Survey Methodology, December 2006
Vol. 32, No. 2, pp. 151-156
Statistics Canada, Catalogue No. 12-001

Bernoulli Bootstrap for Stratified Multistage Sampling

Fumio Funaoka, Hiroshi Saigo, Randy R. Sitter and Tsutom Toida [1]

Abstract

In this article, we propose a Bernoulli-type bootstrap method that can easily handle multi-stage stratified designs where sampling fractions are large, provided simple random sampling without replacement is used at each stage. The method provides a set of replicate weights which yield consistent variance estimates for both smooth and non-smooth estimators. The method's strength is in its simplicity. It can easily be extended to any number of stages without much complication. The main idea is to either keep or replace a sampling unit at each stage with preassigned probabilities, to construct the bootstrap sample. A limited simulation study is presented to evaluate performance and, as an illustration, we apply the method to the 1997 Japanese National Survey of Prices.

Key Words: Complex survey; Linearization; Quantiles; Resampling; Stratification.

1. Introduction

Many large scale surveys are conducted using a stratified multi-stage sampling design. Variance estimation in this type of design can be analytically involved or even impossible. In addition, for publicly released data sets the specific forms of estimators the end-user may wish to obtain variance estimates for are unknown. As a result, resampling methods are often carried out to obtain a set of replicate weights that can be supplied with the data set and used for the purpose of variance estimation for a broad class of possible estimators. The bootstrap is particularly useful since it can handle both smooth and nonsmooth sample statistics under multistage designs. A concise summary of several bootstrap methods for finite population sampling is found in Shao and Tu (1995, pages 232–282) (see also, Gross 1980; Bickel and Freedman 1984; McCarthy and Snowden 1985; Rao and Wu 1988; Kovar, Rao and Wu 1988; Sitter 1992a, b; Booth, Butler and Hall 1994; Shao and Sitter 1996).

If the first-stage sampling fraction is small, there are various bootstrap methods available that treat the first-stage sampling as having been with-replacement for the purposes of variance estimation. In the case where the first-stage sampling fraction is not negligible, there are fewer results available. For bootstrapping in two-stage sampling with simple random sampling (SRS) at each stage see Sitter (1992a, 1992b) and with unequal probabilities Rao and Wu (1988). However, if the first-stage sampling fractions are not negligible no simple bootstrap procedure is available for three or more stages of sampling. In this paper, we propose a new bootstrap method which easily accommodates such cases when the sampling is simple random sampling (SRS) at each stage. We call it a Bernoulli bootstrap (BBE) because of its resemblance to Bernoulli sampling. The National Survey of Prices (NSP) in Japan is used for illustration.

The paper is organized as follows. Section 2 introduces notation for three-stage stratified sampling. Section 3 describes two types of BBE. Section 4 investigates properties of the methods via simulation. Section 5 describes the sampling design of the 1997 NSP and illustrates the use of BBE on the NSP data. Concluding remarks are made in section 6.

2. Stratified Three-Stage Sampling

In stratified random sampling, the finite population, consisting of N primary sampling units (PSU's), is partitioned into H nonoverlapping *strata* of $N_1, N_2, \ldots, N_H$ PSU's, respectively; thus, $\sum_{h=1}^{H} N_h = N$. A simple random sample without replacement (SRSWOR) of PSU's is taken independently from each stratum. The sample sizes within each stratum are denoted by $n_1, n_2, \ldots, n_H$, and the total PSU sample size is $n = \sum_{h=1}^{H} n_h$. At the second stage, a sample of m_{hi} secondary sampling units (SSU's) are selected from PSU i of size M_{hi} within stratum h by SRSWOR. At the third stage, a sample of l_{hij} ultimate sampling units (USU's) are selected from SSU ij of size L_{hij} within stratum h by SRSWOR. A vector of measurements of some unit characteristics is represented as $\mathbf{y}_{hijk} = (y_{1hijk}, y_{2hijk}, y_{\tau hijk})^T$, where the subscripts $hijk$ refer to the stratum label, PSU label, SSU label and USU label, respectively. The population parameter of interest $\theta = \theta(\mathrm{S})$, where $\mathrm{S} = \{\mathbf{y}_{hijk} : h = 1, 2, \ldots, H; i = 1, 2, \ldots, N_h;$

1. F. Funaoka, Professor, Faculty of Economics, Shinshu University, 3-1-1 Asahi, Matsumoto, Nagano, 390-8621, Japan; H. Saigo, Professor, School of Political Science and Economics, Waseda University, 1-6-1 Nishiwaseda Shinjuku, Tokyo, 169-8050, Japan; R.R. Sitter, Professor, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada; T. Toida, Associate Professor, Faculty of Social and Information Studies, Gunma University, 2-4 Aramakicho, Maebashi, Gunma 371-8510, Japan.

