THE THEORY AND PRACTICE OF MAXIMAL BREWER SELECTION WITH POISSON PRN SAMPLING

Phillip S. Kott And Jeffrey T. Bailey, National Agricultural Statistics Service
Phillip S. Kott, NASS, Room 305, 3251 Old Lee Hwy, Fairfax, VA 22030, USA
pkott@nass.usda.gov

K.R.W. Brewer suggests that when estimating the total of a single item for which there is control (auxiliary) data, one employ a ratio or regression estimator and draw the sample using probabilities proportional to the control values raised to a power between 1/2 and 1. Brewer's sample selection scheme can be expanded to multiple targets by drawing overlapping Poisson samples for a number of items simultaneously using permanent random numbers (PRN's). We can call the result an example of "Maximal Brewer Selection" (MBS). This paper develops the theory behind MBS and the calibration estimator rendering it practical. It goes on to describe how this estimation strategy is being used at the National Agricultural Statistics Service.

Key Words: Model; Model-assisted; Randomization; Calibration; Delete-a-group jackknife

1. INTRODUCTION

K.R.W. Brewer's (1963) article in the *Australian Journal of Statistics* is one of the truly remarkable works in the survey sampling literature. It discusses a model-based approach to survey sampling theory, contrasts that approach with the conventional randomization paradigm, and shows how the two can be used in tandem. All this seven years before Royall (1970) set the survey world buzzing with prediction theory (another name for the model-based approach) and almost three decades before the publication of Särndal et al. (1992) made model-assisted survey sampling (which uses models and randomization in tandem) the new conventional wisdom.

This paper builds on one small result in Brewer's impressive opus and some of his work since then. Suppose we are interested in estimating a population (P) total, $T = \sum_P y_i$, with a random sample (S) of size n. We suspect that the y_i follow the model

$$y_i = \beta x_i + k\varepsilon_i, \tag{1}$$

where $E(\varepsilon_i | x_i) = E(\varepsilon_i \varepsilon_j | x_i, x_j) = 0$ $(i \neq j)$, and $Var(\varepsilon_i | x_i) = \sigma_i^2$ is known for all i (but k need not be known).

Equation (1) is a useful model for many establishment surveys. Whether or not it is correct, the following estimator is nearly randomization unbiased for large n (and randomization consistent under a number of sampling designs),

$$t = (\sum_P x_i)\sum_S (y_i/\pi_i)/\sum_S (x_i/\pi_i),$$

where π_i is the selection probability of unit i. Of course, in order for t to be practical, the population sum $\sum_P x_i$ must be known, and the individual x_i must also be known for all units in the *sample*. In what follows, we further require x_i to be known for all units in the *population*. Such an x is called a "control" variable for the *target* variable y.

Brewer showed that when $\pi_i \propto \sigma_i$ the randomization-expected model variance of t was (asymptotically) minimized for fixed sample size n. In this sense, $\pi_i = n\sigma_i / \sum_P \sigma_k$ – if less than or equal to 1 for all i – is the optimal selection scheme given sample size n and estimator t . Godambe (1955) has a similar result for randomization unbiased estimators.

It is sometimes assumed that the σ_i have the form x_i^g, where $0 \leq g \leq 1$. If that is the case, then when g = 1, the optimal selection scheme (i.e., randomization-expected model-variance minimizing) is probability proportional to size, $\pi_i = nx_i / \sum_P x_k$, and t collapses into the Horvitz-Thompson mean-of-ratios estimator $(n^{-1}\sum_S (y_i/\pi_i)$; this is Godambe's 1955 result). When g = 0, the optimal sampling scheme is self-weighting, $\pi_i = n/N$. For establishment surveys, however, g is usually between ½ and 1. Brewer has said (out loud, if not in print) that a sensible value for g in many surveys is 3/4.

Sadly, Brewer's suggestion that the unit selection probabilities be in proportion to some known control value, x_i, raised to the 3/4 power has not been implemented much in practice. One problem is that many real establishment surveys have

multiple targets of interest with varying relevant control values. Recently, however, several survey organizations have come to use calibration estimators in place of traditional expansion and ratio estimators. This has allowed the National Agricultural Statistics Service (NASS) to begin implementing a multivariate version of Brewer's suggestion in its Crops/Stocks Survey (CS). Internally, NASS calls this procedure "multivariate probability proportional to size" sampling. A better name would be "Maximal Brewer Selection" (MBS). This method of sample selection has proven more flexible than the stratification approaches NASS has traditionally used (see Bosecker 1989).

Section 2 fills out the theory of Brewer selection when there is a single target and control. Section 3 describes a simple extension for multiple targets each with its own control variable. Briefly, a Brewer selection probability is assigned to each population unit for every target variable of interest. The largest of these for each unit is then used for the actual sample selection. Section 4 addresses a several questions that NASS needed to resolve before making MBS practical to use. Poisson PRN sampling allows the agency to focus on different combinations of target variables in different survey periods. Section 5 further describes NASS's experience with this new sampling scheme. Section 6 offers some comments including one that describes a method for co-ordinating samples to minimize overlap.

2. BREWER SELECTION

2.1. Some Theory

Suppose we have target values, y_i, which we believe (roughly) obey the model in equal (1). We will call $t_C = \sum_S a_i y_i$, based on a sample S with n members, a *calibration estimator* for T if the *calibration equation*

$$\sum_S a_i x_i = \sum_P x_i \tag{2}$$

is satisfied, and each $a_i = \pi_i^{-1}[1 + O_P(1/\sqrt{n})]$, where π_i is (again) the selection probability of unit i, and O_P refers to an asymptotic probability order *with respect to the randomization rather than the model* (see Isaki and Fuller 1982 for a development of asymptotics in a finite population context). This is a bit of a generalization of the definition of a calibration estimator in Deville and Särndal (1992).

One obvious choice for the a_i is $\pi_i^{-1}(\sum_P x_k / \sum_S [x_k/\pi_k])$. This renders t_C equal to t in Section 1. The choice satisfies the calibration equation, and the a_i are sufficiently close to the π_i^{-1} as long as the design and population are such that $(\sum_S [x_k/\pi_k] - \sum_P x_k)/\sum_P x_k$ is $O_P(1/\sqrt{n})$.

The model variance of t_C as an estimator of T is

$$\begin{aligned} E_\varepsilon[(t_C - T)^2] &= E_\varepsilon[(\textstyle\sum_S a_i y_i - \sum_P y_i)^2] \\ &= E_\varepsilon[(\textstyle\sum_S a_i \varepsilon_i - \sum_P \varepsilon_i)^2] \\ &= \textstyle\sum_S a_i^2 \sigma_i^2 - 2\sum_S a_i \sigma_i^2 + \sum_P \sigma_i^2. \end{aligned} \tag{3}$$

Since each $a_i \approx 1/\pi_i$, $E_\varepsilon[(t_C - T)^2] \approx \sum_S \sigma_i^2/\pi_i^2 - 2\sum_S \sigma_i^2/\pi_i + \sum_P \sigma_i^2$

Technically, the relative difference between the left and right hand sides of the above equation is $O_P(1/\sqrt{n})$. For our purposes, this defines when the two sides of an equation are approximately equal.

The randomization expectation (denoted using the subscript "P") of the model variance of t_C is

$$E_P\{ E_\varepsilon[(t_C - T)^2]\} \approx \sum_P \sigma_i^2/\pi_i - \sum_P \sigma_i^2. \tag{4}$$

Under mild conditions, this is the same as the model expectation of the randomization means squared error of t_C. Isaki and Fuller called that last quantity the "anticipated variance" of t_C, presumably meaning “the anticipation under the model of the randomization mean squared error or variance” (randomization mse and variance are virtually identical

under the designs Isaki and Fuller had in mind). We will use their term here, but keep in mind an alternative meaning for "anticipated variance:" the model variance anticipated before sampling.

If we restrict ourselves to a randomization estimator like t_C, a sensible policy is to choose selection probabilities so that the right hand side of equation (4) is minimized for a given sample size n. Since $n = \sum_P \pi_i$, it is a simple matter to set up a Langrangian equation, the solution to which is $\pi_i = n\sigma_i/\sum_P \sigma_k$. For this solution to be valid each π_i must be no greater than 1. We assume that to be the case for the time being.

The anticipated variance of t_C is (asymptotically) minimized by setting the unit probabilities of selection equal to $n\sigma_i/\sum_P \sigma_k$ *no matter which method it used to draw the sample*. In fact, the same minimum variance is obtained if the sample size itself is allowed to be random with an *expected value* equal to n. Poisson sampling is a simple example of a sampling scheme with a random sample size.

2.2. The Selection Scheme

Suppose we have a working assumption about the σ_i in equation (1). In particular, suppose σ_i is believed to be proportional to x_i^g for some g between 1/2 and 1. Let us reparameterize the model as

$$y_i = \beta(x_i + [\sum_P x_k / \sum_P x_k^g]x_i^g\varepsilon_i), \tag{5}$$

where (again) $E(\varepsilon_i|x_i) = E(\varepsilon_i\varepsilon_j|x_i, x_j) = 0$ $(i \neq j)$, and (now) $Var(\varepsilon_i|x_i) = \sigma^2$. We have chosen this parameterization so that σ is invariant to changes in scale (units of measurement) of the y_i and x_i. Notice that when g =1, σ^2 is the relative variance of y_i under the model. Thus, σ^2 for any g is in some sense a generalized relative variance for y_i.

Observe that σ_i^2 in equation (4) now equals $\beta^2 [\sum_P x_k / \sum_P x_k^g]^2 x_i^{2g}\sigma^2$. Since under the model $T \approx \beta \sum_P x_k$, the relative anticipated variance of t_C is

$$\frac{E_P\{ E_\varepsilon[(t_C - T)^2]\}}{E_\varepsilon(T^2)} \approx \frac{\sum_P x_i^{2g}(\pi_i^{-1} - 1)}{(\sum_P x_i^g)^2}\ \sigma^2.$$

Similarly, the *asymptotic anticipated coefficient of variance* for t_C under the model in equation (5) can be defined as

$$ACV(t_C) = \frac{[\sum_P x_i^{2g}(\pi_i^{-1} - 1)]^{1/2}}{\sum_P x_i^g}\ \sigma. \tag{6}$$

Observe that $ACV(t_C)$ decreases, all other things held constant, as any of the π_i increases.

The right hand side of equation (6) attains its minimum for a fixed expected sample size, $n_E = \sum_P \pi_i$, when $\pi_i = n_E x_i^g/\sum_P x_k^g$ if all these selection probabilities are bounded by 1. Furthermore, at that minimum, $ACV(t_C) \leq \sigma/\sqrt{n_E}$. Near equality holds when all $n_E x_i^g/\sum_P x_k^g << 1$.

Equation (6) further tells us that if we knew σ, we could be assured of meeting meeting an ACV target, say, C. We do this by setting $\pi_i = \min\{1, n_T x_i^g/\sum_P x_k^g\}$ and $n_T \geq (\sigma/C)^2$.

We can call n_T the "targeted sample size." The expected sample size, $n_E = \sum_P \pi_i$, is less than or equal to n_T. Equality holds only when all the $n_T x_i^g/\sum_P x_k^g$ are bounded by 1, which we are *not* requiring. Nevertheless, *setting the selection probabilities at* $\pi_i = min\{1, n_T x_i^g/\sum_P x_k^g\}$ *assures* $ACV(t_C) \leq \sigma/\sqrt{n_T}$ under the model in equation (5) .

In practice σ^2 must be guessed at or estimated from previous data, say by

$$s^2 = \frac{\sum_f w_i x_i^g e_i^2}{\sum_f w_i x_i^g},$$

where f denotes the previous sample, w_i is the weight for unit i in that sample, $e_i = [\sum_F x_k^g / \sum_F x_k](y_i - bx_i)/(bx_i^g)$, $b = \sum_f w_i y_i / \sum_f w_i x_i$, and F is the previous population. Alternatively,

$$s^2 = \frac{(\sum_f w_k x_k^g) \sum_f w_i (y_i - bx_i)^2 / x_i^g}{(\sum_f w_i y_i)^2} .$$

When the model holds, $e_i \approx \varepsilon_i$. That is one justification of our choices for the e_i and s^2. Another follows. If the selection probabilities were $\pi_i = n^* x_i^g / \sum_P x_k^g << 1$ for all i, then the relative randomization variance of t_C as an estimator for $\sum_F y_i$ under Poisson sampling (which is what NASS uses) would be roughly $[\sum_F (y_i - Bx_i)^2/\pi_i]/(\sum_F y_i)^2$, where $B = \sum_F y_i / \sum_F x_i$. This can be reasonably estimated with the sample actually drawn by $[\sum_f w_i (y_i - bx_i)^2/\pi_i]/(\sum_f w_i y_i)^2 = s^2/n^*$. Thus, our choice for defining s^2 is in some way robust to model failure.

We will call the a sample selection procedure where each $\pi_i = \min\{1, n_T x_i^g / \sum_P x_k^g\}$ and $\frac{1}{2} \le g \le 1$ "Brewer selection." This name applies whether or not the choice of n_T depends on σ in equation (5).

3. MULTIPLE TARGETS

Suppose we have M target variables, and y_{im} denotes the unit i y–value for the m'th target. Each target has its own (maybe unique, maybe not) control variable, and x_{im} denotes the unit i x–value for the m'th control. Furthermore, suppose each target/control pair is believed to obey the following model:

$$y_{im} = \beta_m (x_{im} + [\sum_P x_{km} / \sum_P x_{km}^g] x_{im}^g \varepsilon_{im}), \tag{7}$$

where $E(\varepsilon_{im} | x_{im}) = E(\varepsilon_{im}\varepsilon_{jm} | x_{im}, x_{jm}) = 0$ $(i \neq j)$, and $Var(\varepsilon_{im} | x_{im}) = \sigma_m^2$ for all m.

A set of weights, $\{a_i\}$, can often be constructed for a sample S that satisfies the M calibration equations

$$\sum_S a_i x_{im} = \sum_P x_{im}, \quad m = 1, ..., M,$$

such that every $a_i = \pi_i^{-1}[1 + O_P(1/\sqrt{n})]$, where π_i is (again) the selection probability of unit i. Each calibration estimator $t_{C(m)} = \sum_S a_i y_{im}$ provides a model unbiased estimator for $T_m = \sum_P y_{im}$ under the model in equation (7).

One potential way to construct these weights is with the formula inspired by linear regression:

$$a_i = \pi_i^{-1} + (\sum_P \mathbf{x}_k - \sum_S \pi_k^{-1} \mathbf{x}_k)(\sum_S c_k \pi_k^{-1} \mathbf{x}_k' \mathbf{x}_k)^{-1} c_i \pi_i^{-1} \mathbf{x}_i', \tag{8}$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{iM})$ is a row vector, and the choice for the c_i is arbitrary as long as $\sum_S c_k \pi_k^{-1} \mathbf{x}_k' \mathbf{x}_k$ is invertible. Popular choices are $c_i = 1/x_{i1}$ when M = 1 (so t_C becomes t from Section 1), and $c_i = 1$ (when one x_{im} is constant across i). Brewer (1994) suggests $c_i = (1 - \pi_i)/z_i$, where z_i is some composite measure of size across the M controls. We will return to this question of setting the c_i in Section 4.

Given target ACV's (denoted C_m) for all M target variables under the model in equation (7) and known σ-values (σ_m) for each variable, we can be assured of meeting these target ACV's when every

$$\pi_i = \min\{1, \max\{n_{T1} h_{i1}^{(g)}, ..., n_{TM} h_{iM}^{(g)}\}\}, \tag{9}$$

where $n_{Tm} = C_m / \sigma_m$, and $h_{im}^{(g)} = x_{im}^g / \sum_k x_{km}^g$.

Observe that π_i in equation (9) can also be expressed as

$$\pi_i = \max\{\pi_{i1}, ..., \pi_{iM}\}, \tag{10}$$

where $\pi_{im} = n_{Tm} h_{im}^{(g)}$ is Brewer selection for variable m. Consequently, the selection scheme in equation (10) can be called "Maximal Brewer Selection (MBS)." This name applies whether or not each target sample size n_{Tm} is set equal to C_m / σ_m.

4. APPLYING MBS

4.1. Poisson PRN Sampling

Brewer selection can be shown to minimize $ACV(t_C)$ for a fixed n_E under the model in equation (5) and conversely to minimize the expected sample size given a target ACV. Maximal Brewer selection when $M > 1$ does *not* necessarily minimize the expected overall sample size given M target ACV's. Sigman and Monsour (1995) sketch a method for determining selection probabilities that are optimal (i.e., expected-sample-size minimizing) in this sense.

