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

Article

A comparison of sample set restriction procedures

by Jason C. Legg and Cindy L. Yu

June 2010

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2010
Vol. 36, No. 1, pp. 69-79
Statistics Canada, Catalogue No. 12-001-X

A comparison of sample set restriction procedures

Jason C. Legg and Cindy L. Yu [1]

Abstract

For many designs, there is a nonzero probability of selecting a sample that provides poor estimates for known quantities. Stratified random sampling reduces the set of such possible samples by fixing the sample size within each stratum. However, undesirable samples are still possible with stratification. Rejective sampling removes poor performing samples by only retaining a sample if specified functions of sample estimates are within a tolerance of known values. The resulting samples are often said to be balanced on the function of the variables used in the rejection procedure. We provide modifications to the rejection procedure of Fuller (2009a) that allow more flexibility on the rejection rules. Through simulation, we compare estimation properties of a rejective sampling procedure to those of cube sampling.

Key Words: Rejection sampling; Cube sampling; Stratification; Balanced sampling.

1. Introduction

A common practice in survey sampling is to utilize known population information about auxiliary variables to improve estimators of means and totals of characteristics of interest. When population control means or totals for an auxiliary variable are known, regression and other calibration estimators are often utilized. Let $(\boldsymbol{x}_i, y_i, p_i)$, $i = 1, 2, ..., N$, be a sequence of real vectors, where each $\boldsymbol{x}_i$ is a k dimensional vector, and a sample A be selected from $F_N = [(\boldsymbol{x}_1, y_1, p_1), ..., (\boldsymbol{x}_N, y_N, p_N)]$ using a sample design with inclusion probabilities p_i and joint inclusion probabilities p_{ij}. Suppose the population mean of $\boldsymbol{x}_i$, $\overline{\boldsymbol{x}}_N$, is known. Consider the regression estimator of the population mean of the form

$$\overline{y}_{\text{reg}} = \overline{\boldsymbol{z}}_N' \hat{\boldsymbol{\beta}}, \tag{1}$$

where $\boldsymbol{z}_i$ contains design variables and $\boldsymbol{x}_i$, $\overline{\boldsymbol{z}}_N$ is the population mean of $\boldsymbol{z}_i$, and $\hat{\boldsymbol{\beta}}$ is a regression coefficient estimator. For many designs, $\hat{\boldsymbol{\beta}}$ of the form

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i \in A} \boldsymbol{z}_i \, \phi_i \, p_i^{-2} \, \boldsymbol{z}_i' \right)^{-1} \sum_{i \in A} \boldsymbol{z}_i \, \phi_i \, p_i^{-2} \, y_i, \tag{2}$$

where ϕ_i are constants determined by the design, will be asymptotically efficient. Some examples of ϕ_i choices are $\phi_i = (1 - p_i)$ for Poisson sampling and for stratified random sampling, $\phi_{hi} = (N_h - 1)^{-1}(N_h - n_h)$ for element i in stratum h. If we assume there is a vector $\boldsymbol{d}$ such that

$$\phi_i \, p_i^{-2} \, \boldsymbol{z}_i' \boldsymbol{d} = p_i^{-1} \tag{3}$$

for all i, then estimator (1) is design consistent (Fuller 2002). The regression coefficient estimator (2) converges together with

$$\boldsymbol{\beta}_N = \left(\sum_{i=1}^{N} \boldsymbol{z}_i \, \phi_i \, p_i^{-1} \, \boldsymbol{z}_i' \right)^{-1} \sum_{i=1}^{N} \boldsymbol{z}_i \, \phi_i \, p_i^{-1} \, y_i.$$

As an example of applying equation (3), suppose we plan to select a Poisson sample and want to regress on a single covariate x_{1i} through the origin. If we add $(1 - p_i)^{-1} p_i$ into $\boldsymbol{z}_i$ to make $\boldsymbol{z}_i' = (x_{1i}, [1 - p_i]^{-1} p_i)$, then (1) will be design consistent for $\overline{y}_N$ since (3) is satisfied by setting $\boldsymbol{d}' = (0, 1)$. If we further assume that a column of ones is in the column space of the regression variables $\boldsymbol{z}_i$, then for these ϕ_i values, estimator (1) nearly attains the minimum asymptotic variance for design consistent regression estimators under certain regularity conditions (Rao 1994). An alternative approach to constructing a regression estimator is to start with a design consistent estimator, such as the generalized regression estimator of Särndal (1980), and determine the best coefficient given that form of the estimator. Starting with a design consistent form removes the need to satisfy (3). Condition (3) allows estimator (1) to be expressed in the form of a generalized regression estimator (Fuller 2009b, pages 116-117).

When auxiliary information is known at the unit level, the auxiliary information can also be incorporated into the sample design. For example in one classic case, the model with

$$y_i = \beta_0 + \beta_1 x_i + x_i \, \varepsilon_i, \tag{4}$$

$\varepsilon_i \sim \text{ind}(0, \sigma^2)$ and $\text{cov}(\varepsilon_i, x_i) = 0$ is assumed for the population F_N. From Isaki and Fuller (1982), the optimal inclusion probabilities for the regression estimator are those that are proportional to the square root of the design variances, *i.e.*, $p_i \propto x_i$ in this case. A possible sampling procedure is Poisson sampling with inclusion probabilities

1. Cindy L. Yu is an assistant professor in the Department of Statistics and the Center for Survey Statistics and Methodology at Iowa State University, Ames, IA 50010. E-mail: cindyyu@iastate.edu; Jason C. Legg is a postdoctoral researcher at the Center for Survey Statistics and Methodology at Iowa State University, Ames, IA 50010. E-mail: jason-legg@hotmail.com.

$$p_i = \left(\sum_{i=1}^{N} x_i \right)^{-1} n_N \, x_i, \tag{5}$$

where $n_N = \sum_{i=1}^{N} p_i$ is a specified target sample size. A second common design when model (4) is assumed is to stratify the population based on x. Strata are determined by setting the boundaries such that the sum of the sorted x_i values in each stratum are approximately equal. An equal number of units in each stratum are selected. This stratification design has the inclusion probabilities close to (5), and was shown to have an anticipated variance close to the best purposive sample model variance in the two-per-stratum case (Fuller 1981).

Another way to incorporate information from an auxiliary variable into the design is balancing. A sample A is balanced for variable $\boldsymbol{z}$ if

$$\overline{\boldsymbol{z}}_{\mathrm{HT}} = N^{-1} \sum_{i \in A} p_i^{-1} \boldsymbol{z}_i = N^{-1} \sum_{i=1}^{N} \boldsymbol{z}_i = \overline{\boldsymbol{z}}_N. \tag{6}$$

A design is balanced for $\boldsymbol{z}$ if every sample with positive probability is balanced for $\boldsymbol{z}$. Balancing can be thought of as calibration by design. To illustrate the effect of balancing, consider an equal inclusion probability design and $\boldsymbol{z}_i = (1, \, x_i)'$. The conditional prediction variance of $\overline{y}_{\mathrm{reg}}$ under model (4) is

$$V(\overline{y}_{\mathrm{reg}} - \overline{y}_N \,|\, \boldsymbol{x}, \, \overline{x}_{\mathrm{HT}}) = E\{V(\overline{u}_{\mathrm{HT}} \,|\, F_N) \,|\, \boldsymbol{x}, \, \overline{x}_{\mathrm{HT}}\} + (\overline{x}_N - \overline{x}_{\mathrm{HT}})^2 V(\hat{\beta}_1 \,|\, \boldsymbol{x}, \, \overline{x}_{\mathrm{HT}}), \tag{7}$$

where $u_i = x_i \, \varepsilon_i$. For a balanced design, the second term in (7) is 0, which suggests we might improve the estimator by balancing on x. In practice, a combination of balancing and calibration will often outperform either technique used alone.