$j = 1, \ldots, M_{hi}; k = 1, \ldots, L_{hij}\}$, is usually estimated by $\hat{\theta} = \hat{\theta}(s)$, where $s = \{\mathbf{y}_{hijk} : h = 1, 2, \ldots, H; i = 1, 2, \ldots, n_h; j = 1, \ldots, m_{hi}; k = 1, \ldots, l_{hij}\}$. The population total vector is denoted $\mathbf{Y} = (Y_1, \ldots, Y_\tau)^T$. In this case, its unbiased estimate is

$$\hat{\mathbf{Y}} = \sum_{h=1}^{H} \hat{\mathbf{Y}}_{\mathbf{h}} = \sum_{h=1}^{H} (N_h / n_h) \sum_{i=1}^{n_h} \hat{\mathbf{Y}}_{\mathbf{hi}},$$

where $\hat{\mathbf{Y}}_{\mathbf{hi}} = (M_{hi} / m_{hi}) \sum_{j=1}^{m_{hi}} \hat{\mathbf{Y}}_{\mathbf{hij}}$ and $\hat{\mathbf{Y}}_{\mathbf{hij}} = (L_{hij} / l_{hij}) \sum_{k=1}^{l_{hij}} \mathbf{y}_{\mathbf{hijk}}$. This may be written as $\hat{\mathbf{Y}} = \sum_{hijk} w_{hij} \mathbf{y}_{\mathbf{hijk}}$, where $w_{hij} = (N_h / n_h)(M_{hi} / m_{hi})(L_{hij} / l_{hij})$.

For $\tau = 1$, an unbiased estimate of $\mathrm{Var}(\hat{Y})$ is $v(\hat{Y}) = \sum_{h=1}^{H} v(\hat{Y}_h)$, where

$$v(\hat{Y}_h) = \frac{N_h^2 (1 - f_{1h}) s_h^2}{n_h} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}^2 (1 - f_{2hi}) s_{hi}^2}{m_{hi}} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} \frac{L_{hij}^2 (1 - f_{3hij}) s_{hij}^2}{l_{hij}}$$

with $\bar{\hat{Y}}_{h\cdot} = n_h^{-1} \sum_i \hat{Y}_{hi}$, $\bar{\hat{Y}}_{hi\cdot} = m_{hi}^{-1} \sum_j \hat{Y}_{hij}$, $\bar{y}_{hij\cdot} = l_{hij}^{-1} \sum_k y_{hijk}$, $f_{1h} = n_h / N_h$, $f_{2hi} = m_{hi} / M_{hi}$, $f_{3hij} = l_{hij} / L_{hij}$, $s_h^2 = \sum_i (\hat{Y}_{hi} - \bar{\hat{Y}}_{h\cdot})^2 / (n_h - 1)$, $s_{hi}^2 = \sum_j (\hat{Y}_{hij} - \bar{\hat{Y}}_{hi\cdot})^2 / (m_{hi} - 1)$, and $s_{hij}^2 = \sum_k (y_{hijk} - \bar{y}_{hij\cdot})^2 / (l_{hij} - 1)$ (Särndal, Swensson and Wretman 1992, pages 148–149).

3. Proposed Bernoulli Bootstrap

To handle the multi-stage aspect of the sampling within stratum, we propose a multi-stage bootstrap. To simplify ideas, we first introduce a simple version that has some limitations in applicability. We will then subsequently describe a more general form that avoids these difficulties.

A Short Cut BBE

Step I. For each sample PSU, hi, within stratum $h, h = 1, \ldots, H$, we: (a) keep it in the bootstrap sample with probability

$$p_h = \sqrt{1 - \frac{(1 - f_{1h})}{(1 - n_h^{-1})}}; \qquad (3.1)$$

or (b) replace it with one selected randomly from the n_h PSU's. If (a) is the case, go to Step II.

