15	67.54	85.02	68.12	65.98	83.13	95.99
0.99	198.77	179.58	219.16	181.17	172.45	212.38	226.73

Discussion

We began the research with the conjecture that starting with the SCW weights in a regression estimator would produce weights that were almost always positive and that the weights would have desirable properties as measured by the ability to estimate the distribution function of x. To some extent these results support the conjectures. The minimum weights of the SCW regression are larger than those of OLS and comparable to those for raking. Quadratic programming can be used to remove the negative weights in the few samples with negative weights. If no upper bound is imposed, the maximum weights for the SCW weighted regression fall between those of least squares and raking.

It is known that all of the procedures in our simulation study have the same order $n^{-1/2}$ properties. Our simulation and the study of generalized raking procedures done by Deville *et al.* (1993) indicate that there are also modest differences in small samples. No procedure is superior with respect to all criteria. Because of the poor performance for the extreme percentiles, we recommend against the use of the MLE objective function. The quadratic programming and Logit procedure produced weights with marginally smaller sums of squares, marginally smaller maximum weights, and marginally smaller MSE for extreme percentiles than the raking regression. The MLE, SCW regression and raking procedures give marginally larger minimum weights and marginally smaller MSE for the middle percentiles of the x distribution than quadratic programming and Logit procedure. The performances of quadratic programming and Logit procedures in estimating the distribution function of x are comparable.

Appendix

Proof. The ratio of the determinants of estimated covariance matrices in (20) is

$$\frac{\left|\tilde{\Sigma}_{\bar{x}\bar{x},(i)}\right|}{\left|\hat{\Sigma}_{\bar{x}\bar{x}}\right|} = 1 + O_p(n^{-1}) \qquad (26)$$

by assumptions (24) and (25). The difference $\tilde{\mathbf{G}}_{\bar{x}\bar{x},(i)} - \hat{\mathbf{G}}_{\bar{x}\bar{x}}$ is

$$(\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N)\left(\tilde{\Sigma}^{-1}_{\bar{x}\bar{x},(i)} - \hat{\Sigma}^{-1}_{\bar{x}\bar{x}}\right)(\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N)' - 2(\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N)\,\tilde{\Sigma}^{-1}_{\bar{x}\bar{x},(i)}\,\mathbf{d}'_{x_i} + \mathbf{d}_{x_i}\,\tilde{\Sigma}^{-1}_{\bar{x}\bar{x},(i)}\,\mathbf{d}'_{x_i}.$$

By assumptions (23) and (24),

$$\exp\{0.5[(\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N)(\tilde{\Sigma}^{-1}_{\bar{x}\bar{x},(i)} - \hat{\Sigma}^{-1}_{\bar{x}\bar{x}})(\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N)']\} = 1 + O_p(n^{-1}). \qquad (27)$$

Using assumptions (24) and (19), the Taylor expansion at $\mathbf{d}_{x_i} = 0$ gives

$$\begin{aligned}\exp[-(\bar{\mathbf{x}}_{\mathrm{HT}} - \bar{\mathbf{x}}_N)\,\tilde{\Sigma}^{-1}_{\bar{x}\bar{x},(i)}\,\mathbf{d}'_{x_i} + 0.5\,\mathbf{d}_{x_i}\,\tilde{\Sigma}^{-1}_{\bar{x}\bar{x},(i)}\,\mathbf{d}'_{x_i}] &= 1 + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\,\tilde{\Sigma}^{-1}_{\bar{x}\bar{x},(i)}\,\mathbf{d}'_{x_i} + O_p(n^{-1})\\ &= 1 + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\,\hat{\Sigma}^{-1}_{\bar{x}\bar{x}}\,\mathbf{d}'_{x_i} + O_p(n^{-1}). \qquad (28)\end{aligned}$$

Thus, by (26), (27) and (28),

$$[N\tilde{\pi}_{i|x_{\mathrm{HT}}}]^{-1} = (N\pi_i)^{-1}[1 + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\,\hat{\Sigma}^{-1}_{\bar{x}\bar{x}}\,\mathbf{d}'_{x_i}] + O_p(n^{-2}).$$