Balanced sample designs have some additional practical value. For many designs, there is a nonzero probability of selecting a sample that contains undesirable auxiliary variable values. For example, an undesirable sample could be a sample with insufficient sample allocation for domains or a sample with a large number of extreme values of auxiliary variables. Although stratified designs reduce the set of such possible samples by fixing the sample size within each stratum, undesirable samples could still be possible. For example, some stratified samples might have some negative weights from using regression estimators. Balancing can remove poor performing samples by only retaining samples with estimates close to known quantities and with only positive weights for regression estimators.

Balanced sampling was proposed by Royall and Cumberland (1981) as a way to reduce model bias from incorrectly specified polynomial superpopulation models. Valliant, Dorfman and Royall (2000) discuss the implications of balancing from a prediction approach to sampling. Deville and Tillé (2004) investigated methods of selecting balanced samples within the design-based framework described above. See also Tillé (2006 Chapter 8) for a detailed treatment of balancing. In practice, finding a perfectly balanced design may not be possible. Very tight balancing can lead to a design with some extreme joint inclusion probabilities, including zero inclusion probabilities. Therefore, partial balancing is done in practice.

In this paper, we compare design properties through simulation studies of two balancing procedures, the rejective sampling of Fuller (2009a) and the cube sampling of Tillé (2006). We also provide modifications to Fuller's rejective sampling procedure that allow for more flexibility in balancing. In Section 2, the rejective sampling and the cube sampling are described. Properties of the inclusion probabilities of the two balancing procedures are compared in Section 3. In Section 4, some simulation results using balanced samples are presented. In Section 5, we provide adjustments to the rejective procedure. Concluding remarks are made in Section 6.

2. Balanced sampling procedures

Rejection sampling involves discarding any sample that does not meet a specified balancing tolerance. Fuller (2009a) presents one condition for rejecting a sample and Royall and Herson (1973) give another. In Fuller's procedure with the balancing variable vector $\boldsymbol{z}$, a sample is selected under a specified initial design and retained if

$$(\overline{\boldsymbol{z}}_{\mathrm{HT}} - \overline{\boldsymbol{z}}_N)' \, [V(\overline{\boldsymbol{z}}_{\mathrm{HT}} \,|\, F_N)]^{-1} \, (\overline{\boldsymbol{z}}_{\mathrm{HT}} - \overline{\boldsymbol{z}}_N) < \gamma \tag{8}$$

for some constant $\gamma > 0$, where $\overline{\boldsymbol{z}}_{\mathrm{HT}}$ is the Horvitz-Thompson mean estimator for variable $\boldsymbol{z}$, F_N is the given finite population,

$$V(\overline{\boldsymbol{z}}_{\mathrm{HT}} \,|\, F_N) = N^{-2} \sum_{i=1}^{N} \sum_{j=1}^{N} (p_{ij} - p_i \, p_j) \, \boldsymbol{z}_i \, \boldsymbol{z}_i' \, p_i^{-1} p_j^{-1},$$

p_i is the inclusion probability for unit i and p_{ij} is the joint inclusion probability of unit i and unit j under the initial design. Otherwise, the sample is rejected, a new sample is selected under the initial design, and condition (8) is checked for the new sample. If the original design has a central limit theorem, the left side of (8) is asymptotically a χ^2 random variable with degrees of freedom equal to the number of auxiliary variables. An approximate rejection rate can be set using the quantiles of a χ^2 distribution for γ. Choice of a rejection rate will depend on objectives of each individual survey. Low rejection rates may not reduce the variance by a large amount, but provide sufficient comfort to a researcher that a very poor sample will not be selected. On the other hand, high rejection rates could provide large reductions in the variance, but the resulting samples could

have insufficient sample size to accommodate unplanned domain analysis. For example, if a researcher decides to conduct domain analysis on the tail of the distribution of a balancing variable, the joint inclusion probabilities could be small leading to few units in the domain for many samples.

The cube method was developed by Tillé and Deville and is described in Tillé (2006). The cube method attempts to select a balanced sample with predetermined first-order inclusion probabilities. If the first-order inclusion vector does not lead to a balanced design, an additional step of minimizing a cost constraint is used. Unlike the rejection procedure, higher order initial inclusion probabilities are not prespecified. The cost minimization step maintains the specified initial first-order inclusion probabilities.

As a way to understand the cube procedure, Tillé (2006) describes sampling geometrically. The set of all possible samples is defined to be the set of vectors for vertices of an N dimensional unit cube. For example, if $N = 3$, the vertex $(0, 1, 1)$ denotes a sample containing units two and three. Using the balancing equation (6) and desired p_i for $i = 1, ..., N$, a balancing plane is created. Any sample where the balancing plane intersects a vertex of the unit N dimensional cube is a balanced sample. The design is balanced if every point of intersection between the balancing plane and the unit cube is a vertex of the unit cube. The cube sampling procedure begins by selecting a vector on the balancing plane, then a random walk from the initial point to an edge of the unit cube is done. Tillé refers to the random walk step as the flight phase. If the edge point at the end of the random walk is a vertex of the unit cube, the sample is selected. Otherwise, a cost minimization procedure is used to convert the fractional components of the edge vector to integers. The integer components of the edge vector are not changed in the cost minimization step. Tillé refers to the cost minimization step as the landing phase. Rejection sampling with high rejection rates produces results similar to cube sampling.

Other procedures besides rejection and cube sampling can be used to obtain nearly balanced samples. For example, stratification with boundaries determined by the $\boldsymbol{x}$ variables can also introduce some balancing effects to samples (Fuller 1981). Deciding the number of variables to use in the rejection and cube sampling procedures is essentially the same process as deciding how many variables to include in a regression estimator.

Software has been developed for selecting cube samples. For rejection sampling, standard software packages can be used to select a sample and compute (8). A loop needs to be written to complete the procedure. Programs for selecting cube samples have been written for SAS and R. See Rousseau and Tardieu (2004) for SAS and Matei and Tillé (2005) for R, and details of the procedures implemented are addressed in Deville and Tillé (2004). The R program available in the *sampling* library was used in the simulations in this paper. Because the cost minimization step of cube sampling is computationally intensive for more than 20 balancing variables, a variable suppression step is recommended for the landing phase in the programs.

3. Inclusion probabilities

Let π_i be the first-order inclusion probability for unit i and π_{ij} be the joint inclusion probability for unit i and j under a balanced design. Both rejective and cube sampling require initial first-order inclusion probabilities as inputs. The first-order inclusion probabilities are different than the initial values for rejection sampling. For rejection sampling, units closer to the population mean will have a slightly higher inclusion probability than units far from the mean. Cube sampling maintains the first-order inclusion probabilities from the initial specification. That is, for cube sampling $\pi_i = p_i$. Although for rejection sampling $\pi_i \neq p_i$, in general, the estimators considered will still use p_i rather than π_i.