Step II. For each SSU hij in PSU hi of stratum h kept at Step I, we: (c) keep it in a bootstrap sample with probability

$$q_{hi} = \sqrt{1 - \frac{f_{1h}}{p_h^{-1}} \frac{(1 - f_{2hi})}{(1 - m_{hi}^{-1})}}; \qquad (3.2)$$

or (d) replace it with one selected randomly from the m_{hi} SSU's in PSU hi of stratum h. If (c) is the case, go to Step III.

Step III. For each USU $hijk$ in SSU hij in PSU hi of stratum h, we: (e) keep it in the bootstrap sample with probability

$$r_{hij} = \sqrt{1 - \frac{f_{1h}}{p_h^{-1}} \frac{f_{2hi}}{q_{hi}^{-1}} \frac{(1 - f_{3hij})}{(1 - l_{hij}^{-1})}}; \qquad (3.3)$$

or (f) replace it with one randomly selected from the l_{hij} USU's in SSU hij in PSU hi of stratum h.

If we let K_{hij}^* denote the number of times unit $hijk$ appears in the bootstrap resample, then the bootstrap estimate of the total is $\hat{\mathbf{Y}}^* = \sum_{hijk} w_{hij}^* \mathbf{y}_{\mathbf{hijk}}$, where $w_{hij}^* = K_{hij}^* w_{hij}$, and the bootstrap estimate of $V(\hat{\theta})$ is $v_B(\hat{\theta}) = V_*(\hat{\theta}^*)$, where $\hat{\theta}^* = \theta(\hat{\mathbf{Y}}^*)$ and V_* represents the variance under the resampling procedure. Typically, the bootstrap estimate of variance is obtained by Monte Carlo simulation. That is, repeat Steps I–III a large number of times, B, to get $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ and use

$$v_B(\hat{\theta}) = \sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\theta}_{(\cdot)}^*)^2 / B,$$

where $\bar{\theta}_{(\cdot)}^* = \sum_{b=1}^{B} \hat{\theta}^* / B$. In most cases one can replace $\bar{\theta}_{(\cdot)}^*$ by $\hat{\theta}$. This allows the survey methodologist to create a set of replicate weights w_{hij}^* for each bootstrap resample and release these with the data released to the public.

Obviously, the short cut BBE is feasible only when $p_h, q_{hi}, r_{hij} \in [0, 1] \; \forall \; h, i, j$. For instance, it is necessary that $f_{1h} \geq n_h^{-1}$. To handle arbitrary $n_h, m_{hi}, l_{hij} \geq 2$, we may modify each step and change p_h, q_{hi}, r_{hij} accordingly:

A General BBE

Step I′. Choose $(n_h - 1)$ PSU's by SRS with replacement from n_h PSU's in the sample, $h = 1, \ldots, H$. Denote the candidate set by $\{\mathrm{P\tilde{S}U}_{hi} : i = 1, 2, \ldots, n_h - 1\}$. For each PSU i in the sample in stratum h, we: (a) keep it in the bootstrap sample with probability

$$p_h = 1 - \frac{1}{2} \frac{(1 - f_{1h})}{(1 - n_h^{-1})}; \qquad (3.4)$$

or (b) replace it with one selected randomly from $\{\mathrm{P\tilde{S}U}_{hi} : i = 1, 2, \ldots, n_h - 1\}$. If (a) is the case, go to Step II′.

Step II′. For hi kept at Step **I′**, choose $(m_{hi} - 1)$ SSU's by SRS with replacement from m_{hi} SSU's in PSU hi. Denote the candidate set by $\{\mathrm{S\tilde{S}U}_{hij} : j = 1, 2, \ldots, m_{hi} - 1\}$. For each SSU

hij in PSU hi kept at Step $\mathbf{I}'$, we: (c) keep it in the bootstrap sample with probability

$$q_{hi} = 1 - \frac{1}{2} \frac{f_{1h}}{p_h^{-1}} \frac{(1 - f_{2hi})}{(1 - m_{hi}^{-1})}; \qquad (3.5)$$

or (d) replace it with one selected randomly from $\{\tilde{\text{SSU}}_{hij} : j = 1, 2, \ldots, m_{hi} - 1\}$. If (c) is the case, go to Step $\mathbf{III}'$.