Although not optimal, MBS is relatively simple and conveniently flexible when combined with Poisson Permanent-Random-Number (PRN) sampling (Ohlsson 1995 uses the term "PRN;" the concept can be found in Brewer et al. 1972). In such a design, every population unit i is independently assigned a random number p_i – a PRN – from the uniform distribution on the interval [0, 1). Unit i is selected for the sample if and only if $p_i < \pi_i$.

Poisson sampling, whether employing PRN's or not, has the well-known property that the joint selection probability of two distinct units i and k is equal to the product of their individual selection probabilities; that is, $\pi_{ik} = \pi_i\pi_k$. This greatly eases randomization variance estimation. This method of sampling also assures that $\sum_S z_i/\pi_i \approx \sum_P z_i$, since the relative variance of $\sum_S z_i/\pi_i$ is less than $(\sum_P z_i^2/\pi_i)/(\sum_P z_i)^2$, which is $O(1/n)$ under very mild restrictions on the z_i and π_i (see Isaki and Fuller 1982).

Poisson PRN sampling furthermore allows us to think of a sample drawn with MBS inclusion probabilities as the union of M Poisson PRN samples each drawn using the same PRN's and individual Brewer selection probabilities. This is convenient when we are interested in estimates of different combinations of target variables in different surveys.

For example, NASS makes estimates for potatoes in Minnesota in June and December, row crops (e.g., soybeans and corn) in March, June, and December, and small grains (e.g., wheat and barley) in March, June, September, and December. It wants to contact the same farms throughout the year, but has little interest in sampling a farm for the September survey if it has not historically had small grains. Thus, Poisson PRN samples of farms using the same PRN's can be drawn for potatoes, row crops, and small grains, each with its own Brewer selection probabilities. The union of all three is the overall sample in June. Similarly, the union of the row-crops and small-grains samples is the overall sample in March. Bailey and Kott (1997) discuss NASS's use of MBS and Poisson PRN sampling in Minnesota in greater detail.

Two additional points should be made at this time. One is that NASS actually draws the row-crops sample itself using MBS with individual row crops (soybeans, corn, etc.) serving as the target variables. The other is that MBS as practiced by the agency is the result of individual Brewer selections and Poisson PRN sampling. MBS is the cart and the individual Brewer selections the horse.

The overall MBS sample may not be the most efficient (expected-sample-size minimizing) way to meet multiple ACV targets. It is, however, the most efficient way of combining individual Brewer-selected samples.

4.2. A Count Control Variable

One potential target variable in an establishment survey is the number of units in P that still exist during the survey period. An obvious control variable for this target is unity, which can be assigned to each unit in P. Such a control is called a "count variable."

Whether or not the number of units still in existence is really of direct interest to survey managers, setting one component of $\mathbf{x}_i$, say x_{i0} equal to 1 for all i is a sensible policy. For one thing, it assures us that $t_{C(m)}$ will be randomization unbiased when $y_{im} > 0$, but $x_{im} = 0$; that is to say, when survey managers are surprised that unit i has a positive quantity of target variable m.

4.3. Calibration and Variance Estimation

NASS determines its calibration weights by first employing equation (8) with $c_i = 1 - \pi_i$. Brewer (1994) calls such a weighting scheme "cosmetic calibration," because the estimator can be put in prediction form ($t_{C(m)} = \sum_S y_{im} +$

$(\sum_P \mathbf{x}_i - \sum_S \mathbf{x}_i)\mathbf{b}_m$, where $\mathbf{b}_m$ is defined below equation (11)) when $\mathbf{x}_i$ contains a count-variable component. He argues that with cosmetic calibration individual weights rarely fall below unity. Weights below unity are deemed undesirable by many.

Under the weighting that results from employing equation (8) with $c_i = 1 - \pi_i$, when π_i is 1, a_i is also 1. Cosmetic calibration weights lower than unity, although rare, can still occur. NASS uses an iterative process described below that has, so far, successfully eliminated all weights less than unity. When plugging $c_i = 1 - \pi_i$ into equation (8) produces an $a_j < 1$, π_j in the equation is set equal to unity, and the equation run again for all i. This process is continued until all $a_i \geq 1$.

The estimator $t_{C(m)}$ is model unbiased not only under the model in equation (7), but also under the more general model:

$$y_{im} = \mathbf{x}_i\boldsymbol{\gamma}_m + u_{im},$$

where $\boldsymbol{\gamma}_m$ is an unspecified M-vector, and $E(u_{im}|\mathbf{x}_i) = 0$.

In order to be able to estimate the model variance of $t_{C(m)}$, we need to add the assumptions $E(u_{im}u_{jm}|\mathbf{x}_i, \mathbf{x}_j) = 0$, and $E(u_{im}^2|\mathbf{x}_i) = \sigma_{im}^2 < \infty$. In sharp contrast to the design stage, we are allowing the unit variances to be unspecified as long as they are finite.

Following the same reasoning that produced equation (3) leads to

$$E_\varepsilon[(t_{C(m)} - T_m)^2] = \sum_S a_i^2\sigma_{im}^2 - 2\sum_S a_i\sigma_{im}^2 + \sum_P \sigma_{im}^2. \qquad (3')$$

When n is large, we can make use of the near equalities $\sum_S a_i\sigma_{im}^2 \approx \sum_S \sigma_{im}^2/\pi_i \approx \sum_P \sigma_{im}^2$, and conclude

$$E_\varepsilon[(t_{C(m)} - T_m)^2] = \sum_S (a_i^2 - a_i)\sigma_{im}^2 < \sum_S a_i^2\sigma_{im}^2.$$

For a Poisson sample, the randomization mean squared error of $t_{C(m)}$ is

$$E_P[(t_{C(m)} - T_m)^2] \approx \sum_P \ddot{e}_{im}^2(\pi_i^{-1} - 1),$$

where $\ddot{e}_{im} = y_{im} - \mathbf{x}_i\mathbf{B}_m$, and $\mathbf{B}_m = (\sum_P c_k\mathbf{x}_k'\mathbf{x}_k)^{-1}\sum_P c_k\mathbf{x}_k'y_{km}$ (since $\sum_S a_iy_{im} - \sum_P y_{im} = \sum_S a_i\ddot{e}_{im} - \sum_P \ddot{e}_{im} = \sum_S \ddot{e}_{im}/\pi_i + (\sum_P \mathbf{x}_k - \sum_S \pi_k^{-1}\mathbf{x}_k)(\sum_S c_k\pi_k^{-1}\mathbf{x}_k'\mathbf{x}_k)^{-1}\sum_S c_i\pi_i^{-1}\mathbf{x}_i'\ddot{e}_{im} - \sum_P \ddot{e}_{im} \approx \sum_S \ddot{e}_{im}/\pi_i + (\sum_P \mathbf{x}_k - \sum_S \pi_k^{-1}\mathbf{x}_k)(\sum_S c_k\pi_k^{-1}\mathbf{x}_k'\mathbf{x}_k)^{-1}\sum_P c_i\mathbf{x}_i'\ddot{e}_{im} - \sum_P \ddot{e}_{im} = \sum_S \ddot{e}_{im}/\pi_i - \sum_P \ddot{e}_{im}$). When the c_k are all equal, the vector $\mathbf{B}_m$ is often called the "finite-population" or "census" regression coefficient.

An estimator for both the model variance and randomization mean squared of $t_{C(m)}$ is

$$v(t_{(m)}) = \sum_S (a_i^2 - a_i)e_{im}^2. \qquad (11)$$

where $e_{im} = y_{im} - \mathbf{x}_i\mathbf{b}_m$, and $\mathbf{b}_m = (\sum_S c_k\pi_k^{-1}\mathbf{x}_k'\mathbf{x}_k)^{-1}\sum_S c_k\pi_k^{-1}\mathbf{x}_k'y_{km}$ are the sample analogues of $\ddot{e}_{im}$ and $\mathbf{B}_m$, respectively.

4.4. The Delete-a-Group Jackknife

The problem with v in equation (11) is that is requires e_{im} to be calculated separately for each target variable. That is one reason why NASS uses a delete-a-group (DAG) jackknife variance estimator (Kott 1998). The DAG jackknife is also convenient when estimating the variances of domain totals and of ratios.

The Poisson sample is randomly divided into 15 replicate groups, denoted $S_1, S_2, ..., S_{15}$ (some groups can have one more member than others). The complement of each S_r is called the jackknife replicate group $S_{(r)} = S - S_r$. NASS then creates 15 sets of replicate weights. For the rth set: $a_{i(r)} = 0$ when $i \in S_r$; and

$$a_{i(r)} = a_i + (\sum_P \mathbf{x}_k - \sum_{S(r)} a_k\mathbf{x}_k)(\sum_{S(r)} c_ka_k\mathbf{x}_k'\mathbf{x}_k)^{-1}c_ia_i\mathbf{x}_i'$$

otherwise. This choice assures $a_{i(r)} \approx a_i$ for $i \in S_{(r)}$ when 15 is deemed large. Moreover, these two equalities will prove helpful. Under the model, because the ε_{im} are uncorrelated across units,

$$\sum_S a_{i(r)}\varepsilon_{im} - \sum_S a_i\varepsilon_{im} = -\sum_{Sr} a_i\varepsilon_{im} + (\sum_P \mathbf{x}_k - \sum_{S(r)} a_k\mathbf{x}_k)(\sum_{S(r)} c_k a_k \mathbf{x}_k{}'\mathbf{x}_k)^{-1}\sum_{S(r)} c_i a_i \mathbf{x}_i{}'\varepsilon_{im} \approx -\sum_{Sr} a_i\varepsilon_{im}.$$

Even without the model,

$$\begin{aligned}\sum_S a_{i(r)}\ddot{e}_{im} - \sum_S a_i\ddot{e}_{im} &= -\sum_{Sr} a_i\ddot{e}_{im} + (\sum_P \mathbf{x}_k - \sum_{S(r)} a_k\mathbf{x}_k)(\sum_{S(r)} c_k a_k \mathbf{x}_k{}'\mathbf{x}_k)^{-1}\sum_{S(r)} c_i a_i \mathbf{x}_i{}'\ddot{e}_{im} \\ &\approx -\sum_{Sr} a_i\ddot{e}_{im} + (\sum_P \mathbf{x}_k - \sum_{S(r)} a_k\mathbf{x}_k)(\sum_{S(r)} c_k a_k \mathbf{x}_k{}'\mathbf{x}_k)^{-1}\sum_{S(r)} c_i \pi_i^{-1} \mathbf{x}_i{}'\ddot{e}_{im} \\ &\approx -\sum_{Sr} a_i\ddot{e}_{im} + (\sum_P \mathbf{x}_k - \sum_{S(r)} a_k\mathbf{x}_k)(\sum_{S(r)} c_k a_k \mathbf{x}_k{}'\mathbf{x}_k)^{-1}\sum_{S} c_i \pi_i^{-1} \mathbf{x}_i{}'\ddot{e}_{im} \\ &\approx -\sum_{Sr} a_i\ddot{e}_{im} + (\sum_P \mathbf{x}_k - \sum_{S(r)} a_k\mathbf{x}_k)(\sum_{S(r)} c_k a_k \mathbf{x}_k{}'\mathbf{x}_k)^{-1}\sum_{P} c_i \mathbf{x}_i{}'\ddot{e}_{im} \\ &= -\sum_{Sr} a_i\ddot{e}_{im}\end{aligned}$$

when 15 is deemed large.

The DAG variance estimator for $t_{C(m)}$ is :

$$v_J(t_{C(m)}) = (14/15) \sum^{15} (\sum_S a_{i(r)} y_{im} - t_{C(m)})^2, \tag{12}$$

which WESVAR (Westat 1997) calls JK1.

It is easy to see that under the model in equation (7) and the error structure assumed above, the model expectation of $v_J(t_{C(m)})$ when 15 (as well as n) is assumed to be large is approximately $\sum_S a_i^2\sigma_{im}^2$ (since $\sum_S a_{i(r)}y_{im} - t_{C(m)} = \sum_S a_{i(r)}y_{im} - \sum_S a_i y_{im} = \sum_S a_{i(r)}\varepsilon_{im} - \sum_S a_i\varepsilon_{im} \approx -\sum_{Sr} a_i\varepsilon_{im}$).

We sketch below a proof that the randomization expectation of $v_J(t_{C(m)})$ is approximately $\sum_P \ddot{e}_{im}{}^2\pi_i^{-1}$ when $\sum_P \ddot{e}_{im} \approx 0$. This last near equality obtains exactly when $c_i = 1/(\boldsymbol{\gamma}\mathbf{x}_i')$ for some row vector $\boldsymbol{\gamma}$ (since then $\sum_P \ddot{e}_{im} = \sum_P \boldsymbol{\gamma}\mathbf{x}_i' c_i \ddot{e}_{im} = \boldsymbol{\gamma}\sum_P c_i\mathbf{x}_i'\ddot{e}_{im} = 0$). In practice, NASS does not deliberately choose a c_i with this property, however. This can cause the DAG jackknife to be randomization-biased. NASS sets $c_i = 1 - \pi_i$ and includes within $\mathbf{x}_i$ (for calibration purposes) a component $x_{i0} = 1$. Thus, when all the π_i are small, $c_i \approx 1 = 1/(\boldsymbol{\gamma}\mathbf{x}_i')$ for $\boldsymbol{\gamma} = (1, 0, \ldots, 0)$. When some π_i are large, the randomization mean squared error is smaller than $\sum_P \ddot{e}_{im}{}^2\pi_i^{-1}$, so whatever small bias in $v_J(t_{C(m)})$ is caused by $\sum_P \ddot{e}_{im}$ not being near zero is likely to be overwhelmed by $\sum_P \ddot{e}_{im}{}^2\pi_i^{-1}$ being larger than $\sum_P \ddot{e}_{im}{}^2(\pi_i^{-1} - 1)$.

Let n_r be the size of S_r. When 15 is large, $n/n_r \approx 15$ and $14/15 \approx 1$. The replicate group S_r can be viewed as a random subsample of S, and $q_r = (n/n_r)\sum_{Sr} \ddot{e}_{im}/\pi_i$ is a nearly randomization-unbiased estimator of $\sum_S \ddot{e}_{im}/\pi_i$, which is approximately 0 when $\sum_P \ddot{e}_{im} = 0$. The randomization variance of q_r with respect to the subsampling is approximately $(n/n_r)\sum_S \ddot{e}_{im}{}^2/\pi_i$ for each r. Now $\sum_S a_{i(r)}y_{im} - t_{C(m)} \approx \sum_S a_{i(r)}\ddot{e}_{im} - \sum_S a_i\ddot{e}_{im} \approx -\sum_{Sr} \ddot{e}_{im}/\pi_i \approx -q_r/15$. We can conclude that the randomization expectation of $v_J(t_{C(m)})$ in equation (12) with respect to the subsampling when $\sum_P \ddot{e}_{im} = 0$ is approximately $\sum_S \ddot{e}_{im}{}^2/\pi_i$.

5. MORE ON THE NASS EXPERIENCE

NASS prepares different samples in the various US states. NASS integrated its crops, stocks, and livestock surveys in the mid 1980s. Stratified simple random samples were drawn using a *priority* stratification scheme. For example, Stratum 1 might be large hog farms, Stratum 2 large crop farms that are not large hog farms, and so on, depending on the priorities of the target variables. Simple expansion estimates were generated from the sample data. Livestock variables were removed from the integrated Crops/Stocks (CS) Survey in the mid 1990s.

In the 1997/98 growing year, NASS drew a MBS Poisson PRN sample for the CS in one state, Minnesota. This proved very successful (Bailey and Kott 1997) . In 1998/99, this selection method was used in four states. By 1999/2000, 14 states had MBS Poisson PRN samples. Plans are to use MBS exclusively in the following year.

Rather than explicitly adding x_{i0} to the other x_{im} in selection equation (9), NASS has set a minimum value for π_i at roughly 0.01. For the most part, the same control variables have been used in the selection equation and the calibration

(equation (8)), although a count (intercept) variable has been added to every calibration. Figure 1 provides a chart of how many control variables were used in each of the 14 1999/2000 CS states.

NASS has set g in equation (9) equal to 0.75. Brewer (1999) seems to show a slight preference for g = .6. Table 1 reports estimated s-values in one state (PA) based on June 1999 survey data and various values for g. Crop and stock target variables for a single commodity (e.g., corn) use the same control value. One thing to notice is the s seems to increase as the fraction of the sample with positive x-values (called "the commodity population") and positive y-values decreases. The second is that NASS's choice of g = .75 everywhere needs to be explored more thoroughly. In principle, the best choice for g minimizes s asymptotically.