To illustrate differences between initial and final inclusion probabilities, samples of size 20 from a population of 100 units were simulated. The population of x-values was generated as random variables from a standard normal distribution. The rejection procedure used simple random sampling as the initial design and balanced on x. The cube sample procedure used a balancing vector of $\boldsymbol{z}_i = (p_i, x_i)'$, where $p_i = 20/100$ for all i. The inclusion of p_i in the balancing vector for cube sampling was to control the sample size so that the resulting design would be comparable to using an initial design of simple random sample design in the rejection sampling simulation. First-order inclusion probabilities were estimated using a Monte Carlo simulation of size 100,000 (Figure 1). The curve was obtained by nonparametric fitting. An approximate 90% rejection rate was used for the rejection sampling. From rejection sampling theory, first-order inclusion probabilities are approximately a quadratic function of the distance $x_i - \bar{x}_N$ for an equal probability initial sample design (Fuller 2009a). The plot suggests that all first-order inclusion probabilities are 0.2 for the cube sample design. As expected, Figure 1 indicates the cube method maintains the specified first-order inclusion probabilities, but the rejective does not. As a result, the Horvitz-Thompson estimator using the initial inclusion probabilities (p_i) and the rejective samples is biased.

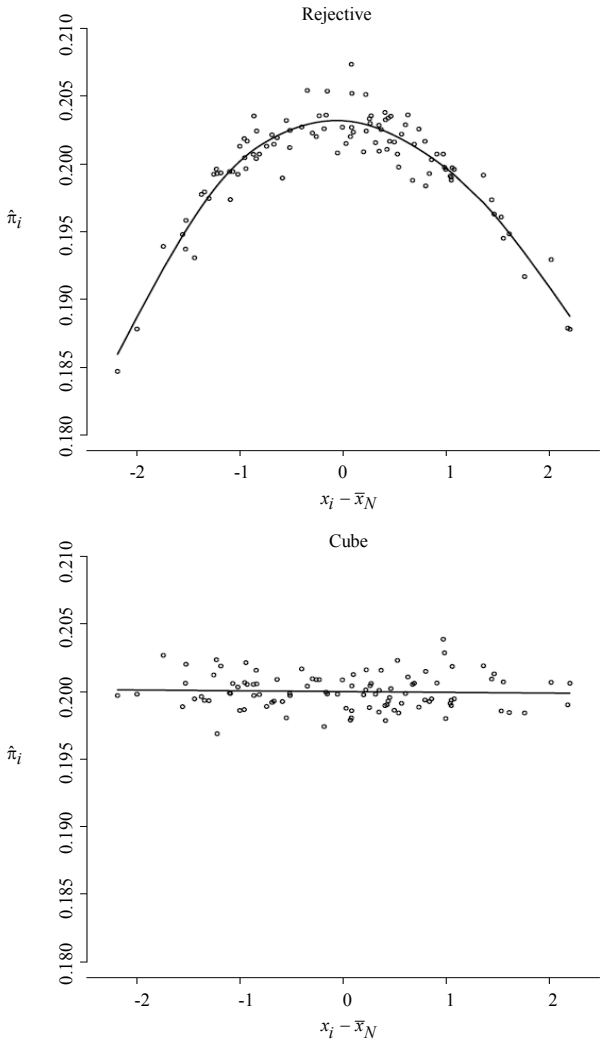

Figure 1 Simulated first-order inclusion probabilities. The balancing variable for the rejective method is $z_i = x_i$, and for the cube method is $z_i = (p_i, x_i)'$, where $p_i = 20/100$

The joint inclusion probabilities for the rejection sampling procedure differ from those of the initial design. A pair of units i and j are likely to have a high joint inclusion probability if $x_i + x_j - 2\bar{x}_N$ is close to zero for an equal probability initial sample design. The joint inclusion probabilities were estimated from simulated samples of size 20 from 100 (Figure 2). The joint inclusion probability for simple random sampling is 0.038. The rejection sampling joint inclusion probabilities are approximately a quadratic function of $x_i + x_j$. The plot of cube sampling joint inclusion probabilities against $x_i + x_j$ appears to have sharper angles than the rejection joint inclusion probabilities. High joint inclusion probabilities for the cube design are associated with pairs of units that are on the far opposite sides of $\bar{x}_N$. That is, for the sample value of $x_i + x_j$, those pairs with a large value of $|x_i| + |x_j|$ have a large probability of inclusion (Figure 3).

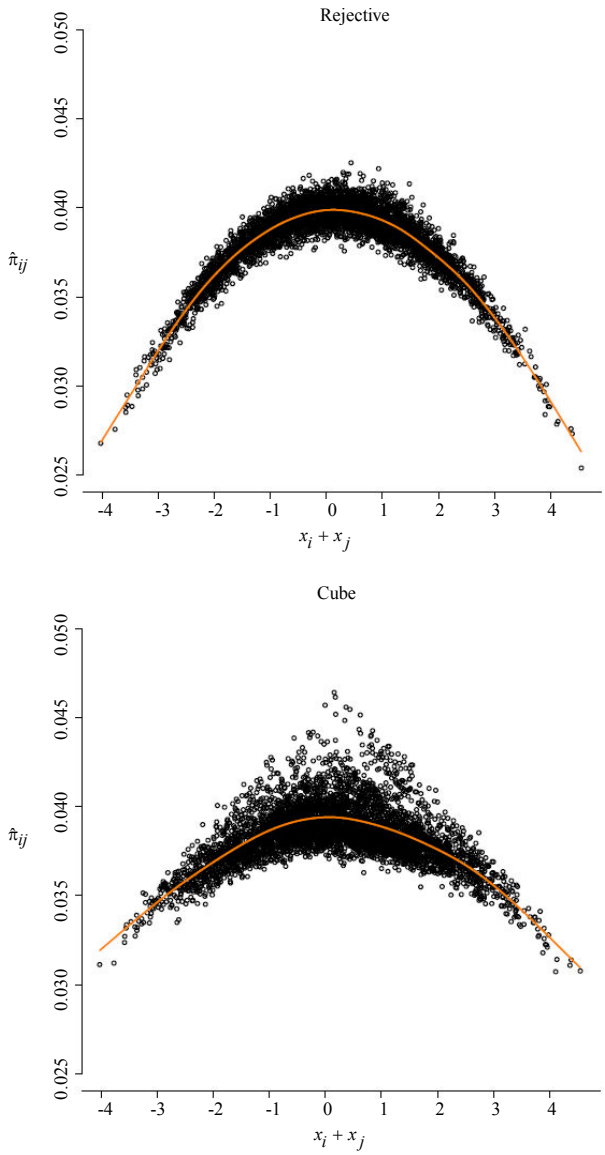

Figure 2 Simulated second-order inclusion probabilities. The balancing variable for the rejective method is $z_i = x_i$, and for the cube method is $z_i = (p_i, x_i)'$, where $p_i = 20/100$

The Horvitz-Thompson estimator using the initial inclusion probabilities under rejection sampling has an $O_p(n^{-1})$ bias while the Horvitz-Thompson estimator under cube sampling is unbiased. The standard Horvitz-Thompson variance estimator is biased for both procedures. Using Monte Carlo methods, the inclusion probabilities can be estimated so that nearly unbiased Horvitz-Thompson estimators can be used. However, for a large population, simulating enough samples to give a precise estimate of the joint inclusion probability for each pair of units is impractical. An alternative approach to variance estimation is to use a regression estimator and the variance estimator for the regression estimator. This is intuitively appealing because balancing is similar to regression through design.