Step III′. For hij kept at Step $\mathbf{II}'$, choose $l_{hij} - 1$ USU's by SRS with replacement from l_{hij} USU's in SSU hij in PSU hi. Denote the candidate set by $\{\tilde{\text{USU}}_{hijk} : k = 1, 2, \ldots, l_{hij} - 1\}$. For each USU $hijk$ in SSU hij in PSU hi, we: (e) keep in the bootstrap sample with probability

$$r_{hij} = 1 - \frac{1}{2} \frac{f_{1h}}{p_h^{-1}} \frac{f_{2hi}}{q_{hi}^{-1}} \frac{(1 - f_{3hij})}{(1 - l_{hij}^{-1})}; \qquad (3.6)$$

or (f) replace it with one randomly selected from $\{\tilde{\text{USU}}_{hijk} : k = 1, 2, \ldots, l_{hij} - 1\}$.

It is readily seen that $p_h, q_{hi}, r_{hij} \in [0, 1] \; \forall \; n_h, m_{hi}, l_{hij} \geq 2$.

The reason for randomly selecting a candidate set in the general BBE can be explained as follows. To fix the idea, consider single-stratum one-stage SRSWOR. Let $\bar{y}^*$ be a bootstrap sample mean under the short cut BBE with some arbitrary $p \in [0, 1]$. Then, it can be shown that $V_*(\bar{y}^*) = n^{-1}(1 - n^{-1}) s^2 (1 - p^2)$, where $s^2 = \sum_i (y_i - \bar{y})^2 / (n - 1)$. Note that $V_*(\bar{y}^*)$ is monotone decreasing with respect to p in $[0, 1]$. So, $\min_{p \in [0,1]} V_*(\bar{y}^*) = 0$ and $\max_{p \in [0,1]} V_*(\bar{y}^*) = n^{-1}(1 - n^{-1}) s^2$. If $f_1 < n^{-1}$, then $\max_p V_*(\bar{y}^*) < v(\bar{y})$. The key idea of the general BBE is that we can make $\max_p V_*(\bar{y}^*)$ greater than $v(\bar{y})$ by putting extra variation into unit replacement through randomly selecting a candidate set.

It can be shown that both the short cut BBE and the general BBE provide consistent variance estimation for smooth functions of estimated population totals. Moreover, under appropriate regularity conditions for the population distribution function, they also provide consistent variance estimation for sample quantiles. In addition, both BBE methods use resample sizes equal to the original sample sizes. This can be a desirable property when we deal with imputed survey data (see Saigo, Shao and Sitter 2001).

It is not difficult to extend the BBE approach to designs with more than three stages. For example, for a four stage stratified design, a USU at the fourth stage within stratum h is kept with probability

$$\sqrt{1 - p_h^{-1} f_{1h} q_{hi}^{-1} f_{2hi} r_{hij}^{-1} f_{3hij} (1 - g_{hijk}^{-1})^{-1} (1 - f_{4hijk})}$$

or replaced in the short cut BBE, where g_{hijk} is the fourth stage sample size and f_{4hijk} is the fourth stage sampling fraction. Further extensions are analogous.

The general BBE randomizes a candidate set in order to merely fix infeasibility of the short cut BBE. This idea has similarities to the approximately Bayesian bootstrap of Rubin and Schenker (1986).

A disadvantage of the general BBE versus the short cut BBE is that the former requires, on the average, $\sum_h \{(n_h - 1) + p_h \sum_i (m_{hi} - 1) + p_h \sum_i q_{hi} \sum_j (l_{hij} - 1)\}$ more random number generations than the latter, where p_h, q_{hi}, and r_{hij} are given by (3.4), (3.5), and (3.6), respectively. This may be time-consuming when the sample sizes and/or the number of strata are large. To reduce random number generations in the general BBE, one can create a candidate set by randomly deleting one unit from the original sample and use

$$p_h = (n_h + 1/2) - \sqrt{(n_h + 1/2)^2 - n_h (1 + f_{1h})}, \qquad (3.7)$$

$$q_{hi} = (m_{hi} + 1/2) - \sqrt{(m_{hi} + 1/2)^2 - f_{h1} p_h^{-1} m_{hi} (1 + f_{2hi})}, \; (3.8)$$

$$r_{hij} = (l_{hij} + 1/2) - \sqrt{(l_{hij} + 1/2)^2 - f_{h1} p_h^{-1} f_{2hi} q_{hi}^{-1} l_{hij} (1 + f_{3hij})}, \qquad (3.9)$$

instead. It can be shown that $p_h, q_{hi}, r_{hij} \in [0, 1]$. The proof for this modified version of the general BBE is similar.