Nonresponse has been handled using the pre-existing imputation scheme, which relies on the old priority stratification. DAG jackknife variances are estimated treating non-response as a second phase of sampling and pretending that respondents were reweighted using the priority strata as the reweighting groups. If the models supporting the imputation scheme are correct, this will (if anything) bias mean squared error estimates upward.

6. COMMENTS

The change in the Crops/Stocks Survey from an estimation strategy featuring stratified simple random sampling and a simple expansion estimator to Poisson PRN sampling with maximal Brewer selection probabilities and a (cosmetic) calibration estimator has proven very successful at NASS. The Agency is currently exploring the use of the new strategy in other surveys as well. In the interest of honest disclosure, NASS actually uses collocated sampling (Brewer et al., 1972) sampling rather than Poisson sampling. This modestly reduces the sample-size variability. Mean squared errors are estimated as if Poisson samples were drawn.

Kott and Fetter (1999) show how Poisson PRN sampling can easily to adapted to limit the number of times a single unit is selected across co-ordinated surveys. Let $\pi_i^{(q)}$ be the unit i selection probability for survey q (= 1, 2, ... Q). Unit i is in the sample for survey q when its PRN, p_i, is in the interval $[\tau_{i,q-1}, \tau_{i,q})$, where $\tau_{i,0} = 0$, and $\tau_{i,f} = \pi_i^{(1)} + ... + \pi_i^{(f)}$. For this *sequential interval Poisson* (SIP) sampling methodology described above to work, $\tau_{i,q}$ cannot exceed unity. Fortunately, it is a simple matter to generalize SIP sampling a bit. We can redefine $\tau_{i,f}$ as $\pi_i^{(1)} + ... + \pi_i^{(f)} - I_{(i)}$, where $I_{(i)}$ is the largest integer less than $\pi_i^{(1)} + ... + \pi_i^{(f)}$. When $\tau_{i,q-1} > \tau_{i,q}$, the interval $[\tau_{i,q-1}, \tau_{i,q})$ is similarly redefined as the union of $[\tau_{i,q-1}, 1)$ and $[0, \tau_{i,q})$.

The larger $I_{(i)}$ the greater the number of survey samples in which unit i can find itself (that number will either be $I_{(i)}$ or $I_{(i)} + 1$) . This is another reason NASS needs to explore the value at which g in equation (9) is set. The smaller the value, the less likely a particular unit with large control values with be selected for a sample.

It may also be that the best choice for g varies by target variable. Worse, $Var(\varepsilon_{im}) \propto x_{im}^{\,g}$ may not even be the appropriate specification. Oddly, this widely used specification began as an approximation of $ax_{im} + bx_{im}^{\,2}$ (see Cochran, 1963, p. 256), which has prompted the belief that ½ must be the lower bound of g in practice. In the NASS application, the quality of control information is better for larger values. Consequently, it is possible that the best g for some target variables is, in fact, less than ½.

REFERENCES

Bailey, J. T. and P.S. Kott (1997), "An Application of Multiple List Frame Sampling for Multi-purpose Surveys, *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 496-500.

Bosecker, R. (1989), *Integrated Agricultural Surveys*, NASS Research Report Number SSB-89-05, Washington, DC: National Agricultural Statistics Service.

Brewer, K.R.W. (1963), "Ratio Estimation and Finite Populations: Some Results Deductible from the Assumption of an Underlying Stochastic Process," *Australian Journal of Statistics*, **5**, pp. 93-105.

Brewer, K.R.W. (1994), "Survey Sampling Inference: Some Past Perspectives and Present Prospects," *Pakistan Journal of Statistics,* **10(1)A**, pp. 213-233.

Brewer, K.R.W. (1999), "Cosmetic Calibration with Unequal Probability Sampling," *Survey Methodology*, forthcoming.

Brewer, K.R.W., Early, L.J., and Joyce, S.F. (1972), "Selecting Several Samples from a Single Population," *Australian Journal of Statistics*, **14**, pp. 231-239.

Cochran, W.G. (1963), *Sampling Techniques*, 2nd Edition, New York: Wiley.

Deville, J-C. and Särndal, C-E. (1992), "Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, **87**, pp. 376-382.

Godambe, V.P. (1955), "A Unified Theory of Sampling from Finite Populations,' *Journal of the Royal Statistical Society*, **B17**, pp. 269-278.

Isaki, and Fuller, W.A. (1982), "Survey Design Under the Regression Superpopulation Model," *Journal of the American Statistical Association*, **77**, 89-96.

Kott, P. S. (1998). *Using the Delete-A-Group Jackknife Variance Estimator in NASS Surveys*, RD Research Report No. RD-98-01, Washington, DC: National Agricultural Statistics Service.

Kott, P.S. and Fetter, M.J. (1999), "Using Multi-phase Sampling to Limit Respondent Burden Across Agriculture Surveys," *Proceedings of the Survey Methods Section, Statistical Society of Canada*, forthcoming.

Ohlsson, E. (1995), "Coordination of Samples Using Permanent Random Numbers," in Cox, B.G., Binder, D.A., Chinnappa, N., Christianson A, Colledge, M.J., and Kott, P.S. (eds) *Business Survey Methods*, New York: Wiley, pp. 153-169.

Royall, R. M. (1970), " On Finite Population Sampling Under Certain Linear Regression Models," *Biometrika* **57**, pp. 377-387.

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

Sigman, R.S. and Monsour, N. J. (1995), " Selecting Samples from List Frames of Businesses," in Cox, B.G., Binder, D.A., Chinnappa, N., Christianson A, Colledge, M.J., and Kott, P.S. (eds) *Business Survey Methods*, New York: Wiley, pp. 133-152.

Westat, Inc. (1997), *A User's Guide to WesVarPC®, Version 2.1*, Rockville, MD: Westat.

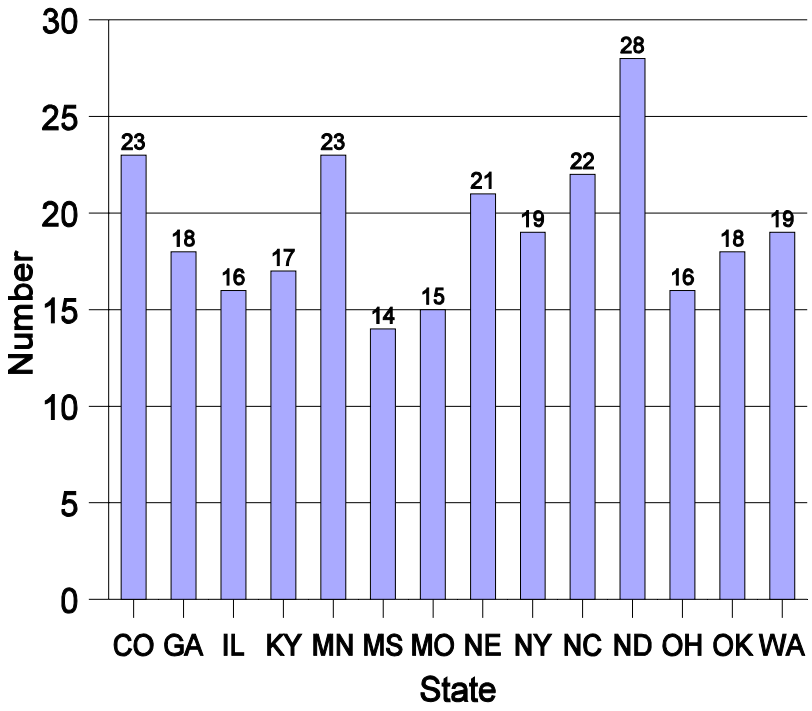

Figure 1: Number of Control Variables in Each State

Table 1: Target Variable Calculations in PA With Different Values for g

Commodity	Commodity Population	Survey: Response Rate	Survey: Positive Reports	Survey: % with Item	s with g = 0.5	s with g = 0.6	s with g =0.75	s with g = 0.9
Alfalfa Acres	18006	84.3%	372	60.7%	1.26	1.26	1.27	1.31
Wheat Stocks	8079	84.4%	29	6.2%	16.88	14.93	12.59	10.82
Barley Acres	5206	84.3%	122	46.0%	1.39	1.40	1.45	1.55
Corn Stocks	21268	82.4%	314	36.1%	2.64	2.51	2.43	2.47
Corn Acres	21268	84.3%	559	78.5%	0.75	0.74	0.76	0.81
Oat Stocks	11824	84.4%	114	22.2%	2.81	2.85	2.95	3.13
Oat Acres	11824	84.3%	250	54.8%	1.39	1.41	1.47	1.55
Other Hay	19478	84.3%	446	65.5%	1.30	1.28	1.27	1.31
Potato Acres	829	84.3%	67	59.4%	0.82	0.82	0.87	0.99
Rye Acres	4210	84.3%	103	40.5%	1.95	1.98	2.08	2.24
Soybean Stocks	7030	83.9%	79	18.4%	3.90	3.73	3.56	3.48
Soybean Acres	7030	84.3%	234	67.1%	0.95	0.95	0.96	1.00
Tobacco Acres	979	84.3%	9	33.3%	1.33	1.35	1.41	1.48
Wheat Acres	3836	84.3%	230	67.9%	0.98	1.03	1.14	1.29

ESTIMATORS FOR USE WITH POISSON SAMPLING AND RELATED SELECTION PROCEDURES

K.R.W. Brewer, Australian National University and Timothy G. Gregoire, Yale University
Ken Brewer, Dept of Statistics and Econometrics, ANU, ACT 0200, Australia. <Ken.Brewer.anu.edu.au>

ABSTRACT

Estimators of total devised for use with Poisson Sampling and related selection procedures have so far been treated as completely unconditional. Their variances have been defined and estimated over the complete range of all possible samples. In this paper we analyze Poisson and collocated samples conditionally on their achieved sample size, as though that size had been fixed beforehand. The performances of the conditional Horvitz-Thompson estimator (known to foresters as 3P7) and of some other related estimators are investigated in this context, both analytically and empirically. The consequences of using these estimators in the contexts of randomly and deliberately ordered systematic sampling are also subjected to scrutiny.

Key Words: Conditional estimation; Cosmetic calibration; Forest surveys; Horvitz-Thompson estimator; Regression estimation; Systematic sampling.

1. UNCONDITIONAL AND CONDITIONAL DESIGN-BASED INFERENCE

Design-based inference is usually unconditional. Expectations and variances are defined over the set of all possible samples that could have been selected given the selection procedure used. Where the sample size n is a random variable, however, estimates of the unconditional variance or mean squared error are not appropriate measures of spread for the particular sample estimate obtained. In Poisson sampling, for instance, the sample size can range from zero to N, (the population size), and the Horvitz-Thompson (HT) estimator , $\hat{Y}_{\bullet HT}$, of the population total $Y_\bullet$, defined by

$$\hat{Y}_{\bullet HT} = \Sigma_{i \in s} Y_i \pi_i^{-1}, \tag{1}$$

where Y_i is the value of the item y for the ith population unit, π_i is its (*a priori*) first order inclusion probability and s is the set of population units included in sample, tends to be proportional to n. Because of this, the design-variance of $\hat{Y}_{\bullet HT}$, defined by

$$V^2(\hat{Y}_{\bullet HT}) = E(\hat{Y}_{\bullet HT} - Y_{\bullet HT})^2, \tag{2}$$

where E is the expectation operator over all possible samples, is typically far more dependent on the variability of the sample size than on the variability in the contributions $Y_i \pi_i^{-1}$.

The estimator referred to in Forestry as 3P7 (Grosenbaugh 1964, 1965), namely

$$\hat{Y}_{\bullet 3P7} = \{E(n)/n\} \Sigma_{i \in s} Y_i \pi_i^{-1} \tag{3}$$

which compensates for the random variability in n, is therefore far preferable to $\hat{Y}_{\bullet HT}$. It can be viewed as the special case of the HT ratio estimator

$$\hat{Y}_{\bullet HTR} = \{X_\bullet / \hat{X}_{\bullet HT}\} \hat{Y}_{\bullet HT}, \tag{4}$$

(Brewer 1963) when the auxiliary variable X_i is equal to π_i for all i. The estimator of variance suggested for it by Brewer, Early and Joyce (1972) was the standard one based on the Taylor approximation to the variance of this ratio estimator. It was, therefore, an estimator of the unconditional variance of that ratio estimator, defined over all possible Poisson samples that could be selected with those π_i, regardless of the achieved sample size. Estimates based on (3) and (4), however, are still likely to be more accurate if $n > E(n)$, than if $n < E(n)$, so the standard variance estimator is more relevant to a sample of size $E(n)$ than to one of size n.

The other obvious way in which 3P7 can be interpreted is as an HT estimator conditioned on the achieved sample size, n. In order for this interpretation to be meaningful, however, it is necessary to postulate the existence of "adjusted inclusion probabilities" (Furnival, Gregoire and Grosenbaugh 1987). The relationship between the actual inclusion probability π_i and the adjusted inclusion probability $\breve{\pi}_i$ is

$$\breve{\pi}_i = \{n / E(n)\} \pi_i, \tag{5}$$

and the 3P7 estimator can conveniently be written as

$$\hat{Y}_{\bullet 3P7} = \Sigma_{i \in s} Y_i \breve{\pi}_i^{-1}. \qquad (6)$$

This is in the form of the HT estimator, but with $\breve{\pi}_i$ taking over the role of π_i. In this paper we will be using the estimator (6) and estimating its variance conditional on the achieved sample size, n, using the adjusted inclusion probabilities, $\breve{\pi}_i$, as though they had been the actual inclusion probabilities.

(Simple random sampling without replacement (*srswor*) theory provides a close analogy to the analysis described here in its use of the poststratification technique. The poststratified estimator is technically a ratio estimator with the auxiliary variable X_i set equal to unity for all i within the stratum, but it is usually analyzed as though the sample within the poststratum had been selected using *srswor* with the sample size fixed.)

2. ESTIMATION OF THE DESIGN-VARIANCE CONDITIONED ON ACHIEVED SAMPLE SIZE

There is a serious problem associated with estimating the variance of the HT estimator, namely that the traditional variance estimators (Horvitz and Thompson 1952, Sen 1953, Yates and Grundy 1953) are crucially dependent on knowledge of the joint or second order inclusion probabilities of pairs of units in sample, π_{ij}. These quantities are usually awkward to evaluate and it is common practice to circumvent the need for them either by invoking the jackknife or by using some other indirect approach (Särndal 1996, Brewer 1999).

In the "conditioned estimation" context described towards the end of Section 1 the problem is even sharper in that the "adjusted" first order inclusion probabilities are themselves fictitious. Any attempt to create similarly fictitious "adjusted" second order inclusion probabilities is bound to be arbitrary (and therefore suspect) and is also likely to be awkward into the bargain.

Here we have avoided this problem by finding estimators of an approximation to what we call the "natural" variance of the HT estimator. By this we mean the variance that estimator has when the π_{ij} are not deliberately manipulated to reduce the HT variance below (or, perversely, to increase it above) that "natural" level. The basic variance estimator we use differs from equation (2.3) in Deville (1999), but has features in common with it and is similar in intent. For the details, see Brewer (2000).

The most important case where the HT estimator does not have that "natural" value is where the sample is selected systematically with unequal probabilities from a deliberately ordered population. This case is also dealt with separately in the same reference.

3. A MONTE CARLO EXERCISE

There are two principal dangers that could arise in the use of the approach described in Sections 1 and 2. Both these dangers stem from the fact that the randomization distribution of an unequal probability sample selected with expected size $E(n)$ and achieved size $n \neq E(n)$ differs from that of an unequal probability sample drawn with fixed size n. When $n < E(n)$, this is typically because fewer small units have been chosen than would have been expected, and *vice versa*.

Suppose then that we observe $n < E(n)$. Let Y_i be the actual survey variable of interest and X_i a known measure believed to be roughly proportional to it. The first danger is that if (as is frequently the case with populations of establishments) the smaller units are relatively more variable than the larger ones in their values of Y_i / X_i, a relative scarcity of small units in the sample might lead to some underestimation of the variance. The second and more important danger occurs when Y_i / X_i is not roughly constant over the whole range of X_i but actually tends to increase (decrease) with X_i. Then a relative scarcity of small units in the sample can lead to an over(under)estimate of that ratio itself, and the contribution to MSE made by squared bias could become important.

Poisson sampling is frequently used in forestry, so this is an important special case. The context in which comparisons are being made here is that of sampling trees from a forest; in particular, a population of 14,443 loblolly pine trees from Alabama (Gregoire and Williams 1992, Gregoire and Valentine 1998) with an aggregate bole volume of $Y_\bullet$ = 8,985.8m^3. Consider the case when $X_i \equiv A_i$, where A_i is the cross-sectional area of the stem at breast height. The Y_i / A_i values in this population are nearly homoskedastic. Homoskedasticity in Y_i / A_i eliminates the first of the two problems described above but the second remains.