Upon using the regression estimator, the bias of the regression estimator under both cube and rejective methods is of the same order. For rejective sampling, Fuller (2009a) gives conditions for the consistency of the variance estimator for the regression estimator. For cube sampling, Deville and Tillé (2005) and Tillé (2006) suggest using the variance estimator for a regression estimator furnishes a good approximation to the variance of the Horvitz-Thompson estimator. The variance estimators proposed by Deville and Tillé (2005) perform well when the joint inclusion probabilities of the resulting cube design are approximately equal to joint inclusion probabilities from a Poisson design. In the simulation studies of Section 4, the variance estimators proposed in Fuller (2009a) and Deville and Tillé (2005) are evaluated.

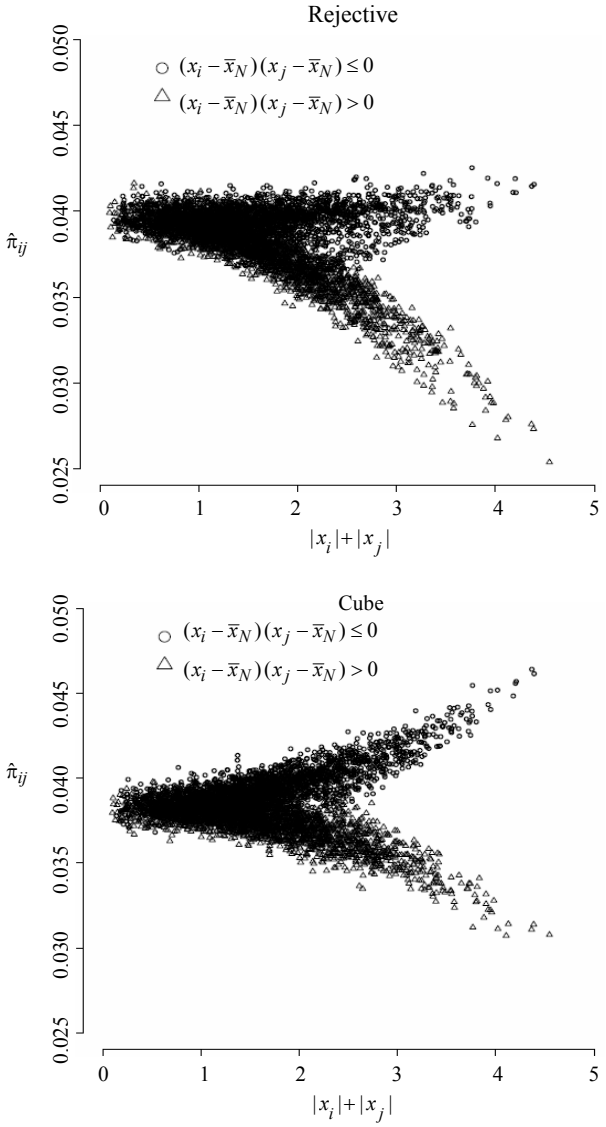

Figure 3 Simulated second-order inclusion probabilities with absolute sums of x. The balancing variable for the rejective method is $z_i = x_i$, and for the cube method is $z_i = (p_i, x_i)'$, where $p_i = 20/100$

4. Simulation of the regression estimator

A population of size 100 was generated from the model

$$y_i = x_i + 0.55\,x_i^2 + x_i\,\varepsilon_i \qquad (9)$$

$\varepsilon_i \sim$ iid $N(0, 0.4)$, where the x_i are fixed values in the range of 0 to 4 (Figure 4). Seventy-two of the x values were randomly simulated values less than 1.15 from a standard exponential distribution. The remaining 28 values, ranging from 0.18 to 4.0, were deterministically added to form the data set of x. The fixed x values were selected to be fairly right skewed so that some large and small strata when stratifying the population on x with approximately equal within-stratum sum of sorted x_i will be produced. The population was held fixed after initial selection. Model (9) contains a quadratic term, and was picked to simulate performance of the design and estimator strategy when model (4) was assumed in design and estimation.

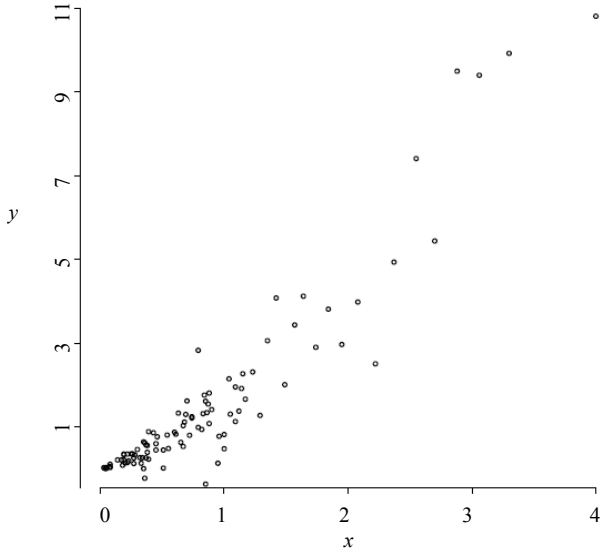

Figure 4 Simulation population under model (9)

We consider Poisson sampling and two-per-stratum stratified random sampling as initial designs. Strata were determined by setting the boundary so that the within stratum sum of sorted x_i was roughly equal for all strata. The sample size was set to 20, and ten strata were formed. The stratum sizes were 35, 15, 11, 9, 8, 7, 5, 4, 3, and 3. The rejection procedure used a stratified two-per stratum sample selection with equal inclusion probabilities within a stratum. The stratum boundaries were chosen this way so that the inclusion probability of unit i is closely proportional to x_i, which is the optimal inclusion probability under model (9) (Ikasi and Fuller 1982). Such a stratified design can also partially balanced on x through a standard design. Balance

in the stratified random sampling design is achieved using a step function to approximate a line. The stratified design will also be partially balanced on x^2. The stratified random sample design is intended to illustrate how much more one can benefit from additional balancing. Two units per stratum were drawn in order to obtain the maximum number of strata while still permitting unbiased variance estimation. Fuller (1981) showed that, in the two-per-stratum case, this stratified design has an anticipated variance close to the best purposive model variance under (4). Initial inclusion probabilities for the Poisson design with expected sample size 20 were set to the initial inclusion probabilities of the stratified design.

The regression estimator considered in this paper is in the form of (1) with $\hat{\boldsymbol{\beta}}$ defined in (2). The regression variable $\boldsymbol{z}$ is a vector of auxiliary variables that contains design variables and x. For the Poisson designs, we used $\boldsymbol{z}_i = (1,\ p_i,\ x_i,\ (1-p_i)^{-1} p_i)'$ as the vector of balancing variables and as the regression variable vector. The first variable provides control for population size, the second variable is a control for sample size, the third variable provides balance on x, and the fourth variable guarantees that the regression estimator is design consistent. See condition (3) for the design consistency of $\bar{y}_{\text{reg}}$ and set $\boldsymbol{d} = (0, 0, 0, 1)'$. For two-per-stratum stratified samples, the vector of balancing variables is $(x_i, I_{1i}, I_{2i}, ..., I_{10i})$ for cube sampling, where I_{hi} are the stratum indicator variables defined as

$$I_{hi} = \begin{cases} 1 & \text{unit } i \text{ in stratum } h \\ 0 & \text{otherwise} \end{cases}$$

for $h = 1, 2, ..., 10$. Only the x variable is included in the rejective balancing procedure since the sample from this initial design is automatically balanced on the stratum indicator variables. The regression variable vector for both balancing procedures is $\boldsymbol{z}_i = (x_i, I_{1i}, ..., I_{10i})'$.