4. A Simulation Study

In this section, we perform limited simulations to examine the BBE for ratio estimation and quantile estimation. For simplicity, we consider two-stage SRSWOR and restrict to a single stratum.

4.1 General Description of Simulation

A single-stratum finite population is generated by the following procedure and fixed over all simulation runs to observe design-based properties of the BBE. First, the average of the auxiliary variables in cluster i is generated by $\mu_i \sim N(\mu, \sigma^2)$ for $i = 1, 2, \ldots, N$. Then, the auxiliary variable x_{ik} of unit k in cluster i is generated by

$$x_{ik} = \mu_i + \varepsilon_{ik} (k = 1, 2, \ldots, M_i; i = 1, 2, \ldots, N), \qquad (4.1)$$

where $\varepsilon_{ik} \sim N(0, (1 - \rho)\sigma^2 / \rho)$. The target variable y_{ik} of unit k in cluster i is obtained by

$$y_{ik} = a + b x_{ik} + e_{ik} (k = 1, 2, \ldots, M_i; i = 1, 2, \ldots, N), \quad (4.2)$$

where $e_{ik} \sim N(0, \sigma^2 / 4)$. The parameter values used are $\mu = 100, \sigma = 10, \rho = 0.1 (0.3), a = 0$, and $b = 1$, and two-stage SRSWOR is used throughout the simulation study.

4.2 Ratio Estimation

Let $N=50, n=15, M_i=20$ and $m_i=3$, for $i=1,\ldots,n$. Consider the ratio estimator of the population total, Y,

$$\hat{Y}_R = \hat{R}X,$$

where $X=\sum_{i=1}^{N}\sum_{k=1}^{M_i} x_{ik}$ is the population total of the x's $\hat{R}=\hat{Y}/\hat{X}$, $\hat{Y}=\sum_{h=1}^{H}\hat{Y}_h=\sum_{h=1}^{H}(N_h/n_h)\sum_{i=1}^{n_h}\hat{Y}_{hi}$, $\hat{X}=\sum_{h=1}^{H}\hat{X}_h=\sum_{h=1}^{H}(N_h/n_h)\sum_{i=1}^{n_h}\hat{X}_{hi}$, $\hat{Y}_{hi}=(M_{hi}/m_{hi})\sum_{k=1}^{m_{hi}}\hat{Y}_{hik}$ and $\hat{X}_{hi}=(M_{hi}/m_{hi})\sum_{k=1}^{m_{hi}}\hat{X}_{hik}$.

For the purpose of comparison, we consider a number of alternate variance estimators that are available in this simple context:

1) The conventional variance estimator is denoted

$$v_0(\hat{Y}_R) = N^2\frac{1-f_1}{n}\frac{\sum_i(\hat{Y}_i-\hat{R}\hat{X}_i)^2}{n-1} + \frac{N}{n}\sum_i\frac{M_i^2(1-f_{2i})s_{d'2i}^2}{m_i}, \quad (4.3)$$

where $f_1=n/N$, $f_{2i}=m_i/M_i$ and

$$s_{d'2i}^2=\sum_j(y_{ij}-\hat{R}x_{ij})^2/(m_i-1).$$

2) The delete 1 PSU at a time jackknife corrected for the first-stage sampling fraction is sometimes used, even though it is not entirely correct,

$$v_{cj}(\hat{Y}_R)=(1-f_1)\frac{n-1}{n}\sum_i(\hat{Y}_{R(i)}-\hat{Y}_{R(\cdot)})^2, \quad (4.4)$$

where $\hat{Y}_{R(i)}$ is the estimator recalculated with the i^{th} PSU removed and $\hat{Y}_{R(\cdot)}=\sum_i\hat{Y}_{R(i)}/n$.