Since Y_i / A_i tends to increase with A_i the ratio estimation bias can be appreciable. To remedy this we experimented with different powers of basal area. We were able to establish empirically that $A^{4/3}$, was a suitable choice for this population of loblolly pine trees. In addition to basal area, we had records of tree heights, but these were poorly determined in comparison with basal area. We therefore used a power of height as a secondary supplementary variable when using regression estimation. Denoting height by H, we had both $Y \propto AH$ and $Y \propto A^{4/3}$, from which it followed logically that $Y \propto H^4$, a relationship that accorded well with our data, so we initially chose H^4 as our secondary supplementary variable; i.e. we used $X_{1i} = A_i^{4/3}$ and $X_{2i} = H_i^4$. (Other species were found to require different models.) We found later, however, that it was even better to use the product of squared diameter and height as a single regressor. This "cylindrical" model could well be appropriate for other species, also.

We used four selection procedures, Deliberately Ordered Systematic (DS), Randomly Ordered Systematic (RS), Poisson (PS), and Collocated (CS). In each case the sampling was with inclusion probabilities proportional to size, $\pi_i \propto A_i$, and without replacement. (In no instance did the largest A_i on the population exceed the skip interval.) The DS samples were selected from the population of 14,443 trees described above, in ascending order of A_i. The RS samples were selected from the same population, but randomly ordered. In Poisson sampling, the ith population unit is independently selected with probability π_i. It is allotted a random number r_i in the interval [0,1) and included in sample if and only if $r_i < \pi_i$. For Collocated sampling the r_i are uniformly spaced over [0,1) instead (Brewer, Early and Hanif 1984).

For each of these four selection procedures we selected from this population 100,000 samples with E(n) = 144 units; and we did the same for samples with E(n) = 48 and E(n) =24 units also. To be circumspect, we repeated the simulation samplings with different random number seeds. Results from these repeated runs varied little; the results presented in the following section were each obtained from a single run of 100,000 samples for each expected sample size.

(The RS and CS runs involved a complete new random ordering of the population prior to the selection of each of the 100,000 samples. Experimental runs were also obtained for RS using the same random population ordering for each sample. The observed (or Monte Carlo) root mean squared errors (RMSEs) differed from ordering to ordering but were typically lower by about 5%, as the first stage component of variance, that between orderings, had been eliminated. The experiment was dropped when the estimated variances were found to be no smaller than for the standard RS runs, and it was realised that the more meaningful definition of the RS variance included the first stage component.)

We next calculated the means and variances (over each set of 100,000 samples) for a number of different estimators of total volume, $Y_\bullet$. The first was the 3P7 of (6) above. The remaining estimators all had certain features in common. They were generalized regression estimators (GREGs) and they were almost in the standard model-assisted survey sampling form (Särndal, Swensson and Wretman 1992 Section 6.4). The only differences from that standard form were (i) that the π_i^{-1} in the formula for the implicit regression coefficient was replaced by $(\pi_i^{-1} - 1)$ and (ii) that Deville and Särndal's (1992) q_k weight, which in this paper is represented by z_i, was chosen to be in the sample space of the regressor variables. These two features combined ensured that the GREG estimators were *cosmetically calibrated* or interpretable in terms of design-based and prediction-based inference simultaneously (Brewer 1999). The formula for a cosmetically calibrated estimator of $Y_\bullet$ is

$$\hat{Y}_{\bullet COSCAL} = \mathbf{1}_n' \boldsymbol{\Pi}_s^{-1} \mathbf{Y}_s + (\mathbf{1}_N' \mathbf{X} - \mathbf{1}_n' \boldsymbol{\Pi}_s^{-1} \mathbf{X}_s) \hat{\boldsymbol{\beta}} \qquad (7)$$

where $$\hat{\boldsymbol{\beta}} = \{\mathbf{X}_s'\mathbf{Z}_s^{-1}(\Pi_s^{-1} - \mathbf{I}_n)\mathbf{X}_s\}^{-1}\{\mathbf{X}_s'\mathbf{Z}_s^{-1}(\Pi_s^{-1} - \mathbf{I}_n)\mathbf{Y}_s\}. \quad (8)$$

Here $\boldsymbol{\Pi}_s$ is the diagonal matrix of the sample π_i, $\mathbf{Y}_s$ is the column vector of the sample y_i, and $\mathbf{X}$, $\mathbf{X}_s$ are the $N \times p$, $n \times p$ population and sample matrices, respectively, of the regressor or supplementary variables.

The cosmetically calibrated estimator is on average marginally less efficient than the standard GREG. Its interest lies partly in the fact that it is readily interpretable both as a design-based and as a prediction-based estimator (Brewer 1999) and partly in the ease with which negative and other undesirably small sample weights can be removed (Brewer 1999, Kott and Bailey 2000). The fact that the efficiencies of the estimators are hardly affected by the choice of $\mathbf{Z}$ is a probable indication that the choice of $\mathbf{Z}$ to be a linear function of the column sums of $\mathbf{X}$ (a necessary requirement for cosmetic calibration) involves no appreciable penalty.

Seven specific estimators in this category are set out below, with their regressor variable sets indicated in brackets: Coscal $(1, A)$ and Coscal (A, H), both with $z_i = A_i$; Coscal $(A^{4/3})$, Coscal $(1, A^{4/3})$, Coscal $(A^{4/3}, H^4)$ and Coscal $(1, A^{4/3}, H^4)$, all four with $z_i = A_i^{4/3}$; and Coscal (D^2H) with $z_i = D^2H_i$, which corresponds to the cylindrical model mentioned earlier. Of these, the first two are the only survivors from a long list of potential alternative estimators, retained less for their performance than for their *a priori* attractiveness. The next two are the best performing estimators that we could construct using only the basal area data. The last three were the best we could construct using all the data available to us, namely basal area and height. In this paper we shall be focusing particularly on 3P7 and the coscal estimator with the cylindrical model.

Since the primary aim of this exercise was to determine whether it was possible to estimate with acceptable accuracy the variance of an estimator conditioned on n, all the variance estimators used were either already suited to this requirement or constructed specifically to meet it. In the first category there were two estimators that we denote by vS and BEJ, both of which had originally been devised to estimate the variance of the estimator 3P7, vS by foresters (Furnival, Gregoire and Grosenbaugh 1987), and BEJ by statisticians (Brewer, Early and Joyce 1972).

The other variance estimators used were specifically devised for this exercise in Brewer (2000). For selection procedures other than DS, a set of variance estimators has been used that is appropriate for the situation where the joint inclusion probabilities have not been manipulated with the aim of reducing the variance of the estimator of total, but retain their "natural" values. The simplest of these other estimators has been adapted from equations (6) and (8) of Brewer (2000). Its formula is

$$\hat{\tilde{V}}^2(\hat{Y}_{\bullet 3P7}) = \Sigma_{i=1}^n (c^{-1} - \breve{\pi}_i)(Y_i\breve{\pi}_i^{-1} - \hat{Y}_{\bullet 3P7}n^{-1})^2 \quad (9)$$

where $\breve{\pi}_i = \pi_i n / E(n)$ is the adjusted inclusion probability and $c = (n-1)/(n - n^{-1}\Sigma_{k=1}^N \pi_k^2)$.

The remaining estimators in this group resemble $\hat{\tilde{V}}^2(\hat{Y}_{\bullet 3P7})$ closely, but in each case the factor $(Y_i\breve{\pi}_i^{-1} - Y_{\bullet}n^{-1})^2$ is replaced by a factor $(Y_i - \hat{\bar{Y}}_i)^2\breve{\pi}_i^{-2}$ where $\hat{\bar{Y}}_i$ is the prediction model estimate of a prediction expectation $\bar{Y}_i$ of Y_i. For instance, in the formula for $\hat{\tilde{V}}^2(1, A)$, the value of $\hat{\bar{Y}}_i$ is arrived at by using Coscal $(1, A)$, the cosmetically calibrated prediction model with the intercept term and the basal area as the regressor variables. Special forms of these estimators are also given in Brewer (2000) for use with DS, where the ordinary variance estimators could be expected to overestimate severely. These are:

$$\hat{\tilde{V}}_{DS}^2(\hat{Y}_{\bullet 3P7}) = \frac{n}{2(n-1)} \Sigma_{k=1}^{n-1} \{\frac{2 - \pi_k - \pi_{k+1}}{2}\}(Y_k\pi_k^{-1} - Y_{k+1}\pi_{k+1}^{-1})^2 \quad (10)$$

and

$$\hat{\tilde{V}}_{DS}^2(\hat{Y}_{\bullet COSCAL}) = \frac{n}{2(n-1)} \cdot \frac{n}{n-p} \Sigma_{k=1}^{n-1} \{\frac{2 - \pi_k - \pi_{k+1}}{2}\} \{(Y_k - \hat{\bar{Y}}_k)\pi_k^{-1} - (Y_{k+1} - \hat{\bar{Y}}_{k+1})\ \pi_{k+1}^{-1}\}^2. \quad (11)$$

Here $\hat{\bar{Y}}_k = \mathbf{X}_k\hat{\boldsymbol{\beta}}$ is the prediction estimator of the prediction mean $\bar{Y}_k = \mathbf{X}_k\boldsymbol{\beta}$ of Y_k.

4. SOME EMPIRICAL RESULTS

The standard deviations of n for E(n) = 144 are 12.0 for Poisson and 7.9 for collocated sampling. The corresponding figures for E(n) = 48 are 6.9 and 4.0. For E(n) =24 they are 4.9 and 3.2.

Table 1. Observed percentage bias of Horvitz-Thompson and coscal estimators of total volume based on 100, 000 samples for each combination of sampling design and expected sample size.

Estimator	Poisson	Collocated	Randomly Ordered Systematic	Deliberately Ordered Systematic
---------- E(n) = 144 ----------				
$\hat{Y}_{\bullet HT}$	0.02	-0.02	0.00	0.00
$\hat{Y}_{\bullet 3P7}$	0.00	-0.05	0.00	0.00
Coscal (1, A)	-0.07	-0.08	-0.07	0.00
Coscal ($A^{4/3}$, H^4)	0.11	0.09	0.11	0.08
Coscal (1, $A^{4/3}$, H^4)	0.12	0.11	0.12	0.09
Coscal (D^2H)	-0.01	-0.02	0.00	-0.01
---------- E(n) = 48 ----------				
$\hat{Y}_{\bullet HT}$	0.02	-0.02	-0.03	0.00
$\hat{Y}_{\bullet 3P7}$	-0.02	-0.07	-0.03	0.00
Coscal (1, A)	-0.27	-0.21	-0.28	-0.05
Coscal ($A^{4/3}$, H^4)	0.32	0.33	0.32	0.29
Coscal (1, $A^{4/3}$, H^4)	0.36	0.37	0.36	0.29
Coscal (D^2H)	-0.03	-0.01	-0.03	0.00
---------- E(n) = 24 ----------				
$\hat{Y}_{\bullet HT}$	0.01	0.01	0.00	-0.01
$\hat{Y}_{\bullet 3P7}$	-0.01	-0.18	0.00	-0.01
Coscal (1, A)	-0.59	-0.47	-0.59	-0.18
Coscal ($A^{4/3}$, H^4)	0.67	0.63	0.61	0.55
Coscal (1, $A^{4/3}$, H^4)	0.78	0.73	0.71	0.55
Coscal (D^2H)	-0.03	-0.04	-0.06	-0.04

The largest observed Monte Carlo biases in the estimator of total are (in absolute magnitude) 0.12% for E(n) = 144, 0.37% for E(n) = 48 and 0.78% for E(n) = 24 (Table 1). There is a strong tendency for them to be positive, especially for the more accurate estimators. The largest percentage contributions of the squared bias term to the MSE are about 1% for E(n) = 144, 5% for E(n) = 48 and 14% for E(n) = 24.

Table 2 displays the root mean square errors (RMSE), expressed as a percentage of $Y_{\bullet}$. The observed RMSEs of 3P7 are of the same order of magnitude for all selection procedures, but naturally lowest for DS. GREG estimation using supplementary variables is capable of reducing the RMSEs well below the 3P7 levels. Regression on D^2H is signally superior to any alternative. The addition of an intercept term appears to increase the biases slightly while hardly changing the RMSEs. Adding an intercept term to the regression on $A^{4/3}$ by itself is also somewhat

counterproductive. In other circumstances, however, the addition of an intercept term can help considerably. One possible inference is that for a good or even reasonably good model, adding the intercept term is unlikely to help, but that adding the intercept to a poor model can sometimes make it better than certain reasonably good models. The topic seems to be worth investigating further.

Table 2: Observed percentage root mean square error of Horvitz-Thompson and coscal estimators of total volume based on 100,000 samples for each combination of sampling design and expected sample size.

Estimator	Poisson	Collocated	Randomly Ordered Systematic	Deliberately Ordered Systematic
---------- E(n) = 144 ----------				
$\hat{Y}_{\bullet HT}$	8.72	6.62	2.87	2.12
$\hat{Y}_{\bullet 3P7}$	2.87	2.73	2.87	2.12
Coscal (1, A)	2.26	2.26	2.26	2.15
Coscal ($A^{4/3}$, H^4)	1.61	1.60	1.61	1.67
Coscal (1, $A^{4/3}$, H^4)	1.59	1.59	1.59	1.69
Coscal (D^2H)	1.42	1.41	1.41	1.48
---------- E(n) = 48 ----------				
$\hat{Y}_{\bullet HT}$	15.07	11.51	5.00	3.74
$\hat{Y}_{\bullet 3P7}$	5.04	4.76	5.00	3.74
Coscal (1, A)	4.01	3.96	3.97	3.67
Coscal ($A^{4/3}$, H^4)	2.86	2.83	2.84	2.90
Coscal (1, $A^{4/3}$, H^4)	2.83	2.81	2.81	2.92
Coscal (D^2H)	2.50	2.50	2.50	2.49
---------- E(n) = 24 ----------				
$\hat{Y}_{\bullet HT}$	21.47	16.30	7.07	5.36
$\hat{Y}_{\bullet 3P7}$	7.22	6.75	7.07	5.36
Coscal (1, A)	5.91	5.72	5.79	5.21
Coscal ($A^{4/3}$, H^4)	4.18	4.08	4.07	3.90
Coscal (1, $A^{4/3}$, H^4)	4.22	4.09	4.08	3.92
Coscal (D^2H)	3.63	3.56	3.54	3.41

The 3P7 variance estimators—vS, BEJ and $\hat{\hat{V}}^2(\hat{Y}_{\bullet 3P7})$—vary in performance from one selection procedure to another (See Table 3). $\hat{\hat{V}}^2(\hat{Y}_{\bullet 3P7})$ in particular is nearly unbiased for PS and RS at all three sample sizes. vS and BEJ are nearly unbiased for PS and RS at the largest E(n). The difference between PS and CS is that under PS the π_{ij} all take the "natural" value $\pi_i\pi_j$, but for Collocated sampling those π_{ij} are modified to a small extent. Hence there is a difference between the PS and the CS variances of the order of 10%, even when conditioned on the achieved sample size, and because this is not reflected in the variance estimator, vS and BEJ overestimate the CS variance to that extent.

The $\hat{\hat{V}}^2$ variance estimators tend to underestimate variance (and this is more noticeable for the smaller sample sizes). This seems to be because the factor $n/(n-p)$ is insufficient for bias correction when the sampling is not

simple random. If so, the problem could be removed by using model-unbiased estimators of the individual squared residuals. The $\hat{v}^2$ (D^2H) estimator is based on this principle, and it seems to work well; we intend to introduce this estimator of variance in greater detail in a later version of this paper.

Table 3: Observed percentage bias of variance estimators of total volume, based on 100,000 samples for each combination of sampling design and expected sample size.