For the initial designs, the variance estimators for $\bar{y}_{\text{reg}}$ are the variance estimators of the mean of $e_i = y_i - \boldsymbol{z}_i' \boldsymbol{\beta}_N$ calculated with $\hat{e}_i$, where $\hat{e}_i = y_i - \boldsymbol{z}_i' \hat{\boldsymbol{\beta}}$. For Poisson sampling, the variance estimator is

$$\hat{V}(\bar{y}_{\text{reg}}) = (n-s)^{-1} n \bar{\boldsymbol{z}}_N' \hat{\boldsymbol{M}}_{zz}^{-1} \sum_{i \in A} \boldsymbol{z}_i\, p_i^{-4} \times (1-p_i)^3\, \hat{e}_i^2\, \boldsymbol{z}_i' \hat{\boldsymbol{M}}_{zz}^{-1} \bar{\boldsymbol{z}}_N, \qquad (10)$$

where

$$\hat{\boldsymbol{M}}_{zz} = N^{-1} \sum_{i \in A} \boldsymbol{z}_i\, p_i^{-2} (1-p_i)\, \boldsymbol{z}_i',$$

and s is the number of variables in $\boldsymbol{z}$. Derivation of (10) is provided in the appendix.

For stratified random sampling with two-per-stratum, the variance estimator for $\bar{y}_{\text{reg}}$ is

$$\hat{V}(\bar{y}_{\text{reg}}) = (H-1)^{-1} H \sum_{h=1}^{H} [(1-W_h)^{1/2} \{0.5 W_h + (\bar{\boldsymbol{z}}_N - \bar{\boldsymbol{z}})\, \hat{\boldsymbol{M}}_{zz,h}^{-1}\, \phi_h\, W_h^2 (\boldsymbol{z}_{h1} - \boldsymbol{z}_{h2})\} \times (\hat{e}_{h1} - \hat{e}_{h2})]^2, \qquad (11)$$

where

$$\hat{\boldsymbol{M}}_{zz,h} = N_h^{-1} \sum_{i \in A_h} \boldsymbol{z}_i\, p_i^{-2}\, \phi_h\, \boldsymbol{z}_i',$$

A_h is the sample set in stratum h, $W_h = n_h / N_h$, $\phi_h = (N_h - 1)^{-1}(N_h - 2)$ for units in stratum h, $\boldsymbol{z}_{hi}$ is the auxiliary variable vector $\boldsymbol{z}_i$ in stratum h,

$$\hat{e}_{hi} = y_{hi} - \bar{y}_h - (\boldsymbol{z}_{hi} - \bar{\boldsymbol{z}}_h)' \hat{\boldsymbol{\beta}},$$

$\bar{y}_h$ and $\bar{\boldsymbol{z}}_h$ are stratum means of y_{hi} and $\boldsymbol{z}_{hi}$, respectively, and $H = 10$ is the number of strata. The derivation of (11) follows the same approach to the one in appendix and has been omitted.

For rejective sampling, the same variance estimators (10) and (11) using the initial design inclusion probabilities, were used to compute the variance estimator of $\bar{y}_{\text{reg}}$ for rejective samples. Fuller (2009a) proved that the large sample properties of the regression estimator for the rejective sample are the same as those of the regression estimator for the original inclusion procedure under some regularity conditions. For cube sampling, a variance estimator proposed by Deville and Tillé (2005) was evaluated for $\bar{y}_{\text{reg}}$ using cube samples.

Let $p(\cdot)$ denote the initial design and $\pi(\cdot)$ be the resulting scheme after balancing. The number of samples selected was 30,000 for each Monte Carlo simulation under initial designs, cube sampling and rejective sampling with both 90% and 95% rejection rates. The Horvitz-Thompson estimator $\bar{y}_{\text{HT}}$ and the regression estimator $\bar{y}_{\text{reg}}$ were constructed using initial inclusion probabilities p_i. Note that for rejection sampling, the Horvitz-Thompson estimator using the initial inclusion probabilities is not the Horvitz-Thompson estimator under the balanced designs. For each initial design, the following quantities were computed in the simulation studies.

- $V_p(\bar{y}_{\text{HT}})$ (or $V_p(\bar{y}_{\text{reg}})$): Monte Carlo variance of the Horvitz-Thompson estimator (or the regression estimator) using samples from initial designs.
- $V_\pi(\bar{y}_{\text{HT}})$ (or $V_\pi(\bar{y}_{\text{reg}})$): Monte Carlo variance of the Horvitz-Thompson estimator (or the regression estimator) for balanced samples.
- $\text{bias}_\pi(\bar{y}_{\text{HT}})$ (or $\text{bias}_\pi(\bar{y}_{\text{reg}})$): Monte Carlo bias of the Horvitz-Thompson estimator (or the regression estimator) using balanced samples.

For cube samples,

- $\hat{V}_{DT}(\bar{y}_{\text{reg}})$: estimated variance of the regression estimator using the variance estimators in Deville and Tillé (2005) and each cube sample.
- $\text{ave}(\hat{V}_{DT}(\bar{y}_{\text{reg}}))$: Monte Carlo average of $\hat{V}_{DT}(\bar{y}_{\text{reg}})$ using all cube samples.

Deville and Tillé (2005) recommend several variance estimators based on a Poisson sampling approximation with corrections for known constraints in the design variance. The first three estimators in Deville and Tillé (2005) have minor differences, therefore only the second estimator was used in the simulation studies. Deville and Tillé (2005) also propose the fourth estimator, but that estimator requires solving a nonlinear equation system, which would have been computationally expensive to add to the simulation. However, the fourth estimator could perform better than the other cases for stratified designs, since their fourth estimator reproduces the variance of a stratified random sample when the balancing vector contains stratum indicators.

For rejective samples,

- $\hat{V}(\bar{y}_{\text{reg}})$: estimated variance of the regression estimator using equation (10) (or (11)) for the Poisson (or two-per-stratum stratified) initial design and each balanced sample.
- $\text{ave}(\hat{V}(\bar{y}_{\text{reg}}))$: Monte Carlo average of $\hat{V}(\bar{y}_{\text{reg}})$ using all balanced samples.

In the simulations, $\hat{V}(\bar{y}_{\text{reg}})$ was also computed for cube samples, for comparison.

Table 1 reports the estimates for the Poisson design. The variance of the Horvitz-Thompson mean under initial Poisson sampling with expected sample size 20 and no balancing is $V_p(\bar{y}_{\text{HT}}) = 0.08$. The variances in Table 1 are standardized by $V_p(\bar{y}_{\text{HT}})$, and the biases are standardized by $\sqrt{V_p(\bar{y}_{\text{HT}})}$. The Horvitz-Thompson estimator is unbiased under the cube method designs, because cube sampling retains the first order inclusion probabilities. The Horvitz-Thompson estimator using initial design inclusion probabilities is biased under rejective sampling since the inclusion probabilities differ from the initial design inclusion probabilities, as indicated in Figure 1. The bias of the regression estimator under rejective sampling is less than the bias of the Horvitz-Thompson estimator with initial design inclusion probabilities. The bias of $\bar{y}_{\text{reg}}$ under both cube and rejective procedures is of the same order. Increasing the rejection rate increases the bias of $\bar{y}_{\text{reg}}$ for the rejection designs. However, the biases in $\bar{y}_{\text{reg}}$ under both balancing procedures and rejection rates are negligible relative to the Monte Carlo variances. For the Horvitz-Thompson estimator using initial design inclusion probabilities, the gain from using the balanced sample is substantial for both cube and rejective methods. The mean squared errors are further reduced by using the regression estimator along with either balancing procedures. The gain from using the regression estimator is larger for rejective sampling than for cube sampling, likely due to the cube method achieving tighter balance than the rejective method. Both procedures lead to similar variances for the regression estimator. The variance of the regression estimator under the Poisson initial design is $V_p(\bar{y}_{\text{reg}}) = 0.249$ (relative to $V_p(\bar{y}_{\text{HT}})$). By comparing 0.249 to the fourth row of Table 1, we can see that the gain from using the balanced samples on the regression estimator is moderate. The result is consistent with the finding in Fuller (2009a) that the variance reduction in $\bar{y}_{\text{reg}}$ by using rejective samples is due to a second order correction. The variance estimator of $\bar{y}_{\text{reg}}$ using (10) has small bias for both cube and rejective samples ($\text{ave}(\hat{V}(\bar{y}_{\text{reg}}))$ in Table 1). The variance estimator $\hat{V}_{DT}(\bar{y}_{\text{reg}})$ proposed in Deville and Tillé (2005) performed similarly as $\hat{V}(\bar{y}_{\text{reg}})$ in (10) since the second variance estimator in Deville and Tillé (2005) is very close to (10) for Poisson sampling. This result supports the claim that the Poisson approximation assumption in the variance estimators of Deville and Tillé (2005) is satisfied for the Poisson design case.