3) An externally weighted jackknife (see Folsom, Bayless and Shah 1971) can be derived that corrects for both stages of sampling as

$$v_{ewj}(\hat{Y}_R)=(1-f_1)\frac{n-1}{n}\sum_i(\hat{Y}_{R(i)}-\hat{Y}_{R(\cdot)})^2 + f_1\sum_i(1-f_{2i})\frac{m_i-1}{m_i}\sum_j(\hat{Y}_{R(ij)}-\hat{Y}_{R(\cdot\cdot)})^2, \quad (4.5)$$

where $\hat{Y}_{R(i)}$ is the i^{th} jackknife pseudo value by deleting PSU i, $\hat{Y}_{R(ij)}$ is the ij^{th} jackknife pseudo value by deleting unit j in PSU i, $\hat{Y}_{R(\cdot)}=\sum_i\hat{Y}_{R(i)}/n$, and $\hat{Y}_{R(\cdot\cdot)}=\sum_j\hat{Y}_{R(ij)}/m_i$.

4) A model-assisted variance estimator is also available (see Särndal, Swensson and Wretman (1992), equation (8.10.6)),

$$v_{ma}(\hat{Y}_R)=(X/\hat{X})^2v_0(\hat{Y}_R). \quad (4.6)$$

We use $B=100$ bootstrap resamples in each of $S=1{,}000$ simulation runs. The true MSE's are approximated by 10,000 simulation runs and we use Monte Carlo estimates of percent relative bias and coefficient of variation of the various variance estimators as measures of their relative performance, as well as, empirical coverage probabilities of 90% confidence intervals.

We see in Table 1 that $v_{\text{BBE}}, v_0, v_{ewj}$ and v_{ma} perform comparably and well, except that the CV of the resampling methods are a bit higher than the non-resampling methods, as is typical. The delete 1 PSU at a time jackknife performs poorly.

To investigate the conditional properties, we ordered the 1,000 simulation runs on $X/\hat{X}$ and divided the runs into 20 equally sized groups. For each group the average of each variance estimator is calculated. Figure 1 plots these grouped averages for each variance estimator (excluding v_{cj} since it has large negative bias) versus the grouped average $X/\hat{X}$, for $\rho=0.3$. The true MSE is included in the plot, as well. This is a similar plot to that used by Royall and Cumberland (1981a, 1981b). One can see that v_{BBE} tracks the true MSE much like v_{ewj} and v_{ma}, whereas v_0 does not. Thus, the BBE seems to have a desirable conditional property.

Table 1
Comparison of Variance Estimators for $\hat{Y}_r$

ρ		% Bias	CV	Coverage (90%)
0.1	v_0	−1.70	0.28	89.2
	v_{BBE}	−0.62	0.33	88.9
	v_{ewj}	−0.33	0.30	89.4
	v_{cj}	−26.55	0.39	80.5
	v_{ma}	−0.39	0.30	89.4
0.3	v_0	−0.67	0.28	86.6
	v_{BBE}	−1.63	0.33	86.5
	v_{ewj}	−0.74	0.29	86.5
	v_{cj}	−26.85	0.39	80.2
	v_{ma}	−0.87	0.29	86.4

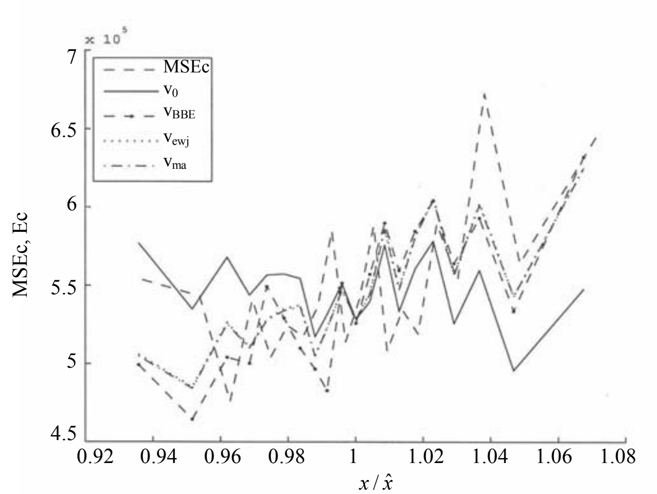

Figure 1. MSEc and Ec(v) for the ratio estimation.