Estimator	Poisson	Collocated	Randomly Ordered Systematic	Deliberately Ordered Systematic
---------- E(n) = 144 ----------				
vS	0.6	11.5	0.6	84.2
BEJ	-1.0	9.8	-1.0	81.4
$\hat{\tilde{V}}^2(\hat{Y}_{\bullet 3P7})$	0.4	11.0	-0.3	2.1
$\hat{\tilde{V}}^2(A^{4/3}, H^4)$	-1.8	-1.1	-1.7	-17.2
$\hat{v}^2$ (D^2H)	0.1	0.7	0.0	-6.7
---------- E(n) = 48 ----------				
vS	-1.4	10.9	0.0	81.6
BEJ	-3.8	8.2	-2.3	77.4
$\hat{\tilde{V}}^2(\hat{Y}_{\bullet 3P7})$	0.4	11.7	-0.3	3.2
$\hat{\tilde{V}}^2$ ($A^{4/3}$, H^4)	-4.1	-3.5	-4.8	-15.6
$\hat{v}^2$ (D^2H)	0.1	1.4	0.1	1.7
---------- E(n) = 24 ----------				
vS	-4.0	9.5	-0.4	78.1
BEJ	-8.3	4.7	-4.0	70.4
$\hat{\tilde{V}}^2(\hat{Y}_{\bullet 3P7})$	0.2	11.6	-0.2	2.8
$\hat{\tilde{V}}^2$ ($A^{4/3}$, H^4)	-8.7	-6.4	-7.1	-1.7
$\hat{v}^2$ (D^2H)	-0.4	1.6	0.8	9.3

5. TESTING WHETHER THE ESTIMATORS MEASURE CONDITIONAL VARIANCE

The estimators selected for this analysis were 3P7 regarded as an conditional version of the Horvitz-Thompson estimator and of the best performing of the Coscal estimators, namely Coscal (D^2H). Results are presented for individual quintiles of the sample sizes in Tables 4 through 9. They confirm that the conditional variance estimators used in this study do faithfully reflect the variability of the conditional variances with n. They also indicate, however, the price being paid for the reduction in variance when using CS instead of PS; there is an appreciable quintile-specific bias to the 3P7 estimator, as high as 2% for E(n) = 24. For Coscal (D^2H), however, the CS bias is more consistent over quintiles, smaller, and often actually smaller than the quintile-specific biases to be found in PS.

Table 4: Performance of $\hat{Y}_{\bullet 3P7}$ and Coscal (D^2H) estimators of total volume based on 100, 000 Poisson samples of expected size E(n) = 144; results are summarized by quintiles of increasing sample size.

Quintile	Average n	$\hat{Y}_{\bullet 3P7}$				Coscal (D^2H)			
		Bias%	SE%	Est. SE%	RMSE%	Bias%	SE%	Est. SE%	RMSE%
1	127.6	0.04	3.04	3.05	3.04	-0.01	1.53	1.53	1.53
2	137.6	0.03	2.93	2.93	2.93	-0.01	1.47	1.47	1.47
3	143.9	0.00	2.87	2.87	2.87	-0.01	1.44	1.43	1.44
4	150.2	-0.02	2.78	2.81	2.78	-0.02	1.40	1.40	1.40
5	160.9	-0.03	2.72	2.71	2.72	0.00	1.36	1.36	1.36

Table 5: Performance of $\hat{Y}_{\bullet 3P7}$ and Coscal (D^2H) estimators of total volume based on 100, 000 collocated samples of expected size E(n) = 144; results are summarized by quintiles of increasing sample size.

Quintile	Average n	$\hat{Y}_{\bullet 3P7}$				Coscal (D^2H)			
		Bias%	SE%	Est. SE%	RMSE%	Bias%	SE%	Est. SE%	RMSE%
1	133.0	-0.88	2.78	3.00	2.92	-0.06	1.49	1.50	1.49
2	139.8	-0.34	2.71	2.91	2.73	-0.02	1.45	1.46	1.45
3	143.9	-0.04	2.67	2.87	2.67	-0.02	1.43	1.43	1.43
4	148.1	0.27	2.62	2.82	2.63	0.00	1.40	1.41	1.40
5	155.1	0.75	2.56	2.76	2.67	0.03	1.38	1.38	1.38

Table 6: Performance of $\hat{Y}_{\bullet 3P7}$ and Coscal (D^2H) estimators of total volume based on 100, 000 Poisson samples of expected size E(n) = 48; results are summarized by quintiles of increasing sample size.

Quintile	Average n	$\hat{Y}_{\bullet 3P7}$				Coscal (D^2H)			
		Bias%	SE%	Est. SE%	RMSE%	Bias%	SE%	Est. SE%	RMSE%
1	38.6	-0.03	5.55	5.60	5.55	-0.05	2.78	2.80	2.78
2	44.2	0.03	5.18	5.20	5.18	-0.03	2.61	2.61	2.61
3	47.8	-0.02	5.05	5.01	5.05	-0.03	2.53	2.51	2.53
4	51.6	-0.02	4.79	4.82	4.79	-0.03	2.41	2.41	2.41
5	57.8	-0.04	4.55	4.56	4.55	-0.02	2.29	2.28	2.29

Table 7: Performance of $\hat{Y}_{\bullet 3P7}$ and Coscal (D^2H) estimators of total volume based on 100, 000 collocated samples of expected size E(n) = 48; results are summarized by quintiles of increasing sample size.

Quintile	Average n	$\hat{Y}_{\bullet 3P7}$				Coscal (D^2H)			
		Bias%	SE%	Est. SE%	RMSE%	Bias%	SE%	Est. SE%	RMSE%
1	41.7	-1.50	5.05	5.41	5.27	0.22	2.66	2.70	2.66
2	45.5	-0.55	4.80	5.15	4.83	-0.01	2.56	2.57	2.56
3	47.9	-0.08	4.62	5.01	4.63	-0.02	2.48	2.51	2.48
4	50.4	0.49	4.49	4.87	4.52	0.02	2.43	2.44	2.43
5	54.5	1.28	4.33	4.68	4.51	0.05	2.34	2.34	2.67

Table 8: Performance of $\hat{Y}_{\bullet 3P7}$ and Coscal (D^2H) estimators of total volume based on 100, 000 Poisson samples of expected size E(n) = 24; results are summarized by quintiles of increasing sample size.									
Quintile	Average n	$\hat{Y}_{\bullet 3P7}$				Coscal (D^2H)			
		Bias%	SE%	Est. SE%	RMSE%	Bias%	SE%	Est. SE%	RMSE %
1	17.4	0.05	8.39	8.35	8.39	-0.04	4.24	4.19	4.24
2	21.3	0.02	7.46	7.51	7.46	-0.04	3.76	3.77	3.76
3	23.9	0.01	7.08	7.10	7.08	-0.02	3.53	3.55	3.53
4	26.5	-0.02	6.80	6.74	6.80	-0.02	3.39	3.37	3.39
5	31.0	-0.03	6.19	6.24	6.19	-0.04	3.11	3.12	3.11

Table 9: Performance of $\hat{Y}_{\bullet 3P7}$ and Coscal (D^2H) estimators of total volume based on 100, 000 collocated samples of expected size E(n) = 24; results are summarized by quintiles of increasing sample size.									
Quintile	Average n	$\hat{Y}_{\bullet 3P7}$				Coscal (D^2H)			
		Bias%	SE%	Est. SE%	RMSE%	Bias%	SE%	Est. SE%	RMSE %
1	19.6	-2.29	7.40	7.94	7.74	-0.19	3.92	3.96	3.93
2	22.2	-0.85	6.89	7.40	6.94	-0.03	3.65	3.69	3.65
3	23.9	-0.11	6.59	7.11	6.59	-0.02	3.54	3.56	3.54
4	25.7	0.57	6.33	6.84	6.36	0.01	3.40	3.42	3.40
5	28.6	1.76	5.95	6.47	6.20	0.04	3.22	3.25	3.22

6. SUMMARY

Estimation of the volume of timber available from a forest can be substantially improved by using regression estimators based on careful modelling of individual tree volume as a non-linear function of basal area and height.

Monte Carlo experiments have confirmed that collocated and even Poisson sampling, despite the random nature of their sample sizes, do not detract appreciably from the efficiency with which the volume of available timber in a forest can be measured—provided of course that the estimator used is one that adjusts appropriately for the achieved sample size.

Although some improvements appear possible, the conditional variances of the estimator 3P7 and of generalized regression estimators can already be estimated reasonably well, for both Poisson and collocated sampling, using the estimators of approximate conditional variance suggested in Brewer (2000). These estimators performed particularly well for 3P7 under Poisson sampling. When used with collocated sampling, they tended to overestimate.

7. REFERENCES

Brewer, K.R.W., (1963), "Ratio estimation and finite populations: some results deducible from the assumption of an underlying stochastic process," *Australian Journal of Statistics*, **5**, pp. 93-105.

Brewer, K.R.W., (1999), "Cosmetic Calibration for Unequal Probability Samples," *Survey Methodology,* **25**, pp. 205-212.

Brewer, K.R.W., (2000), "Deriving and Estimating an Approximate Variance for the Horvitz-Thompson Estimator Using only First Order Inclusion Probabilities," Contributed paper, Second International Conference on Establishment Surveys, Buffalo, NY, June 17-21.

Brewer, K.R.W., Early, L.J., and Hanif, M. (1984), "Poisson, Modified Poisson and Collocated Sampling," *Journal of Statistical Planning and Inference*, **10**, pp. 15-30.

Brewer, K.R.W., Early, L.J., and Joyce, S. F., (1972), "Selecting Several Samples from a Single Population," *Australian Journal of Statistics*, **14**, pp. 231-239.

Deville, J-C., (1999), "Variance Estimation for Complex Statistics and Estimators: Linearization and Residual Techniques," *Survey Methodology*, **25**, pp. 193-203.

Deville, J-C., and Särndal, C-E., (1992). "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, **87**, pp. 376-382.

Furnival, G.M., Gregoire, T.G. and Grosenbaugh, L.R., (1987), "Adjusted Inclusion Probabilities with 3P Sampling," *Forest Science*, 33, pp. 617-631.

Gregoire, T.G. and Valentine, H.T., (1999), "Composite and Calibration Estimation Following 3P Sampling," *Forest Science*, **45**, pp. 179-185.

Gregoire, T.G. and Williams, M., 1992, "Identifying and Evaluating the Components of Non-Measurement Error in the Application of Standard Volume Equations," *The Statistician*, **41**, pp. 509-518.

Grosenbaugh, L.R., (1964), "Some Suggestions for Better Sample-tree-measurement," *Proceedings of the Society of American Foresters*, Oct. 1963, Boston MA, Bethesda MD: Society of American Foresters.

Grosenbaugh, L.R., (1965), "Three-pee Sampling Theory and Program THRP for Computer Generation of Selection Criteria," *USDA Forest Service Research Paper PSW-21*. 53p.

Horvitz, D.G. and Thompson, D.J., (1952), "A Generalization of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association*, **47**, pp. 663-685.

Kott, P.S. and Bailey, J.T., (2000), "The Theory and Practice of Maximal Brewer Selection with Poisson PRN Sampling," Invited paper, Second International Conference on Establishment Surveys, Buffalo NY, June 17-21.

Särndal, C-E., (1996), "Efficient Estimators with Simple Variance in Unequal Probability Sampling," *Journal of the American Statistical Association*, **91**, pp. 1289-1300.

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

Sen, A.R., (1953), "On the Estimate of the Variance In Sampling with Varying Probabilities," *Journal of the Indian Society of Agricultural Statistics*, **5**, pp. 119-127.

Yates, F. and Grundy, P.M., (1953), "Selection Without Replacement From Within Strata With Probability Proportional To Size," *Journal of the Royal Statistical Society*, *Series B*, **15**, pp. 235-261.

A User's Guide to Pareto πps Sampling

Bengt Rosén

Statistics Sweden, Box 24 300, S - 104 51 Stockholm, Sweden. bengt.rosen@scb.se

ABSTRACT

A vehicle for utilization of auxiliary information is to employ a πps scheme, i.e. to sample with inclusion probabilities proportional to given size values. Pareto πps is a scheme for selection of a list sample with predetermined size. It has a number of attractive properties, in particular the following. As regards point estimate accuracy it is, in our best understanding, optimal among schemes which admit objective assessment of sampling errors. Simple procedures for variance estimation and non - response adjustment are available. The scheme admits efficient sample coordination by permanent random numbers. A sampling - estimation strategy with particularly good properties is obtained by combining Pareto πps and generalized regression estimation.

Key words **: Pareto πps, order πps, point and variance estimation, non - response adjustment.**

1 Introduction

The following situation will be considered. Information about a population characteristic is to be gained from observations on a probability sample from the population U = (1, 2 ,... , N). List sampling is used, from a frame that one - to - one corresponds with the units in U, and also contains unit - wise auxiliary information. Until Section 7 the auxiliary data are assumed to be size values $\mathbf{s} = (s_1, s_2, \dots, s_N)$, $s_k > 0$, which typically are positively correlated with the study variable $\mathbf{y} = (y_1, y_2, \dots, y_N)$. (Example : unit = enterprise, **y** = sales during Mach 2000, **s** = number of employees.) As is well known , estimation precision for a population total (or mean) often benefits from use of a ***πps scheme***, i.e. a scheme with sample inclusion probabilities $\pi_1, \pi_2, \dots, \pi_N$ such that ;

π_k is proportional to s_k , $k = 1, 2, ..., N.$ (1.1)

A well - known πps scheme is Poisson sampling , which has simple sample selection and estimation procedures. However, it and various other πps schemes have the drawback of random sample size. It is generally desirable that a sampling scheme has ***fixed sample size***, and we confine to schemes with that property. Then (1.1) leads to the following ***desired inclusion probabilities*** $\lambda_1, \lambda_2, \dots, \lambda_N$, where n is sample size and $\tau(\mathbf{s}) = s_1 + s_2 + ... + s_N$;

$$\lambda_k = n \cdot s_k / \sum_{j=1}^{N} s_j = n \cdot s_k / \tau(\mathbf{s}) , \quad k = 1, 2, \dots , N. \tag{1.2}$$

Formula (1.2) may yield λ : s exceeding 1, which is incompatible with being probabilities. If so, some adjustment has to be made, usually by introducing a "take all" stratum. In the sequel is presumed that $\lambda_k < 1$, $k = 1, 2, \dots, N$.

A scheme with inclusion probabilities according to (1.2) has the following Horvitz - Thompson estimator for the population total $\tau(\mathbf{y}) = y_1 + y_2 + ... + y_N$;

$$\hat{\tau}(\mathbf{y}) = \sum_{k \in \text{Sample}} y_k / \lambda_k = (\tau(\mathbf{s}) / n) \cdot \sum_{k \in \text{Sample}} y_k / s_k . \tag{1.3}$$

A "perfect" πps scheme shall satisfy $\pi_k = \lambda_k$, $k = 1, 2, \dots, N$. We will be a bit more generous, though, and accept a sampling scheme as a ***πps schemes*** if (1.4) below is met ;

$\pi_k \approx \lambda_k$ holds with good approximation for $k = 1, 2, ..., N.$ (1.4)

The literature offers several πps schemes , and the statistician must choose. The main ***desired properties of a πps scheme***, besides having fixed sample size, are listed below.

- The scheme has ***simple sample selection***. (1.5)
- The scheme yields ***good estimation accuracy***. (1.6)
- The scheme ***admits objective assessment of sampling errors*** (consistent variance estimation). (1.7)
  - Variance estimators are available, the simpler the better.
  - Variance estimates never become negative.
- The scheme ***admits sample coordination*** over time and between surveys. (1.8)

The πps scheme which is most frequently used in practice is ***systematic πps***. This is in fact is a whole family of schemes generated by different frame ordering rules , whereby **r*andom frame order*** (rfo) and ***frame ordered by size*** (sfo) are chief possibilities. ***Sunter πps***, Sunter (1977) , is highlighted in Särndahl et al. (1992).

This paper focuses on ***Pareto πps***, a member in the family of ***order πps schemes*** specified in Section 2.2. The aim is to (hopefully) demonstrate that Pareto πps meets the desires formulated above particularly well and, accordingly, to recommend it for practical use. Rigorous justifications of subsequent claims require quite sophisticated theory, though, which is not presented in this paper. We confine to just giving references.

2. Order πps schemes, notably Pareto πps

2.1. Pareto πps

DEFINITION 2.1 : ***Pareto πps*** with ***size values*** $\mathbf{s} = (s_1, s_2, \ldots, s_N)$ and ***sample size*** n generates a sample as follows.

1. ***Desired inclusion probabilities*** $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ are computed by (1.2).
2. Independent random variables $R_1, R_2, \ldots, R_N$ with uniform distributions on the interval $[0, 1]$ are realized, and ***ranking variables*** Q are computed as follows;

$$Q_k = \frac{R_k \cdot (1-\lambda_k)}{\lambda_k \cdot (1-R_k)}, \quad k = 1, 2, \ldots, N. \tag{2.1}$$

3. ***The sample*** consists of the units with the ***n smallest Q - values***.

The somewhat fancy name Pareto is explained in next section. When all s_k (or equivalently all λ_k) agree, Pareto πps is nothing but simple random sampling (SRS).

All order πps schemes and, hence, also Pareto πps are based on asymptotic considerations. They are approximate πps schemes in the (1.4) sense, having perfect πps properties only for "infinite" samples. In particular, desired and factual inclusion probabilities do not agree exactly for finite samples. However, as discussed in Section 5, these imperfections have negligible effects in most practical survey situations, even for quite small sample sizes.