By assumptions (18), (23) and (25), and using the fact that $E\{\tilde{\pi}^{-2}_{i|x_{\mathrm{HT}}}\}$ is bounded,

$$\begin{aligned}\bar{y}_{p\tilde{\pi}} &= \bar{y}_{\mathrm{HT}} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\,\hat{\boldsymbol{\theta}} + O_p(n^{-1})\\ &= \bar{y}_{\mathrm{HT}} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\,\boldsymbol{\theta}_N + O_p(n^{-1}). \qquad (29)\end{aligned}$$

If one is an element of $\mathbf{x}_i$ or $\mathrm{Var}\{\sum_{i=1}^{n}\pi_i^{-1}\} = 0$, and if $\mathbf{M}_{\bar{x}\bar{y}} = \Sigma_{\bar{x}\bar{y}}$, the SCW-estimator for the population mean of vector $\mathbf{q}_i = (1,\ \mathbf{x}_i)$ satisfies

$$\bar{\mathbf{q}}_{p\tilde{\pi}} = N^{-1}\sum_{i=1}^{n}\tilde{\pi}^{-1}_{i|\bar{x}_{\mathrm{HT}}}\,\mathbf{q}_i = (1,\ \bar{\mathbf{x}}_N) + O_p(n^{-1}), \qquad (30)$$

because the $\boldsymbol{\theta}$ for $\mathbf{x}$ is the identity matrix. By (30),

$$(\bar{\mathbf{x}}_c,\ \bar{y}_c) = N\left[\sum_{i=1}^{n}\tilde{\pi}^{-1}_{i|\bar{x}_{\mathrm{HT}}}\right]^{-1}(\bar{\mathbf{x}}_{p\tilde{\pi}},\ \bar{y}_{p\tilde{\pi}}) = (\bar{\mathbf{x}}_{p\tilde{\pi}},\ \bar{y}_{p\tilde{\pi}}) + O_p(n^{-1}). \quad (31)$$

Thus,

$$\begin{aligned}\bar{y}_{w\mathrm{reg}} &= \bar{y}_c + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_c)\,\hat{\boldsymbol{\beta}}_{c,1}\\ &= \bar{y}_{p\tilde{\pi}} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{p\tilde{\pi}})\,\hat{\boldsymbol{\beta}}_{c,1} + (\bar{y}_c - \bar{y}_{p\tilde{\pi}}) + (\bar{\mathbf{x}}_{p\tilde{\pi}} - \bar{\mathbf{x}}_c)\,\hat{\boldsymbol{\beta}}_{c,1}\\ &= \bar{y}_{p\tilde{\pi}} + O_p(n^{-1})\\ &= \bar{y}_{\mathrm{HT}} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\mathrm{HT}})\,\hat{\boldsymbol{\theta}} + O_p(n^{-1}),\end{aligned}$$

by (30), (31) and (29).

Acknowledgements

This research was partly supported by Cooperative Agreement 43-3AEU-3-80088 between Iowa State University, the USDA National Agricultural Statistics Service and the U.S. Bureau of the Census, and by Cooperative Agreement 68-3A75-14 between the USDA Natural Resources Conservation Service and Iowa State University. We thank the Associate editor and referees for comments that improved the paper.

References

Bardsley, P., and Chambers, R.L. (1984). Multipurpose estimation from unbalanced samples. *Applied Statistics*, 33, 290-299.

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.

Deming, W.E., and Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11, 427-444.

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 procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

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

Huang, E.T., and Fuller, W.A. (1978). Nonnegative regression estimation for sample survey data. *Proceedings of the Social Statistics Section*, American Statistical Association, 300-305.

Husain, M. (1969). Construction of Regression Weights for Estimation in Sample Surveys. Unpublished M.S. thesis, Iowa State University, Ames, Iowa.

Rao, J.N.K., and Singh, A.C. (1997). A ridge shrinkage method for range restricted weight calibration in survey sampling. *Proceedings of the section on survey research methods*, American Statistical Association, 57-64.

Stephan, F.F. (1942). An alternative method of adjusting sample frequency tables when expected marginal totals are known. *Annals of Mathematical Statistics*, 13, 166-178.

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

Tillé, Y. (1999). Estimation in surveys using conditional inclusion probabilities: complex design. *Survey Methodology*, 25, 57-66.