4.3 Quantile Estimation

For quantile estimation, we set $N = 100, n = 30,$ $M_i = 100$ and $m_i = 10,$ for $i = 1, \ldots, n.$ We use $B = 500$ bootstrap resamples in each of $S = 5,000$ simulation runs. The true MSE's are approximated by 50,000 simulation runs. Only the results for v_{BBE} and v_{ewj} when $\rho = 0.1$ are summarized in Table 2 because those when $\rho = 0.3$ are similar. We see that the BBE method performs quite well, with a slight upward bias, while the externally weighted jackknife method has serious bias because of its inconsistency in variance estimation for quantiles.

Table 2
Performance of v_{BBE} and v_{ewj} for the 0.10, 0.25, 0.50, 0.75 and 0.90 quantiles

	v_{BBE}			v_{ewj}		
Quantile	%Bias	CV	Coverage (90%)	%Bias	CV	Coverage (90%)
0.10	8.40	0.51	87.7	51.87	1.93	81.3
0.25	6.21	0.42	88.2	21.19	1.28	83.3
0.50	2.53	0.37	87.4	14.27	1.00	83.0
0.75	6.23	0.42	87.8	28.07	1.33	83.4
0.90	6.32	0.50	88.0	54.47	2.05	80.3

5. Application to the 1997 National Survey of Prices in Japan

The objective of the NSP is to analyze price formations of major consumers' goods, such as food, clothes and home appliances. To this end, quantile estimation plays a central role, and many quantile estimates based on several post-stratifications are included in the NSP reports.

The stratified multistage sampling used in NSP 1997 is summarized as follows:

Stratification. Municipalities form the PSU's and are stratified into 537 strata, first according to prefectures and economic sphere that each municipality forms and then further by their population sizes.

First Stage Sampling. These PSU's are selected via SRSWOR independently within each stratum. An overview of the first-stage sampling fractions is given in Table 3.

Second Stage Sampling. In a selected municipality, all the large scale outlets are enumerated. In other words, single stage cluster sampling is employed for large scale outlets. For small scale outlets, on the other hand, a sampled municipality is divided into survey areas (SSU's) each consisting of about 100 outlets. Systematic sampling is used to sample survey areas. The sampling fractions at the second stage are between 0.1 and 1.0.

Third Stage Sampling. In each selected survey area, 40 outlets (USU's) are chosen by ordered systematic sampling with respect to the types of outlets and the annual sales reported in the 1994 Census of Commerce.

Strictly speaking, there is no valid variance formula for the NSP data because it contains systematic sampling. For estimating variance, however, it is assumed that systematic sampling can be approximated by SRSWOR. Even under this simplified condition, there is no closed variance formula for sample quantiles. In fact, no variance estimates are associated with estimated price quantiles in the NSP report, while the average prices are reported with their variance estimates.

In this section, we apply the short cut BBE to the NSP data, assuming that systematic sampling can be approximated by SRSWOR. Some strata have only one PSU. In addition, $f_{1h} < n_h^{-1}$ in some strata. Such strata are grouped into adjacent strata so that p_h given by (3.1) is in $[0, 1].$ After grouping, there are more than 280 strata. The effect of reforming strata is assumed to be negligible.

Table 3
The First Stage Sampling Fractions in NSP 1997

Area Category	Population Size	# of PSU's	Sampling Fraction	Sample Size
Cities	$\geq 100{,}000$	221	1/1	221
Cities	$50{,}000 - 99{,}999$	220	2/3	179
Cities	$< 50{,}000$	224	1/3	80
Towns and villages	$\geq 40{,}000$	32	1/5	4
Towns and villages	$< 40{,}000$	2,536	1/15	187

After reforming strata, the short cut BBE is employed in those strata composed by cities. On the other hand, the with-replacement bootstrap (Shao and Tu 1995, page 247) using resample size $(n_h - 1)$ is used in those composed of towns and villages, where the first stage sampling fractions are small. The quantile estimates and their standard errors for selected commodities in small-sized outlets are shown in Table 4. Note that the prices of a given commodity are discrete. However, we apply the bootstrap as if prices of commodities are continuous. This approximation should be acceptable for many commodities, but not for very inexpensive ones, since in such a case, a large percentage of observations concentrate on a specific price and the estimated standard error can be 0.