Pareto πps certainly meets desire (1.5), as is illustrated by the SAS - program below. It selects a Pareto πps sample with sample size SAMPZ from the records in the SAS data set FRAME, equipped with desired inclusion probabilities in the variable LAMB.

```
Data RANKING; Set FRAME; R = ranuni(SEED);
Q = R*(1 - LAMB) / (LAMB*(1 - R)); Proc SORT; By Q; Run;
Data SAMPLE; Set RANKING; If _N_ <= SAMPZ then output; Run;
```

2.2 Order πps

As stated, Pareto πps is a particular member in the family of order πps schemes introduced in Rosén (1997b), which is defined below. $H(\cdot)$ denotes the distribution function of a probability distribution with density.

DEFINITION 2.2 : ***Order πps*** with ***size values*** $\mathbf{s} = (s_1, s_2, \ldots, s_N)$, ***sample size*** n and ***shape distribution*** $H(\cdot)$ generates a sample by the same type of steps as in Definition 2.1, with the following modification of step 2. The r***anking variables*** Q are computed by (2.2) below, where H^{-1} denotes inverse function;

$$Q_k = H^{-1}(R_k) / H^{-1}(\lambda_k), \quad k = 1, 2, \ldots, N. \tag{2.2}$$

It is by no means obvious that Definition 2.2 leads to πps schemes even in the (1.4) sense. However, Rosén (2000) proves that this is the case under general conditions on the shape distribution H.

Pareto πps is the particular order πps scheme given by the shape distribution;

$$H(t) = t/(1+t), \quad \text{which has density } h(t) = 1/(1+t)^2, \ 0 \le t < \infty. \tag{2.3}$$

Definition 2.2 introduces a whole family of sampling schemes, different H yield different schemes (although not in an entirely one - to - one fashion). To distinguish order πps schemes they are named by their shape distribution. The distribution in (2.3) seems to have no well - established name, though. After literature consultation we follow Feller (1966) and call it a Pareto distribution. Hence, the name Pareto πps.

Hitherto studies of order πps schemes have paid special attention to, besides Pareto πps, ***uniform order πps*** given by the uniform shape distribution $H(t) = \min(t, 1)$, $0 \le t < \infty$, and ***exponential order πps*** given by the exponential shape distribution $H(t) = 1 - e^{-t}$, $0 \le t < \infty$. The former scheme was first studied by Ohlsson (1990, 1998).

He calls it ***sequential Poisson sampling*** . Ohlsson's work provided background for the generalized notion "order πps". As regards Pareto πps , the author and Saavedra (1995) independently came across that scheme. Saavedra calls it ***odds ratio sequential Poisson sampling***. Exponential order sampling is considered in the literature under the name ***successive sampling***, see e.g. Rosén (1997a).

3 Estimation from observations on Pareto πps samples

Procedures for point and variance estimation are, of course, crucial in practical application of a sampling scheme. The following discussion is confined to the key estimation problem, estimation of the characteristic "population total", denoted $\tau(\mathbf{y}) = y_1 + y_2 + ... + y_N$. As is well known , if this problem can be handled the estimation problem is solved for the vast majority of practically interesting characteristics , including ratios and domain characteristics . We first consider the ideal situation when all sampled units respond.

3.1 Estimation under full response

3.1.1 Point estimation

Since a πps scheme is presumed to satisfy (1.4) it is natural to use the estimator (1.3) , which is re - stated in (3.1) below. However, under (1.4) it is not a "perfect", unbiased HT - estimator, rather a "quasi" HT - estimator, afflicted with some bias. The bias issue for Pareto πps is discussed in Section 5, with the conclusion that the bias is negligible almost always in practice.

$$\hat{\tau}(\mathbf{y}) = \sum_{k \in \text{Sample}} y_k / \lambda_k \tag{3.1}$$

3.1.2 Estimator variance

At least in survey planning it is of interest to have an expression for the theoretical estimator variance. The following approximate variance formula, derived in Rosén (1997a , b), is asymptotically correct ;

$$V[\hat{\tau}(\mathbf{y})] \approx \frac{N}{N-1} \cdot \sum_{k=1}^{N} \left(\frac{y_k}{\lambda_k} - \sum_{j=1}^{N} y_j \cdot (1-\lambda_j) \Big/ \sum_{j=1}^{N} \lambda_j \cdot (1-\lambda_j) \right)^2 \cdot \lambda_k \cdot (1 - \lambda_k) . \tag{3.2}$$

3.1.3 Assessment of sampling error

Formula (3.2) is theoretical and involves $\mathbf{y}$ - values for all population units. In practice an estimator $\hat{V}[\hat{\tau}(\mathbf{y})]$ of $V[\hat{\tau}(\mathbf{y})]$ must be exhibited , which together with approximate normal distribution for the estimator justifies the following type of (approximate) ***level 100 · (1 - α) % confidence interval*** for $\tau(\mathbf{y})$, $\delta_{\alpha/2}$ denoting the standard normal 1 - $\alpha/2$ fractile ;

$$\hat{\tau}(\mathbf{y}) \pm \delta_{\alpha/2} \cdot \sqrt{\hat{V}[\hat{\tau}(\mathbf{y})]} . \tag{3.3}$$

Consistent estimation of $V[\hat{\tau}(\mathbf{y})]$ is provided by, see Rosén (1997b) ;

$$\hat{V}[\hat{\tau}(\mathbf{y})] = \frac{n}{n-1} \cdot \sum_{k \in \text{Sample}} \left(\frac{y_k}{\lambda_k} - \sum_{j \in \text{Sample}} \frac{y_j \cdot (1-\lambda_j)}{\lambda_j} \Big/ \sum_{j \in \text{Sample}} (1-\lambda_j) \right)^2 \cdot (1 - \lambda_k) . \tag{3.4}$$

Formula (3.4) may look cumbersome at first glance. However, it is quite innocent from a computation point of view, as is seen from the formulas below. Note that only "single - summations" are involved.

$$\hat{V}[\hat{\tau}(\mathbf{y})] = \frac{n}{n-1} \cdot (A - B^2/C), \quad \text{where} \tag{3.5}$$

$$A = \sum_{k \in \text{Sample}} (y_k/\lambda_k)^2 \cdot (1 - \lambda_k) , \quad B = \sum_{k \in \text{Sample}} (y_k/\lambda_k) \cdot (1 - \lambda_k) , \quad C = \sum_{k \in \text{Sample}} (1 - \lambda_k) . \tag{3.6}$$

Formulas (3.4) - (3.6) show that Pareto πps meets both desires in (1.7), simple and non - negative variance estimation. Moreover, in Rosén (1997b) is proved that $\hat{\tau}(\mathbf{y})$ is asymptotically normally distributed under Pareto πps.

3.2 Estimation when non - response occurs

Practical surveys are seldom ideal , in particular non - response almost always occurs. Then formulas (3.1) and (3.4) cannot be used straight away, some adjustment for non - response has to be made. The following results are

taken from Rosén & Lundqvist (1998), where adjustment procedures for uniform, exponential and Pareto πps are presented, with theoretical as well as numerical justifications.

Non-response adjustment must be based on some model (= assumption) about response behavior. The simplest, and most commonly used, is the simple MAR (Missing At Random) model stated below.

***Simple MAR model*:** Sampled units respond independently, all with the same response propensity. (3.7)

Under (3.7) adjustment is achieved by a recipe which somewhat sweepingly can be formulated as follows.

Use (3.1) and (3.4) with "number of sampled units" exchanged for "number of responding units". (3.8)

More precisely, under (3.7) point estimation is carried out as follows. With

n' = number of responding units, (3.9)

modified inclusion probabilities: $\lambda'_k = (n'/n)\cdot\lambda_k = n'\cdot s_k / \sum_{j=1}^{N} s_j$, $k = 1, 2, \ldots, N$, (3.10)

$\mathcal{R}$ sample = the collection of responding units, (3.11)

the following counterpart to the estimator in (3.1) works with negligible bias;

$$\hat{\tau}(\mathbf{y}) = \sum_{k \in \mathcal{R}\text{sample}} y_k / \lambda'_k \,. \tag{3.12}$$

Moreover, the confidence interval (3.3) works with the variance estimator;

$$\hat{V}[\hat{\tau}(\mathbf{y})] = \frac{n'}{n'-1} \cdot \sum_{k \in \mathcal{R}\text{sample}} \left(\frac{y_k}{\lambda'_k} - \sum_{j \in \mathcal{R}\text{sample}} \frac{y_j \cdot (1-\lambda'_j)}{\lambda'_j} \Big/ \sum_{j \in \mathcal{R}\text{sample}} (1-\lambda'_j) \right)^2 \cdot (1-\lambda'_k) \,. \tag{3.13}$$

Formulas (3.5) and (3.6) have the following counterparts;

$$\hat{V}[\hat{\tau}(\mathbf{y})] = \frac{n'}{n'-1} \cdot (A' - B'^2/C') \,, \quad \text{with } A', B' \text{ and } C' \text{ as in (3.5)-(3.6) with } \lambda \text{ changed to } \lambda'. \tag{3.14}$$

A more elaborate response model runs as follows.

***MAR model with several response homogeneity groups*:** With the population partitioned into (known) disjoint groups $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_G$, the following holds. Sampled units respond independently, and units in the same group have the same response propensity, which may vary between groups. (3.15)

Under response model (3.15), adjustment for non-response is achieved by the following post-stratification procedure. For $g = 1, 2, \ldots, G$, n_g denotes the number of sampled units from $\mathcal{G}_g$, n'_g the number of responding units and $\mathcal{R}_g$sample the set of responding units. Set $\mathcal{R}$sample = $\mathcal{R}_1$sample $\cup$ $\mathcal{R}_2$sample $\cup$ ... $\cup$ $\mathcal{R}_G$sample. Modify the λ:s according to (3.16) below, where $g_{(k)}$ is the index for the group to which unit k belongs. The somewhat intriguing operation min($\cdot$, 1) is introduced for the following reason. Without it, it can happen, although only in exceptional cases, that one or more λ'_k becomes greater than 1.

$$\lambda'_k = \min\left(n'_{g(k)} \cdot \lambda_k \Big/ \sum_{j \in G_{g(k)}} \lambda_j \,, 1\right) = \min\left(n'_{g(k)} \cdot s_k \Big/ \sum_{j \in G_{g(k)}} s_j \,, 1\right), \quad k = 1, 2, \ldots, N. \tag{3.16}$$

After these λ-modifications, (3.12) works and also the confidence interval (3.3) with the variance estimator;

$$\hat{V}[\hat{\tau}(\mathbf{y})] = \sum_{g=1}^{G} \frac{n'_g}{n'_g - 1} \cdot \sum_{k \in \mathcal{R}_g\text{sample}} \left(\frac{y_k}{\lambda'_k} - \sum_{j \in \mathcal{R}_g\text{sample}} \frac{y_j \cdot (1-\lambda'_j)}{\lambda'_j} \Big/ \sum_{j \in \mathcal{R}_g\text{sample}} (1-\lambda'_j) \right)^2 \cdot (1-\lambda'_i) \,. \tag{3.17}$$

Numerical computation of variance estimates by (3.17) is simplified by expansion of the squares, which leads to formulas analogous to (3.5) and (3.6). Moreover, formulas (3.13)-(3.17) show that both desires in (1.7) are met also under non-response adjustment.

4 Estimation accuracy

The accuracy of an estimator depends on its variance and possible bias. As already mentioned, $\hat{\tau}(\mathbf{y})$ in (3.1) is afflicted with some bias which, however, is negligible in almost all practical contexts. Therefore, we confine the discussion of estimation accuracy to estimator variance. In the sequel notions as "optimal", "better", etc. relate to $V[\hat{\tau}(\mathbf{y})]$. The following result from Rosén (1997b) explains why Pareto πps is of special interest.

Pareto πps is asymptotically (as $n \to \infty$) optimal among order πps schemes with the same size values and sample size, uniformly over study variables **y**. Moreover, optimality holds not only for estimation of totals, it holds for all "usual" types of characteristics, including ratios and domain characteristics. (4.1)

The result (4.1) tells that Pareto πps at least asymptotically performs better than other order πps schemes. However, there are other "competing" schemes, notably those at the end of Section 1, as well as small sample situations. To compare schemes we use the measure ***relative*** (to Pareto πps) ***variance increase*** (RVI);

$$\textbf{\textit{RVI}}(\text{for scheme } S) = \frac{V[\hat{\tau}(\mathbf{y})] \text{ under } \pi\text{ps scheme } S}{V[\hat{\tau}(\mathbf{y})] \text{ under Pareto } \pi\text{ps}} - 1 . \qquad (4.2)$$

Factors that affect RVI significantly are: (i) The sampling fraction n/N. (ii) The relation between the size and study variables **s** and **y**. As regards this relation, ideal for πps is when **s** and **y** are exactly proportional, then $\tau(\mathbf{y})$ is in fact estimated without error. In that case the $(\mathbf{s}, \mathbf{y})$-values lie along a straight line through the origin. In practice this never occurs, though, the $(\mathbf{s}, \mathbf{y})$-values ***scatter*** more or less around a ***trend***, which is assumed to be increasing. Chief possibilities for trend type are ***proportional*** trend (= linear through the origin), ***convex*** and ***concave*** trend (when **y** grows relatively faster respectively slower than **s**). If the trend is flat or decreasing, πps sampling is non-favorable compared with SRS.

Table 4.1 presents examples of RVI-values for different schemes in certain sampling situations (N, **s** and **y**). As seen, the compared schemes are uniform and exponential order πps, Sunter πps, systematic πps(rfo) and πps(sfo) and SRS. Pareto πps is covered implicitly with all RVI = 0. SRS is included as a benchmark. The figures in the table, which come from Rosén (1997b), are based on Monte Carlo simulations. The considered sampling situations were generated with $(\mathbf{s}, \mathbf{y})$-relation according to model (4.3) below, where α determines the trend shape (proportional for $\alpha = 1$, convex for $\alpha > 1$, concave for $\alpha < 1$) and σ the magnitude of scatter.

$$s_k = k, \quad y_k = c \cdot (s_k^{\alpha} + \sigma \cdot Z_k \cdot \sqrt{s_k^{\alpha}}), \quad c > 0, \quad \sigma \geq 0, \quad \text{the } Z_k \text{ being iid } N(0,1), \quad k = 1, 2, ..., N. \qquad (4.3)$$

Table 4.1 **RVI - values (in %) for situations (4.3) with N = 100 and α, σ as stated below**

Sampling fraction	Uniform πps			Exponential πps			Sunter πps			Systematic πps(rfo)			Systematic πps(sfo)			SRS		
	α = 1.5 σ = 2	α = 1 σ = 2	α = 0.7 σ = 0.5	α = 1.5 σ = 2	α = 1 σ = 2	α = 0.7 σ = 0.5	α = 1.5 σ = 2	α = 1 σ = 2	α = 0.7 σ = 0.5	α = 1.5 σ = 2	α = 1 σ = 2	α = 0.7 σ = 0.5	α = 1.5 σ = 2	α = 1 σ = 2	α = 0.7 σ = 0.5	α = 1.5 σ = 2	α = 1 σ = 2	α = 0.7 σ = 0.5
0.1	0.2	$4 \cdot 10^{-4}$	0.1	0.1	$1 \cdot 10^{-4}$	0.03	5.8	2.6	25	2.2	2.4	3.4	-74	+3.6	-34	1083	450	246
0.2	1.3	$2 \cdot 10^{-3}$	0.7	0.3	$6 \cdot 10^{-4}$	0.2	62	58	43	1.0	-3.5	-0.4	-75	-16	-52	1095	446	229
0.3	4.2	$6 \cdot 10^{-3}$	1.9	0.9	$2 \cdot 10^{-3}$	0.5	262	181	103	0.7	0.6	-0.6	-73	+88	-9.2	1130	440	213
0.4	12	0.02	4.7	2.3	$7 \cdot 10^{-3}$	1.7	653	280	141	21	-2.2	7.3	-86	-21	-52	1222	432	197
0.5	39	0.04	12	4.7	0.03	2.0	1457	401	180	29	-3.3	13	-73	+12	-47	1538	421	186

The main conclusions from the full collection of numerical findings in Rosén (1997b) are stated in (4.4) - (4.6).

Under proportional $(\mathbf{s}, \mathbf{y})$-trend the order πps schemes perform very similarly, with a slight edge for Pareto πps. Under non-proportional trend Pareto πps performs better than the other, the edge is small, though, for small sampling fractions but may be considerable for high ones. (4.4)

This result tells : (i) The optimality result (4.1) holds not only asymptotically , in essence it holds for quite small samples. (ii) Among order πps schemes, Pareto πps provides an "insurance without premium", it never performs worse than other order πps schemes and in some situations considerably better.