Table 1
Properties of samples based on Poisson sampling of expected size 20. $V_p(\bar{y}_{\text{HT}}) = 0.08$ and $V_p(\bar{y}_{\text{reg}})/V_p(\bar{y}_{\text{HT}}) = 0.249$

	Cube	Rej. 90%	Rej. 95%
$\text{bias}_\pi(\bar{y}_{\text{HT}})/\sqrt{V_p(\bar{y}_{\text{HT}})}$	-0.002	-0.016	-0.007
$\text{bias}_\pi(\bar{y}_{\text{reg}})/\sqrt{V_p(\bar{y}_{\text{HT}})}$	-0.002	0.002	0.005
$V_\pi(\bar{y}_{\text{HT}})/V_p(\bar{y}_{\text{HT}})$	0.142	0.270	0.220
$V_\pi(\bar{y}_{\text{reg}})/V_p(\bar{y}_{\text{HT}})$	0.131	0.136	0.129
$\text{ave}(\hat{V}(\bar{y}_{\text{reg}}))/V_p(\bar{y}_{\text{HT}})$	0.122	0.123	0.121
$\text{ave}(\hat{V}_{DT}(\bar{y}_{\text{reg}}))/V_p(\bar{y}_{\text{HT}})$	0.120	-	-

In Table 2, estimates under the initial two-per-stratum stratification design are reported. The variance of the Horvitz-Thompson mean under the initial stratification design is $V_p(\bar{y}_{\text{HT}}) = 0.011$ and all estimates are standardized by this value. Since stratification in this initial design controls for most of the effect of x on y, the regression estimator is not a large improvement over the Horvitz-Thompson estimator using initial design inclusion probabilities. The bias and variance of $\bar{y}_{\text{HT}}$ are close to those of $\bar{y}_{\text{reg}}$ under both cube and rejective methods. The larger estimated bias in $\bar{y}_{\text{HT}}$ under cube sampling is due to Monte Carlo error. The gain from balancing on x is not large, compared to the gain in the Poisson example. However, with this highly controlled initial stratified design, in which

the initial samples are already partially balanced on x, there still can be a modest benefit from additional balancing and using $\bar{y}_{\text{reg}}$ estimators. This result is seen for $\bar{y}_{\text{reg}}$ by comparing the fourth row of Table 2 to the variance of $\bar{y}_{\text{reg}}$ under the initial design $V_p(\bar{y}_{\text{reg}}) = 0.987$. Therefore, in this case a good strategy is to combine stratification, balancing, and regression, which is a similar conclusion drawn in Deville and Tillé (2004). The variance estimator $\hat{V}(\bar{y}_{\text{reg}})$ using (11) gives estimates on average for the regression estimator variances under both cube and rejective procedures that are close to the true variances. However, the variance estimator $\hat{V}_{DT}(\bar{y}_{\text{reg}})$ proposed by Deville and Tillé (2005) performed poorly for cube sampling. A possible reason is that the Poisson sampling approximation in the second variance estimator of Deville and Tillé (2005) assumes joint inclusion probabilities that are far from the actual joint inclusion probabilities in the small strata. The joint inclusion probabilities in the small strata are closer to those of stratified random sampling than Poisson sampling. This issue might explain why $\hat{V}(\bar{y}_{\text{reg}})$ in (11) using the initial two-per-stratum inclusion probabilities is less biased than $\hat{V}_{DT}(\bar{y}_{\text{reg}})$ in this case.

Table 2
Properties of samples based on stratified sampling of size 20. $V_p(\bar{y}_{\text{HT}}) = 0.011$ and $V_p(\bar{y}_{\text{reg}})/V_p(\bar{y}_{\text{HT}}) = 0.987$

	Cube	Rej. 90%	Rej. 95%
$\text{bias}_\pi(\bar{y}_{\text{HT}})/\sqrt{V_p(\bar{y}_{\text{HT}})}$	-0.028	0.014	0.010
$\text{bias}_\pi(\bar{y}_{\text{reg}})/\sqrt{V_p(\bar{y}_{\text{HT}})}$	-0.013	0.014	0.010
$V_\pi(\bar{y}_{\text{HT}})/V_p(\bar{y}_{\text{HT}})$	0.910	0.866	0.813
$V_\pi(\bar{y}_{\text{reg}})/V_p(\bar{y}_{\text{HT}})$	0.929	0.865	0.813
$\text{ave}(\hat{V}(\bar{y}_{\text{reg}}))/V_p(\bar{y}_{\text{HT}})$	0.907	0.881	0.775
$\text{ave}(\hat{V}_{DT}(\bar{y}_{\text{reg}}))/V_p(\bar{y}_{\text{HT}})$	0.792	-	-

To assess large sample properties of the balancing procedures, the size of the Poisson simulation was quadrupled. The population was replicated four times and a sample of expected size 80 was selected. The Horvitz-Thompson variance of a mean under the Poisson design is $V_p(\bar{y}_{\text{HT}}) = 0.020$ and the regression estimator variance is $V_p(\bar{y}_{\text{reg}}) = 0.132$. The resulting relative variances and biases are close to the results for samples of size 20 (Table 3). The simulation results agree with the theoretical result of Fuller (2009a) that the regression estimator is an $O_p(n^{-1/2})$ estimator after rejection of the type used in this paper. Although it has not been proven here, regression estimator after cube sampling appears to possesses similar properties to the regression estimator using rejection sampling.