Table 4
Sample Quantiles (Standard Errors) of Selected Commodities for Small Outlets in NSP

Commodity	p	0.10	0.25	0.5	0.75	0.90
Rice (5kg)[a]	Sample quantile	239.4	255.2	278.3	299.1	315.0
(10 yens)	(standard error)	(0.24)	(0.53)	(0.21)	(0.02)	(0.61)
Instant Coffee (1 bottle)[b]	Sample quantile	714	788	859	893	914
(yen)	(standard error)	(0.13)	(0.40)	(0.00)	(2.68)	(1.43)
Beer (24 cans)[c]	Sample quantile	467.3	500.0	536.8	549.4	549.4
(10 yens)	(standard error)	(1.01)	(0.64)	(0.82)	(0.00)	(0.00)
PC[d]	Sample quantile	248.8	260.4	299.3	346.5	375.9
(1,000 yens)	(standard error)	(2.03)	(0.35)	(3.25)	(7.17)	(1.48)

The specified brands [a]Koshihikari; [b]Nescafe Gold Blend, 100g; [c]Sapporo (Nama) Black Label, 350ml; [d]NEC PC9821 NW133/D14.

6. Conclusions

The bootstrap is useful for estimating variances in complex surveys, particularly when quantile estimation is important. We have proposed two Bernoulli-type bootstrap methods that can easily handle multi-stage stratified SRSWOR designs where sampling fractions are large: the short cut BBE and the general BBE. In both methods, a sampling unit at a given stage is either kept or replaced with preassigned probabilities to construct a bootstrap sample. The general BBE has an advantage in that it can handle any combination of sample sizes ≥ 2 although it requires more random number generations than the short cut BBE. As an illustration, we applied the short cut BBE to Japanese 1997 National Survey of Prices data.

Acknowledgements

The second author was supported by the Japan Statistical Association. The third author was supported by a grant from the Natural Science and Engineering Research Council of Canada. The authors thank the Statistics Bureau, Ministry of Public Management, Home Affairs, Posts and Telecommunications, and the Ministry of Economy, Trade, and Industry, Japan, for providing the 1997 NSP data.

References

Bickel, P.J., and Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statististics*, 12, 470-482.

Booth, J.G., Butler, R.W. and Hall, P. (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association*, 89, 1282-1289.

Folsom, R.E., Bayless, D.L. and Shah, B.V. (1971). Jackknifing for variance components in complex sample survey designs. *Proceedings of the Social Statistics Section*, American Statistical Association, 36-39.

Gross, S. (1980). Median estimation in sample surveys. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 181-184.

Kovar, J.G., Rao, J.N.K. and Wu, C.F.J. (1988). Bootstrap and other methods to measure errors in survey estimates. *Canadian Journal of Statististics*, 16, Supplement, 25-45.

McCarthy, P.J., and Snowden, C.B. (1985). The bootstrap and finite population sampling. *Vital and Health Statistics*, Serie 2, 95, Public Health Service Publication, 85-1369, Washington, DC: U.S. Government Printing Office.

Rao, J.N.K., and Wu, C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.

Royall, R.M., and Cumberland, W.G. (1981a). An empirical study of the ratio estimator and estimators of its variance. *Journal of the American Statistical Association*, 76, 66-77.

Royall, R.M., and Cumberland, W.G. (1981b). The finite population linear regression estimator: An empirical study. *Journal of the American Statistical Association*, 76, 924-930.

Rubin, D.B., and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statististical Association*, 81, 366-374.

Saigo, H., Shao, J. and Sitter, R.R. (2001). A repeated half-sample bootstrap and balanced repeated replication for randomly imputed data. *Survey Methodology*, 27, 189-196.

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

Shao, J., and Sitter, R.R. (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association*, 91, 1278-1288.

Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer-Verlag.

Sitter, R.R. (1992a). A resampling procedure for complex survey data. *Journal of the American Statistical Association*, 87, 755-765.

Sitter, R.R. (1992b). Comparing three bootstrap methods for survey data. *Canadian Journal of Statististics*, 20, 135-154.