Next comparison is made with schemes outside order πps, which admit objective assessment of sampling error (as all order πps schemes do), with Sunter πps and systematic πps(rfo). The chief finding is as follows.

(4.4) holds with "order πps" exchanged for "πps scheme which admits consistent variance estimation". (4.5)

The findings from comparison with systematic πps(sfo) (i.e. systematic πps with frame ordered by the **s** - values) are more complex. The strong side of this scheme is that it sets two variance reducing forces in action , πps and "implicit stratification". Its well - known weakness is that it does not admit assessment of sampling errors. The following rather confusing comparison picture arose.

Systematic πps(sfo) often yields dramatically better estimation accuracy than Pareto πps , notably in situations with non - proportional $(\mathbf{s}, \mathbf{y})$ - trend,. In situations with fairly proportional $(\mathbf{s}, \mathbf{y})$ - trend the ranking between systematic πps(sfo) and Pareto πps seems to be erratic, they take turn to be best. (4.6)

To the best of our understanding it is hard to tell in advance which of Pareto πps and systematic πps(sfo) that is advantageous in a particular sampling situation .

5 Approximation accuracy, notably estimator bias

5.1 Some basic notions

As mentioned several times , since Pareto πps is based on asymptotic considerations desired and factual inclusion probabilities do not agree exactly, which in turn afflicts $\hat{\tau}(\mathbf{y})$ in (3.1) with some bias. When discussing these issues we use the performance measures in (5.1) and (5.2) below. First a comment on notation. Inclusion probabilities depend on population size N, size values $\mathbf{s} = (s_1, s_2, \dots, s_N)$ and sample size n. In the sequel these parameters often are exhibited in notations like $\pi_k(n; N; \mathbf{s})$ and $\lambda_k(n; N; \mathbf{s})$.

Maximal absolute relative error for inclusion probabilities :

$$\Psi(n; N; \mathbf{s}) = \max_k \left| \pi_k(n; N; \mathbf{s}) / \lambda_k(n; N; \mathbf{s}) - 1 \right| . \tag{5.1}$$

Absolute relative estimator bias : $\mathrm{AREB}[\hat{\tau}(\mathbf{y})] = \left| \mathrm{E}[\hat{\tau}(\mathbf{y})] / \tau(\mathbf{y}) - 1 \right| .$ (5.2)

The above concepts are related as follows , which is demonstrated e. g. in Rosén (2000 a) ;

$$\mathrm{AREB}[\hat{\tau}(\mathbf{y})] \le \Psi(n; N; \mathbf{s}) \cdot \left(\sum_{k=1}^{N} | y_k | \Big/ \tau(\mathbf{y}) \right) . \tag{5.3}$$

If the ***study variable is non - negative***, i.e. if $y_k \ge 0$, $k = 1, 2, \dots, N$, which is the case in most practical surveys , the last factor in (5.3) equals 1, and (5.3) takes the simple form ;

$$\mathrm{AREB}[\hat{\tau}(\mathbf{y})] \le \Psi(n; N; \mathbf{s}) . \tag{5.4}$$

The AREB bounds in (5.2) and (5.3) are often fairly conservative , as discussed in Rosén (2000 a).

$\Psi(n ; N ; \mathbf{s})$ is defined in terms of approximation accuracy for inclusion probabilities, a rather theoretical topic. However, by (5.3) and (5.4) Ψ also provides information about estimator bias, which makes it interesting also from survey practical point of view. Before entering bias questions we present some results about the asymptotic behavior of inclusion probabilities.

5.2 Asymptotic behavior of inclusion probabilities

Rosén (2000 a) proves that (5.5) below holds under general conditions for Pareto, uniform and exponential πps ;

$$\max_k \left| \pi_k(n; N; \mathbf{s}) / \lambda_k(n; N; \mathbf{s}) - 1 \right| \quad \text{is (at most) of order} \quad O(\log n / \sqrt{n}) . \tag{5.5}$$

Results of type (5.5) are used to study the asymptotics of inclusion probabilities, in the usual frame - work for finite population asymptotics : A sequence of populations with sizes tending to infinity. In particular, the result (5.6) below is proved for the schemes mentioned above. It tells that desired and factual inclusion probabilities agree asymptotically. It is conjectured that (5.6) holds generally for any order πps scheme.

$$\pi_k(n;N;\mathbf{s}) / \lambda_k(n;N;\mathbf{s}) \to 1 \quad \text{as } n \to \infty \quad \text{(and hence also } N \to \infty\text{)}. \tag{5.6}$$

5.3 On estimator bias

5.3.1 Factors that affect the bias magnitude

When considering use of Pareto πps in practice , a crucial question for the statistician is ;

Will the Pareto πps point estimator bias be negligible in my particular survey situation ? (5.7)

In search for answers to (5.7) , available theoretical bounds of type (5.5) , regrettably only add to the general ex- perience that theoretical error bounds seldom are sharp enough to yield practically valuable information on the small sample performance of a large sample procedure. The bound O(logn/√n) in (5.5) is too crude for that purpose. In our understanding , practically useful information can only be gained by numerical investigations, exact computations or/and Monte Carlo simulations. The results presented in the sequel come from Aires & Rosén (2000), which reports on an extensive numerical study of exactly computed Pareto πps inclusion probabilities.

Answers to (5.7) are with necessity a bit involved , since the bias depends on several factors. The study variable is of course one of them. On this point we confine to the case with non - negative study variables , which is the most common in practice. By (5.4) , Ψ can then be interpreted as an AREB bound. Other factors that affect whether the bias is negligible or not are : (i) The tolerance limit for "negligible". (ii) The variation of the size values. (iii) The population size. (iv) The sample size. These factors are discussed below.

Tolerance limit for negligibility

There is of course no unanimous answer to how large a "negligible" bias may be. This depends on the intended use of the statistic and on the magnitude of sampling errors and other survey errors. We believe that most survey statisticians regard 1% , and even 2% , as a negligible relative bias.

Dependence on size values

When all size values are equal , Pareto πps is SRS with $\pi_k = \lambda_k = n/N$. Hence, for bias to be at hand the size values must show variation. In the following , the size values **s** are presumed to be ***normed*** so that average size is 1, i.e. so that (5.8) below is met. A normed **s** is referred to as a ***size pattern***.

$$(1/N) \cdot \sum_{k=1}^{N} s_k = 1 . \tag{5.8}$$

The maximal and minimal normed s - values are denoted s_{max} and s_{min}. The ***size pattern range*** is specified by the interval $[s_{min} , s_{max}]$. Another aspect on s - value variation is the ***size pattern shape***, which concerns how size values spread over $[s_{min} , s_{max}]$. Following Aires & Rosén (2000), where precise definitions are given , three "extremal" shape types are considered. (i) The s - values are fairly ***evenly spread*** over $[s_{min} , s_{max}]$. (ii) The majority of s - values lie in the ***middle*** of $[s_{min} , s_{max}]$. (ii) The majority of s - values lie at the ***boundaries*** of $[s_{min} , s_{max}]$.

Figures 5.1 illustrates how $\Psi(\cdot ; N ; \mathbf{s})$ - sequences may differ for different pattern shapes with the same N, s_{min} and s_{max} . In particular it illustrates the following general circumstance. When the sample size is not "very small", the boundary shape is most adverse to good agreement between desired and factual inclusion probabilities.

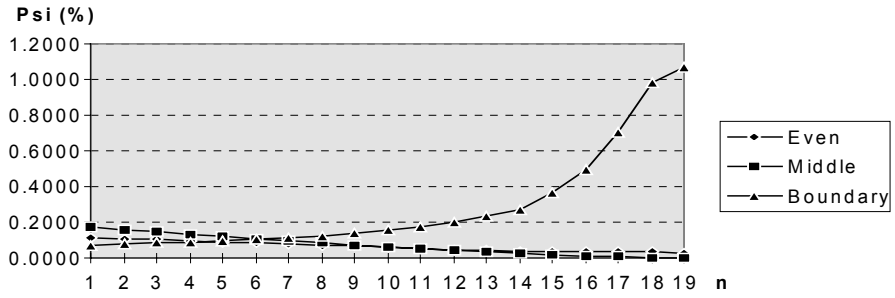

Figure 5.1 Psi-sequences for N=100, s_{min}=0.1, s_{max}=5

The statements in (5.9) and (5.10) below are based on our experience of size value patterns met in practice.

We believe that s_{max} seldom is larger than 5 and that s_{min} seldom is smaller than 0.1. (5.9)

The boundary shape is very unusual in practice. Most practical size pattern shapes resemble the even spread shape, in the sequel referred to as "lying ***in the vicinity of even spread***". (5.10)

Some background for (5.9) and (5.10) is as follows. (i) The surveyor disposes of the size values, preliminary values may be modified. If the frame comprises units with very small preliminary size values, such units are often either definition-wise excluded from the survey population or given larger **s**-values in the sample selection.

(ii) If size values vary very much over the entire population, there are often subject matter grounds for stratification by size before sampling, followed by drawing independent samples from the strata. (An example is an enterprise survey with "number of employees" as size. It is usually natural to divide into strata of type "very big", "big" and "small" enterprises. Mostly the "very big" stratum is totally inspected.) The strata then take population roles, and s_{max} and s_{min} in the strata are usually considerably smaller / larger than in the entire population.

Table 5.1 below introduces, for later use, a broad categorization of size value patterns.

Table 5.1 Some categories of size value patterns

	Category A	Category B	Category C	Category D
Size pattern shape	In the vicinity of even spread.	In the vicinity of even spread.	No restriction.	No restriction.
s_{max}	≤ 5	≤ 10	≤ 5	≤ 10
s_{min}	≥ 0.1	≥ 0.05	≥ 0.1	≥ 0.05
Comments on occurrence in practice	Most practical situations are believed to fall in Category A.	"Normal" pattern shape, while s_{max} and/or s_{min} may be extreme	"Normal" s_{max} and s_{min}, while shape may be extreme (e.g. of boundary type).	Pattern shape as well as s_{max} and/or s_{min} may be extreme.

Dependence on population size

The numerical findings show that for given size value pattern **s** and sample size n, $\Psi(n; N; \mathbf{s})$ decreases as the population size N increases, i.e. desired and factual inclusion probabilities come closer to each other.

Dependence on sample size

The assumption $\lambda_k < 1$ constrains sample sizes as stated in (5.11), where $[\cdot]$ denotes integral part, and - "less than". The quantity n_m is called the ***maximal sample size***, and an n which satisfies (5.11) is said to be ***admissible***.

$$n \leq n_m = n_m(N; \mathbf{s}) := [N/s_{max} -]. \quad (5.11)$$

Since Pareto πps is based on asymptotic considerations, one expects in the first round that conditions for small Ψ (hence, for small bias) would be of the type "provided that n is *at least* ...". However, as illustrated in Figure 5.1, conditions for small Ψ, which encompass all types of size pattern shapes, inclusive the unpleasant boundary shape type, rather are of the form "provided that n is *at most* ..." (For pattern shapes in the vicinity of even

spread, though , Ψ typically decreases as n increases .) This aspect is handled technically as follows. An α is specified, $0 < \alpha < 1$, and is used to determine an ***α-maximal sample size*** $n_{m,\alpha}$ as follows;

$$n_{m,\alpha}(N;\mathbf{s}) := [\alpha \cdot N / s_{max}] . \qquad (5.12)$$

Conditions to the effect "sample size is *at most* ..." are formulated by stating that n must not exceed $n_{m,\alpha}$.

5.3.2 Conditions which imply negligible estimator bias

The numerical findings in Aires & Rosén (2000) are condensed in Table 5.2 below. More detailed information is provided in the full report. Population sizes smaller than 25 were not considered in that study.

Table 5.2. Sample sizes that imply negligible bias

N = population size, α states that sample size must not exceed $n_{m,\alpha}$ in (5.12). n_0 specifies a sufficiently large sample size, under the α-restriction, for negligible bias with specified tolerance. Study variables are presumed to be non-negative.

Size Pattern category See Table 5.1	Tolerance limit for negligibility								
	2%			1%			0.5%		
	N	α	n_0	N	α	n_0	N	α	n_0
A	≥ 25	1	1	≥ 40	1	1	≥ 80	1	1
				[25 , 40)	1	3	[50 , 80)	1	3
							[40 , 50)	1	4
B	≥ 80	1	1	≥ 80	1	1	≥ 125	1	1
	[25 , 80)	1	2	[40 , 80)	1	3	[100 , 125)	1	3
							[80 , 100)	1	4
C	≥ 100	1	1	≥ 125	1	1	≥ 175	1	1
	≥ 80	0.9	1	≥ 100	0.9	1	≥ 150	0.9	1
	≥ 40	0.8	1	≥ 80	0.8	1	≥ 100	0.8	1
	≥ 25	0.5	1	≥ 40	0.5	1	≥ 80	0.5	1
D	≥ 150	0.9	1	≥ 175	0.8	1	≥ 125	0.5	1
	≥ 125	0.8	1	≥ 80	0.5	1	[100 , 125)	0.5	3
	≥ 80	0.5	1						
	[50 , 80)	0.5	2						

Earlier remarks imply that the sufficient sample sizes n_0 in Table 5.2 in most practical situations are conservative and , hence, "overly safe". In particular, one should not conclude that the bias necessarily is larger than "guaranteed" for sample sizes that are smaller then stated n_0 - values. However, even with the above conservative bounds the conclusion is that the bias in almost all practical situations is negligible for all admissible sample sizes.

6 On sample coordination and adjustment for overcoverage

Ohlsson (1990, 1995, 1998) emphasizes that uniform order πps has the attractive properties which are discussed below. These properties are shared by all order πps schemes, hence also by Pareto πps.

For an order πps scheme positive ***coordination of samples*** (to achieve great sample overlap) drawn at different occasions in time from the "same" (but updated) frame is obtained by associating ***permanent random numbers*** to the frame units, to be used at successive draw occasions , i.e. by letting the R_k in Definition 2.2 be permanent. Similar technique can be used for positive or negative coordination of simultaneously drawn samples to different surveys. Negative coordination is achieved for example if R_k in one sample selection is exchanged for $1 - R_k$ in another selection.

When the frame contains ***overcoverage*** (out - of - scope units) , a sample of predetermined size from the (unobservable) list of in - scope units can be selected as follows . Order the frame units by the Q : s in Definition 2.2 , and start observing them in that order. Exclude successively encountered out - of - scope units until a sample of in- scope units of prescribed size is obtained. This procedure yields an order πps sample from the in - scope units.

One looses full control of the inclusion probabilities, though, since the size sum over the in-scope list is unknown. However, if the task is to estimate a ratio $\tau(\mathbf{y})/\tau(\mathbf{x})$, as is the case in e.g. price index surveys, this does not matter since the unknown size sums in the estimates of nominator and denominator cancel. In the general case the unknown size sum is readily estimated.

7 Pareto πps as component in optimal sampling - estimation strategies

So far available auxiliary information has consisted of size values $\mathbf{s} = (s_1, s_2, \dots, s_N)$. Here we turn to a more elaborate situation, with auxiliary information in conjunction with a superpopulation model. The general version of the simple model below provides background for generalized regression estimation, as described in Chapter 6 in Särndal et al. (1992). The study variable $\mathbf{y}$ relates to (here one-dimensional) auxiliary data $x_1, x_2, \dots, x_N$ according to the following superpopulation model;

$$y_k = \beta \cdot x_k + \varepsilon_k, \quad k = 1, 2, \dots, N, \tag{7.1}$$

where $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ satisfy the following conditions, with $\mathcal{E}$, $\mathcal{V}$ and $\mathcal{C}$ for superpopulation expectation, variance and covariance: $\mathcal{E}[\varepsilon_k] = 0$, $\mathcal{V}[\varepsilon_k] = \sigma_k^2$ and $\mathcal{C}[\varepsilon_k, \varepsilon_l] = 0$, $k \neq l$, $k, l = 1, 2, \dots, N$. The parameters $\sigma_1, \sigma_2, \dots, \sigma_N$ are part of the auxiliary information, and are regarded as known modulo a proportionality factor.

Some notation. Subscript HT signifies Horvitz-Thompson estimators. Algebraic operations on variables shall be interpreted as component-wise. For $\mathbf{y} = (y_1, y_2, \dots, y_N)$ and $\mathbf{x} = (x_1, x_2, \dots, x_N)$, $\mathbf{y} \cdot \mathbf{x} = (y_1 \cdot x_1, y_2 \cdot x_2, \dots, y_N \cdot x_N)$, $\mathbf{y}/\mathbf{x} = (y_1/x_1, y_2/x_2, \dots, y_N/x_N)$, $\mathbf{x}^2 = (x_1^2, x_2^2, \dots, x_N^2)$.