Table 3
Properties of samples based on Poisson sampling of expected size 80. $V_p(\bar{y}_{\text{HT}}) = 0.02$ and $V_p(\bar{y}_{\text{reg}})/V_p(\bar{y}_{\text{HT}}) = 0.132$

	Cube	Rej. 90%	Rej. 95%
$\text{bias}_\pi(\bar{y}_{\text{HT}})/\sqrt{V_p(\bar{y}_{\text{HT}})}$	0.002	-0.006	-0.007
$\text{bias}_\pi(\bar{y}_{\text{reg}})/\sqrt{V_p(\bar{y}_{\text{HT}})}$	0.002	0.000	-0.001
$V_\pi(\bar{y}_{\text{HT}})/V_p(\bar{y}_{\text{HT}})$	0.127	0.267	0.224
$V_\pi(\bar{y}_{\text{reg}})/V_p(\bar{y}_{\text{HT}})$	0.122	0.124	0.123
$\text{ave}(\hat{V}(\bar{y}_{\text{reg}}))/V_p(\bar{y}_{\text{HT}})$	0.121	0.121	0.121
$\text{ave}(\hat{V}_{DT}(\bar{y}_{\text{reg}}))/V_p(\bar{y}_{\text{HT}})$	0.121	-	-

5. Adjustments to the rejection procedure

Fuller's rejection sampling procedure treats all balancing variables with the same importance. For a large number of balancing variables, exact balance on all variables cannot be expected and the approximation could be poor for some important variables. Therefore, a practitioner may want to have tighter balance on a subset of the balancing variables. As an example, a researcher may want to use Poisson sampling for simplicity but also have some control on the random sample size. A random sample size can complicate study planning and is a large contributor to the variance of estimators. Balanced sampling can be used to reduce the variation in sample sizes by balancing on the variable p_i, which is the initial first-order inclusion probability. For Fuller's rejection procedure, the variance of the sample size increases when the number of balancing variables increases and the rejection rate is held constant. The rejection procedure can be altered so that the p_i balance is tighter than the balance for other variables.

One approach to increasing the balancing on a subset of variables is to change the rejection test function. The order of the approximation to the first and second-order inclusion probabilities in Fuller (2009a) remains the same when the variance matrix in the rejection quadratic form is replaced with a symmetric positive definite matrix of the same order.

To determine weights for weighted rejection sampling, it is convenient to transform the balancing variables so that $V(\bar{z}_{\text{HT}} \mid F_N)$ is a diagonal matrix. The weighted rejection sampling test statistic is

$$\sum_{q=1}^{m} c_q V(\bar{z}_{\text{HT},q} \mid F_N)^{-1}(\bar{z}_{\text{HT},q} - \bar{z}_{N,q})^2, \qquad (12)$$

where m is the number of balancing variables, z_q is the q^{th} balancing variable, and c_q are selected weights. The weight on the first variable $z_{1i} = p_i$ can be set large relative to the weights on other variables to reduce variation in sample size. The transformation is the Gramm-Schmidt transformation using the design variances under the initial design. Balancing

is done on the transformed variables, but the first variable is not transformed. The transformed variables have uncorrelated Horvitz-Thompson estimators. Balancing on the transformed variables will still balance the original variables since each transformed variable is a residual from a regression operation on preceeding variables.

Equation (12) can be paralleled to the penalty term of the distance function underlying ridge calibration. See Rao and Singh (1997), Beaumont and Bocci (2008), and Chambers (1996). Specifically, selection of the c_q weights is similar to the problem of selecting appropriate costs in ridge calibration. Thus, rejection sampling using (12) can be viewed as incorporating ridge calibration at the design stage.

A second way to produce tighter balance on a subset of variables is to do rejection separately for subsets. A test statistic is produced for each subset and a sample must be accepted by all of the tests to be accepted. In the Poisson case, one test statistic may reject if the sample size is not within a specified tolerance of the expected sample size. This second approach requires some additional assumptions beyond those in Fuller (2009a), but a similar argument can be used to justify the procedure.

To prove the convergence properties of the multiple test rejection procedure, it is convenient to consider two subsets of balancing variables and think of rejection being done sequentially on each subset. We call the two subset rejection procedure a two-step rejective sampling procedure. Suppose $\boldsymbol{z}_i' = (\boldsymbol{z}_{1i}', \boldsymbol{z}_{2i}')$ is the balancing vector and the original design is denoted as $p(\cdot)$. The procedure is as follows.

Step 1: Select a sample using $p(\cdot)$ and reject samples with the balancing condition (8) on the first subset $\boldsymbol{z}_1$,

$$Q_1 = (\overline{\boldsymbol{z}}_{\mathrm{HT},1} - \overline{\boldsymbol{z}}_{N,1})' V(\overline{\boldsymbol{z}}_{\mathrm{HT},1} | F_N)^{-1} (\overline{\boldsymbol{z}}_{\mathrm{HT},1} - \overline{\boldsymbol{z}}_{N,1}) < \gamma_1.$$

Step 2: Use the accepted sample from step 1 to check the balancing condition (8) on the second subset $\boldsymbol{z}_2$,

$$Q_2 = (\overline{\boldsymbol{z}}_{\mathrm{HT},2} - \overline{\boldsymbol{z}}_{N,2})' V(\overline{\boldsymbol{z}}_{\mathrm{HT},2} | F_N)^{-1} (\overline{\boldsymbol{z}}_{\mathrm{HT},2} - \overline{\boldsymbol{z}}_{N,2}) < \gamma_2.$$

Reject the sample if the condition is not satisfied and repeat Step 1.

In both weighted and two-step procedures, trial and error is likely needed to choose γ's in practice. In the weighted procedure, the quadratic form becomes a sum of multiples of χ^2 random variables, which makes selection of γ more difficult than in the unweighted case. We used moment matching approximations to select γ's that provide rejection rates close to desired, but then resorted to small simulations to determine the rejection rate as a function of γ. For the two-step procedure, we used a χ^2 approximation to select a γ_1 that gave approximately the desired rejection rate at the first step, and used second χ^2 approximation to select an initial γ_2 that gave approximately the desired rejection rate at the second step. The second parameter γ_2 was adjusted in order to achieve the target overall rejection rate. The choice of γ's in the two-step procedure is subjective because many combinations of γ_1 and γ_2 can produce the same overall rate. In practice, a practitioner likely will set a tight bound for the first variable subset and loose bounds on the remaining balancing variables.

The large sample mean and variance of the regression estimator under the two-step rejective sample are the same as those of the regression estimator for the original design. Also, the usual estimator of variance under the original design for the regression estimator is appropriate for the two-step rejective sample. The proof of this statement is an extension of the proof in Fuller (2009a) and can be provided upon request.

To examine some properties of the two procedures, the Monte Carlo simulations for the Poisson initial sample design were repeated with the variable p_i separated from the other three variables. The balancing vector was transformed so that the variance matrix of the Horvitz-Thompson total estimators was diagonal. For the weighting procedure, the weight on the p_i component of the quadratic form was set to 1.5, the weights on the other components were set to 1, and γ was set to 0.627. This weighting procedure restricted the samples to those with sample sizes ranging from 18 to 22. For the two-step procedure, any sample with a sample size outside of the range from 18 to 22 was rejected in the first step and then the quadratic form for the remaining three variables was checked using a γ of 0.63 for the second step. Given the good performance of the variance estimator $\hat{V}(\overline{y}_{\mathrm{reg}})$ in (10), Table 4 only contains its Monte Carlo averages values $\mathrm{ave}(\hat{V}(\overline{y}_{\mathrm{reg}}))$.

Table 4
Properties of rejection samples with adjustments based on Poisson sampling of expected size 20, and 95% rejection rate

	Weighted	Two-step
$\mathrm{bias}_\pi(\overline{y}_{\mathrm{HT}})/\sqrt{V_p(\overline{y}_{\mathrm{HT}})}$	-0.005	-0.014
$\mathrm{bias}_\pi(\overline{y}_{\mathrm{reg}})/\sqrt{V_p(\overline{y}_{\mathrm{HT}})}$	0.003	0.002
$V_\pi(\overline{y}_{\mathrm{HT}})/V_p(\overline{y}_{\mathrm{HT}})$	0.210	0.217
$V_\pi(\overline{y}_{\mathrm{reg}})/V_p(\overline{y}_{\mathrm{HT}})$	0.132	0.132
$\mathrm{ave}(\hat{V}(\overline{y}_{\mathrm{reg}}))/V_p(\overline{y}_{\mathrm{HT}})$	0.121	0.121
$V_\pi(n)$	1.237	1.902

Results for expected sample size of 20 and a rejection rate near 95% were similar for the two adjustment procedures (Table 4). The Horvitz-Thompson estimator for the weighted procedure performed slightly better than the Horvitz-Thompson estimator for the two-step procedure. A reason for this discrepancy is that the weighted procedure

had much less variation in sample sizes ($V_\pi(n)$ in the last row of Table 4). Additional simulations with larger expected sample sizes gave similar relative variances. The regression estimator performed at roughly the same efficiency for the two procedures. The Horvitz-Thompson estimators using the initial design inclusion probabilities for these adjustment procedure performed slightly better than the Horvitz-Thompson estimator for the rejection procedure that did not place additional control on the sample size.

6. Discussion

Rejection sampling and cube sampling produce roughly equally performing regression estimators. Balancing provides major gains when the initial design provides little control on the auxiliary values entering samples. A well stratified sample design provides many of the benefits of balancing on a continuous variable. However, further balancing after stratification can still yield small mean squared error gains for regression estimators. Additionally, balancing could be used to prevent negative weights produced by regression estimators (Fuller 2009a).

For the simulations, the rejection rate was fixed at 90% for the larger population. When the population and sample sizes are increased, the rejection rate can be increased while still maintaining a large set of possible samples. Additional simulations were carried out with rejection rates near 99%, but the results were not presented since the differences between the results with 95% and with 99% were very small and the bias of $\bar{y}_{\text{reg}}$ remained negligible. The marginal variance reduction due to balancing decreases as the balancing condition is tightened.

In some special cases, an investigator may want to balance tightly on some variables and weakly on others. Gains can be made by choosing different weights for different variables or by dividing the variables into separate test sets. The weighted and two-step rejection procedures performed comparably, so the decision between procedures will largely be based on the ease of implementation.

Acknowledgements

This work was supported by Cooperative Agreement No. 68-3A75-4-122 between the USDA Natural Resources Conservation Service and the Center for Survey Statistics and Methodology at Iowa State University. The authors thank Wayne A. Fuller, the associate editor and two anonymous referees for helpful comments that improve the paper.

Appendix

Start with

$$V(\bar{y}_{\text{reg}} \mid F_N) = V(\bar{y}_{\text{reg}} - \bar{y}_N \mid F_N).$$

Let

$$\bar{y}_N = \bar{\boldsymbol{z}}_N' \boldsymbol{\beta}_N$$

and note

$$y_i = \boldsymbol{z}_i' \boldsymbol{\beta}_N + e_{Ni}$$

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i\in A} \boldsymbol{z}_i \phi_i p_i^{-2} \boldsymbol{z}_i'\right]^{-1} \sum_{i\in A} \boldsymbol{z}_i \phi_i p_i^{-2} (\boldsymbol{z}_i' \boldsymbol{\beta}_N + e_i)$$

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_N + \left[N^{-1}\sum_{i\in A} \boldsymbol{z}_i \phi_i p_i^{-2} \boldsymbol{z}_i'\right]^{-1} N^{-1}\sum_{i\in A} \boldsymbol{z}_i \phi_i p_i^{-2} e_i. \quad (13)$$

Under assumptions (design consistency standard assumptions)

$$N^{-1}\sum_{i\in A} \boldsymbol{z}_i \phi_i p_i^{-2} \boldsymbol{z}_i' = N^{-1}\sum_{i\in U} \boldsymbol{z}_i \phi_i p_i^{-1} \boldsymbol{z}_i' + O_p(n^{-1/2}).$$

Write

$$N^{-1}\sum_{i\in U} \boldsymbol{z}_i \phi_i p_i^{-1} \boldsymbol{z}_i' = \boldsymbol{M}_{zz,N}.$$

Use the same argument to expand the $N^{-1}\sum_{i\in A} \boldsymbol{z}_i \phi_i p_i^{-2} e_i$ term. Then the expansion of (13) is

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_N + \boldsymbol{M}_{zz,N}^{-1} N^{-1}\sum_{i\in A} \boldsymbol{z}_i \phi_i p_i^{-2} e_i + O_p(n^{-1}).$$

For construction of confidence intervals for $\bar{y}_N$ it is enough to consider the variance of the linearized term. Therefore consider in the notation of Särndal, Swensson, and Wretman (1992),

$$\text{AV}(\bar{y}_{\text{reg}}) = \bar{\boldsymbol{z}}_N' \boldsymbol{M}_{zz,N}^{-1} V(\bar{\boldsymbol{b}}_{\text{HT}} \mid F_N) \boldsymbol{M}_{zz,N}^{-1} \bar{\boldsymbol{z}}_N$$

where

$$\boldsymbol{b}_i = \boldsymbol{z}_i \phi_i p_i^{-1} e_i.$$

The variance of the HT estimator for the mean of b_i under Poisson sampling is

$$\sum_{i\in U} (1 - p_i) p_i^{-1} \boldsymbol{b}_i \boldsymbol{b}_i'.$$

Next apply that $\phi = 1 - p_i$ to obtain the asymptotic variance approximation to the linearized part of $\bar{y}_{\text{reg}}$

$$\text{AV}(\bar{y}_{\text{reg}}) = \bar{\boldsymbol{z}}_N' \boldsymbol{M}_{zz,N}^{-1} \sum_{i\in U} (1 - p_i)^3 p_i^{-3} \boldsymbol{z}_i e_i^2 \boldsymbol{z}_i' \boldsymbol{M}_{zz,N}^{-1} \bar{\boldsymbol{z}}_N.$$

The variance estimator is obtained by replacing the population totals with HT estimators under Poisson sampling and incorporating a degree of freedom correction to the front of $n/(n-s)$ due to the small sample size.

References

Beaumont, J.-F., and Bocci, C. (2008). Another look at ridge calibration. *Metron*, 66, 1, 5-20.

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

Deville, J.-C., and Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91, 893-912.

Deville, J.-C., and Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128, 569-591.

Fuller, W.A. (1981). An empirical Study of the ratio estimator and estimators of its variance: Comment. *Journal of the American Statistical Association*, 76, 78-80.

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

Fuller, W.A. (2009a). Some design properties of a rejective sampling procedure. Forthcoming *Biometrika*.

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

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

Matei, A., and Tillé, Y. (2005). Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size. *Journal of Official Statistics*, 21, 543-570.

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

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, Washington, D.C., 57-65.

Rousseau, S., and Tardieu, F. (2004). La macro SAS CUBE d'échantillonnage équilibré, Documentation de l'utilisateur. Technical report, INSEE, Paris.

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

Royall, R.M., and Herson, J. (1973). Robust estimation in finite populations I. *Journal of the American Statistical Association*, 68, 880-889.

Särndal, C.-E. (1980). On π inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, 67, 639-650.

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

Tillé, Y. (2006). *Sampling Algorithms*. New York: Springer Science+ Business Media, Inc.

Tillé, Y., and Matei, A. (2005). The R package Sampling. *The Comprehensive R Archive Network*, *http://cran.r-project.org/*, *Manual of the Contributed Packages*.

Valliant, R., Dorfman, A.H. and Royall, R.M. (2000). *Finite Population Sampling and Inference*: *A Prediction Approach*. New York: John Wiley & Sons, Inc.