Next we state a result due to Cassel et al (1976). An admissible (sampling-estimation) ***strategy*** is a pair $[P, \hat{\tau}(\mathbf{y})]$ of a sample design P and a linear, design unbiased estimator $\hat{\tau}(\mathbf{y})$. They showed that ***optimal strategies***, relative to minimization of $\mathcal{E}(V[\hat{\tau}(\mathbf{y})])$, are characterized by the following properties;

P is a πps scheme with size values proportional to $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)$. (7.2)

The estimator $\hat{\tau}(\mathbf{y})$ is of the form $\hat{\tau}(\mathbf{y})_{HT} + \beta \cdot [\tau(\mathbf{x}) - \hat{\tau}(\mathbf{x})_{HT}]$. (7.3)

Since the Cassel et al. paper much effort has been devoted to the estimator part of the optimal strategy, leading to the generalized regression estimator (GREG);

$$\hat{\tau}(\mathbf{y})_{GREG} = \hat{\tau}(\mathbf{y})_{HT} + \hat{B} \cdot [\tau(\mathbf{x}) - \hat{\tau}(\mathbf{x})_{HT}] \quad \text{where} \quad \hat{B} = \hat{\tau}(\mathbf{y} \cdot \mathbf{x}/\boldsymbol{\sigma}^2)_{HT} / \hat{\tau}(\mathbf{x}^2/\boldsymbol{\sigma}^2)_{HT}. \tag{7.4}$$

However, only little attention has been paid to the design part, the πps scheme. A possible reason may be shortage of πps schemes with good properties. Since Pareto πps provides a nice πps scheme, at least in the author's opinion, it is of interest to revisit the optimal strategy problem by studying the performance of the strategy [Pareto πps($\boldsymbol{\sigma}$), $\hat{\tau}(\mathbf{y})_{GREG}$]. The Cassel et al. result gives background for the conjecture that this strategy is close to "universally optimal" under the above superpopulation model.

Since Pareto πps does not admit exact HT-estimation, the GREG estimator in (7.4) has to be modified a bit. In the sequel πps($\mathbf{s}$) indicates estimation in accordance with (3.1). The modified GREG estimator is;

$$\hat{\tau}(\mathbf{y})_{GREG}^{\pi ps(\sigma)} = \hat{\tau}(\mathbf{y})_{\pi ps(\sigma)} + \hat{B} \cdot [\tau(\mathbf{x}) - \hat{\tau}(\mathbf{x})_{\pi ps(\sigma)}] \quad \text{where} \quad \hat{B} = \hat{\tau}(\mathbf{y} \cdot \mathbf{x}/\boldsymbol{\sigma}^2)_{\pi ps(\sigma)} / \hat{\tau}(\mathbf{x}^2/\boldsymbol{\sigma}^2)_{\pi ps(\sigma)}. \tag{7.5}$$

Theoretical and numerical results on comparison between [Pareto πps($\boldsymbol{\sigma}$), $\hat{\tau}(\mathbf{y})_{GREG}^{\pi ps(\sigma)}$] and various "competing" strategies are presented in Rosén (2000 b). To make a long story short, the findings support the conjecture that the strategy in fact is close to being universally optimal when the superpopulation model is correct.

8 Summarizing conclusions

In Section 4 is argued that Pareto πps should be preferred among πps schemes which admit objective assessment of sampling error. The choice of πps design then stands between Pareto πps and systematic πps(sfo) (= with frame ordered by size). The latter scheme has the following pros and cons. Often it yields more accurate point estimates than Pareto πps, but the opposite also occurs, and it is hard to tell in advance which will be the case in a specific survey situation. On the (very) negative side stands that systematic πps(sfo) deprives assessment of sampling error as well as sample coordination by (permanent) random numbers. We believe that under these premises Pareto πps is seen as the best alternative in most practical survey contexts.

References

Aires, N. (1999). Algorithms to Find Exact Inclusion Probabilities for Conditional Poisson Sampling and Pareto πps Sampling Designs. *Methodology and Computing in Applied Probability* **4** *459 - 473*.

Aires, N. & Rosén, B. (2000). On Inclusion Probabilities and Estimator Bias for Pareto πps Sampling. Statistics Sweden R&D Report 2000 : 2.

Cassel, C. M., Särndal C. - E. & Wretman, J. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika* **63** *615 - 620*.

Feller, W. (1966). An Introduction to Probability Theory and Its Applications, Volume II. Wiley, New York.

Ohlsson, E. (1990). Sequential Poisson Sampling from a Business Register and its Application to the Swedish Consumer Price index. Statistics Sweden R&D Report 1990 : 6.

Ohlsson, E. (1995). Coordination of Samples Using Permanent Random Numbers. In *Business Survey Methods*. John Wiley & Sons, New York.

Ohlsson, E. (1998). Sequential Poisson Sampling. *Journal of Official Statistics*, **14** *149 - 162*.

Rosén, B. (1997a). Asymptotic Theory for Order Sampling. *J Stat Plan Inf*, **62** *135 - 158*.

Rosén, B. (1977b). On Sampling with Probability Proportional to Size. *J Stat Plan. Inf.*, **62** *159 - 191*.

Rosén, B. (2000a). On Inclusion Probabilities for Order Sampling. To appear in *J Stat Plan. Inf.*

Rosén, B. (2000b). Generalized Regression Estimation and Pareto πps. Statistics Sweden R&D Report.

Rosén, B. & Lundqvist P. (1998). Estimation from Order πps Samples with Non - Response. Statistics Sweden R&D Report 1998 : 5.

Saavedra (1995). Fixed Sample Size PPS Approximations with a Permanent Random Number. *1995 Joint Statistical Meetings American Statistical Association*. Orlando, Florida.

Sunter A. B. (1977). List Sequential Sampling with Equal or Unequal Probabilities without Replacement. *Applied Statistics* **26** *261 - 268*.

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

ESTIMATION STRATEGIES USING VARIANTS OF POISSON SAMPLING
DISCUSSION

Pedro J. Saavedra, ORC Macro
ORC Macro, 11785 Beltsville Dr., Calverton, MD 10705
saavedra@macroint.com

This discussion is written from the perspective of a practitioner. And prior to addressing the specific papers, I want to describe my own experience with some of the topics (and authors) presented here. I work for a consulting firm, and for a number of years one of my main clients has been the Department of Energy. Back in the early 1980s a group of statisticians including myself were given the task of designing a sample for the EIA-782, a petroleum product price and volume survey of retailers and resellers that required a given CV for different states, regions and products. We were working with the Petroleum Marketing Division at EIA, and there was concern that the existing design was not sufficiently robust given the large number of estimates to be derived from the sample. To deal with this problem we developed a design consisting of a series of stratified samples linked by a Permanent Random Number (Saavedra, 1988). The PRN not only allowed us to maximize the overlap between the samples, hence reducing the total number of units, but it also permitted us to rotate the sample between cycles.

Originally the idea was for each sample to stand alone, but a policy decision was made that every response had to be used in every estimate. Hence, it was decided that any company that was selected for any State or product would be selected for all. This design had two major deficiencies. The first one was that it had a variable sample size. The second was that there is no known analytic formula that permits us to calculate probabilities of selection, and hence to assign weights. Since this was a list sample, the second deficiency could be partially resolved through simulating repeated draws. This design was used for close to a dozen years, and I was fortunate to be involved as a contractor in much of the work associated with the design.

Five or six years ago I was given the task of redesigning the EIA-782. The major concern was the need to reduce the sample size to about 2,000 companies. I had been dissatisfied with the number of de facto certainties -- units which were among two or three in one stratum for one product in one State. I decided that I would prefer to assign a probability of selection to each company and sample with unequal probabilities (one would say PPS, but it is not clear what measure of size one might use). Naturally, I began by perusing the most comprehensive list of PPS methods in my library -- Brewer and Hanif's (1983) monograph on PPS sampling.

I required a PPS method with a fixed sample size which could be used with a Permanent Random Number in order to permit rotation of the sample. I figured that if Brewer and Hanif had listed no such method, it had to be because none was available (and indeed, none was at the time of its publication). However, I knew of one statistician who frequent faced similar problems to the ones I faced in working for EIA, so I called Phil Kott at the Department of Agriculture. As it happened, Phil had just received a paper by Ebjorn Ohlsson -- a chapter in an upcoming book -- in which Ohlsson described such a method (Ohlsson, 1995). And even though the method did not yield exact probabilities, he cited evidence that it came very close. The method was called Sequential Poisson Sampling (SPS) and it was a form of order sampling which simply calculated probabilities as if one were to use Poisson sampling, divided a random number (possibly a PRN) by the probability of selection and then selected the n smallest quotients. If a Poisson sample were selected with exactly the expected sample size and using the same random numbers SPS would select the identical sample.

Ohlsson's method seemed to do exactly what I needed, but there was something that intuitively bothered me. The PRN one used in Poisson sampling was always between 0 and 1, and this resembled a probability. But then we had a ratio of "probabilities" and in most statistical procedures taking the ratio of odds had proved far more effective than taking the ratio of probabilities. So I tried several variants of SPS, did some simulations, and found that indeed, the formula (r-rp)/(p-rp) worked better that r/p or than other options. Indeed, at some point in the simulations it even seemed to do better than an exact method.

I presented the results at the Joint Meetings in Orlando (Saavedra, 1995) and proposed the use of this approach in the EIA-782. However, there was understandable reluctance to use a method that had not been analytically shown to be sound. Fortunately, at this time Ohlsson notified me that around the time I was presenting in Orlando, Bengt Rosen had not only published results (Rosen, 1995a, 1995b) demonstrating that a method identical to mine

improved Ohlsson's method, but showed that this approach optimized a class of sampling methods just as I suspected. Rosen called his method "Pareto Sampling" for which those of my colleagues who tried to articulate "Odds-Ratio Sequential Poisson Sampling" will be eternally grateful. Armed with Rosen's paper, I completed the design of the new EIA-782 sample. In the first cycle we used simulations to achieve the probabilities of selection, but in the second cycle we implemented the Chromy Algorithm (1987) in what to our knowledge have been its first implementation with Pareto Sampling (Saavedra &Weir, 1999).

Now to turn our attention to the papers. Rosen's paper is a delight from a practitioner's perspective. Ordinarily one sees theoretical papers when a new methodology is developed and one has to sit down and figure out how and when to apply the details. Rosen has taken care of this for the practitioner. Nevertheless, I will bring up a few practical issues. Rosen discusses the problem of adjustment for nonresponse. A related issue is the degree to which one can use order sampling to continue to sample until a fixed number of respondents or a fixed number of in-scope units is obtained. This can be an issue in conjunction with the previous one, but it is also an issue in its own right. Suppose one had an establishment survey and wished to obtain n respondents that sell product x. The frame lists the volume of sales, but is several years old. Suppose one wants to obtain estimated prices and other information that would ordinarily be obtained through a ratio estimator. Can we use Pareto sampling to reach exactly n respondents? If so, how should we adjust our weights, variance estimates and so forth? My usual answer is that I will try several thousand simulations, but experience tells me that asking Bength Rosen may be more productive.

Rosen mentions the advantage that systematic sampling has by using implicit stratification. Indeed, there is something that approximates an implicit stratification method that may be used in Pareto sampling, but it is only effective under certain conditions. I am referring to collocation within subdomains. Suppose I want to use implicit stratification by States. Under systematic sampling one can guarantee that each would be represented proportional to its size (or the sum of the size measures). Under Pareto sampling one can at least reduce the likelihood of extreme oversampling or undersampling. The procedure is simple. You begin with a random number (be it a PRN or a newly assigned one). If there are m units in State j one divides the segment between 0 and 1 into m equal segments, and assigns to each unit a random number within one of the segments, so as to preserve the order of the units. Thus the unit in position q will receive the number (q-r)/m where r is a second random number assigned to the unit. This final random number becomes the number used in Pareto sampling, and its rank order correlation with the original random number will be 1.0 within categories.

Turning our attention to Brewer and Gregoire, the problem that they address is an old one for this practitioner, though I must admit to not always recognizing it in practice. The EIA-782 estimates both volumes and prices, and even though the emphasis is largely on prices (which being the result of ratio estimation are not as affected as total volumes) there is also a concern for volumes. During the design stage of this survey we simulated two versions of the Horvitz-Thompson ratio estimator with postratification. Essentially we applied the equivalent of 3P7 at the stratum level and we applied the traditional postratified estimator (setting the auxiliary variable to unity). Even though we were using Pareto at the national level, sampling for each post-stratum had variable sample size and could thus benefit from adjustments of the HT estimators. At this stage we were doing simulations where we had the actual population totals to compare the results with, and we found that both improved the point estimates, but were not very different in their improvement. The 3P7 estimator was marginally better than the population adjustment estimator. We did not consider the issue of variance estimation at that time.

The final comment overlaps the Kott and Bailey paper. One of the things I like about Phil Kott's work is that I can always count on him to not assume that there is a unique target estimate of interest or a unique auxiliary variable in the frame. In our petroleum sales surveys we start with a frame that has volumes for ten product/end use categories per company per State. We then estimate a somewhat larger number, and estimate most of them for every state and region. We use the Chromy algorithm to obtain probabilities of selection. As I read the Brewer and Gregoire paper I do not know whether to envy them for having a single estimand, or (more constructively) to hope that at ICES III they will present a multivariate version of their paper.

The Kott and Bailey paper discusses methods of extending Brewer selection probability to a multivariate situation. Kott and Bailey propose the derivation of univariate probabilities of selection and the use of the maximum probability for sampling selection. To a certain extent this is what was done in the first cycle of the new EIA-782 design. However, an approach that seems to work better is the use of the Chromy algorithm. Most people are familiar of the use of Chromy to establish optimal allocations when there are different target variables and many

strata. But the algorithm, or at least the implementation which Zayatz and Sigman (1995) have put in place in the census may also be used for unequal probability sampling (a term more accurate than PPS when there is no measure of size that the probabilities are proportional to).

In the Chromy algorithm each unit becomes a stratum. When the auxiliary variable represents a total we assume the standard deviation to be proportional to the value of the auxiliary variable (in our case the yearly volume of sales of the product at the frame level). Then a fractional allocation capped at 1 is assigned to the units, and this becomes the probability of selection. We in fact tested the algorithm by simulating 1,000 samples, estimating the totals from the sample and calculating the mean square errors of the estimates.

One point regarding model based estimators which ought to be mentioned here is that what works for the entire sample and population works for strata when one uses substratification schemes. This greatly multiplies the number of estimators one might consider. When we moved from linking strata to PPS, we still resorted to poststratification to refine our estimates. Unlike the approach Kott and Bailey present, we did not calibrate the weights to a single weight, but adjusted the Horvitz-Thompson estimators separately for each target variable and area. Petroleum surveys are more complex than agricultural surveys by virtue of the fact that a company can sell in several states.

In conclusion, the three papers presented today can be related to one practitioner's work in one survey. It is entirely possible that one or more of the papers could influence the development of this one survey as all three of the presenters (as well as the chair of this session) already have.

Brewer, K.R.W. and Hanif, M., (1983), Sampling with Unequal Probabilities, New York: Springer-Verlag.

Chromy, J. (1987) "Design Optimization with Multiple Objectives," Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 194-199

Ohlsson, E. (1995), "Coordination of Samples Using Permanent Random Numbers", Survey Methods for Business, Farms and Institutions, edited by Brenda Cox, New York: Wiley.

Rosen, B. (1995a) "On Sampling with Probability Proportional to Size", R&D Report 1995:1, Stockholm, Statistics Sweden.

Rosen, B. (1995b) "Asymptotic Theory for Order Sampling", R&D Report 1995:1, Stockholm, Statistics Sweden.

Saavedra, P. J. (1988) "Linking Multiple Stratifications: Two Petroleum Surveys". 1988 Joint Statistical Meetings, American Statistical Association, New Orleans, Louisiana.

Saavedra, P.J. (1995) "Fixed Sample Size PPS Approximations with a Permanent Random Number", 1995 Joint Statistical Meetings, American Statistical Association, Orlando, Florida.

Saavedra, P.J. and Weir, P. (1999) "Application of the Chromy Allocation Algorithm with Pareto Sampling", 1999 Joint Statistical Meetings, American Statistical Association, Baltimore, MD.

Zayatz, L. and Sigman, R. (1995) Chromy_Gen: General-Purpose Program for Multivariate Allocation of Stratified samples Using Chromy's Algorithm, Economic Statistical Methods Report series ESM-9502, June 1995, Bureau of the Census.