United States Department of Agriculture
Forest Service
Rocky Mountain Research Station
Research Note RMRS–RN–4
September 1998

Poisson Sampling – The Adjusted and Unadjusted Estimator Revisited

Michael S. Williams, Hans T. Schreuder, and Gerardo H. Terrazas

Abstract—The prevailing assumption, that for Poisson sampling the adjusted estimator $\hat{Y}_a$ is always substantially more efficient than the unadjusted estimator $\hat{Y}_u$, is shown to be incorrect. Some well known theoretical results are applicable since $\hat{Y}_a$ is a ratio-of-means estimator and $\hat{Y}_u$ a simple unbiased estimator. We formalize an additional realistic situation for high-value timber estimation for which $\hat{Y}_u$ is more efficient. Here $y_i \approx \beta x_i$ for all but a few units in a population for which y_i is very large and x_i very small. This is a common situation in estimating the net volume of high-value standing timber such as that found in the Pacific Northwest region of the United States. Basically this means that $\hat{Y}_a$ is sensitive to some types of valid data. The generalized regression estimator and an approximate Srivastava estimator are compared with $\hat{Y}_a$ and $\hat{Y}_u$ and shown to be not as sensitive to such data points. Simulations on a small population illustrate these ideas.

Keywords: poisson sampling, adjusted estimator, unadjusted estimator, generalized regression estimator, approximate Srivastava estimator

Williams, M.S.; Schreuder, H.T.; Terrazas, G.H. 1998. **Poisson sampling – The adjusted and unadjusted estimator revisited.** Res. Pap. RMRS–RN–4. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 10 p.

Introduction

Hajek (1958, 1964) introduced Poisson sampling into the statistical literature. It is defined as a sampling design in which the sample units have unequal probabilities of selection, π_i. In addition the units in the population are independent and the sample size, n, is a random variable. Hajek proposed the unbiased estimator for the population total given by

$$\hat{Y}_u = \sum_{i=1}^{n} \frac{y_i}{\pi_i} \qquad [1]$$

where $y_i =$ the value of interest for unit $i, \pi_i = n_e x_i / X$ is the probability of selecting unit i based on a covariate x_i, n is the achieved sample size with expected sample size $E[n] = n_e$, and X is the sum of all x_i in the population.

Grosenbaugh (1964) introduced 3-P sampling into the forestry literature as an alternative to Poisson sampling. He used a slightly biased estimator defined by

$$\hat{Y}_a = \frac{n_e}{n} \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \frac{n_e}{n} \hat{Y}_u, \qquad [2]$$

which will be referred to as the adjusted estimator.

The literature contains many examples showing that $\hat{Y}_a$ is more efficient than $\hat{Y}_u$ (Grosenbaugh 1964, Schreuder et al. 1968, Furnival et al. 1987). Recently, occasions were observed where $\hat{Y}_u$ is at least as efficient as $\hat{Y}_a$. These situations arose for high-value timber stands in the Pacific Northwest where an accurate estimate of net volume was desired. In order to obtain a good estimate of net volume in a stand, cutting down and destructively measuring trees for their actual volume may be desirable. In this case, Poisson sampling may be appropriate where an ocular estimate of net volume while the tree is still standing is the covariate x_i. Thus, sampling is proportional to estimated standing tree net volume. It can be very difficult to ascertain net volume on a standing tree and even experienced timber cruisers severely overestimate or underestimate net volume of some trees, particularly for larger trees. However, these estimates are valid data. Thus, there is no reason to remove these points from an estimate net volume. It is in situations where the ocular estimate of net volume (x_i) is very small and the actual net volume (y_i) is large that problems with the adjusted estimator arise.

Literature Review

Poisson sampling is a strategy based on using unequal probabilities of selection for each of the elements in the sample. If the sample membership indicator I_i is given by

$$I_i = 1 \text{ if the } i^{th} \text{ element is in the sample}$$
$$I_i = 0 \text{ otherwise}$$

for each $i = 1, 2, \ldots N$, then

$$P[I_i = 1] = \pi_i$$

and

$$P[I_i = 0] = 1 - \pi_i.$$

The probability that any sample s is chosen under Poisson sampling is given by

$$P[s] = \prod_{i \in s} \pi_i \prod_{i \notin s} (1 - \pi_i).$$

In Poisson sampling the sample size n is random with mean $E[n] = \sum_{i=1}^{N} \pi_i$ and variance $V[n] = \sum_{i-1}^{N} \pi_i (1 - \pi_i)$.

Under Poisson sampling the variance for the unadjusted estimator given in [1] is given by

$$V(\hat{Y}_u) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \sum_{i=1}^{N} y_i^{\,2} \quad . \qquad [3]$$

The adjusted estimator, given in [2], was proposed by Grosenbaugh (1964). This estimator is slightly biased, but generally has a smaller variance than the unadjusted estimator. The approximate variance for the adjusted estimator is given by

$$V(\hat{Y}_a) \doteq [\sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \frac{Y^2}{n_e}](1 + \frac{V(n)}{n_e^2}), \qquad [4]$$

where

$$V(n)=\sum_{i=1}^{N}\pi_i-\sum_{i=1}^{N}\pi_i^2. \qquad [5]$$

Since its introduction, $\hat{Y}_a$ has been considered more efficient than $\hat{Y}_u$ (Grosenbaugh 1964, Schreuder et al. 1968, Furnival et al. 1987). Schreuder et al. (1968) compared the efficiency of the adjusted and unadjusted estimators using two highly correlated populations. When the correlation coefficient (ρ_{xy}) between x_i and y_i exceeded .95, they found the ratio of Monte Carlo estimated variances, V_u / V_a, for the unadjusted and adjusted estimators to range from 4.20 to 9.38 for one population and 6.74 to 12.17 for the other population. The differing ratios were generated using different sampling fractions. Van Deusen (1987) states that the 3-P adjusted estimator has nearly optimal properties under a regression superpopulation model and can be expected to perform well except under extreme deviations for this model. Schreuder (1987) describes a situation where the model between x_i and y_i violate the superpopulation model. The results in Van Deusen (1987) and Schreuder (1987) indicate that $\hat{Y}_a$ can perform poorly, but the performance of $\hat{Y}_u$ in these situations is undocumented.

Furnival et al. (1987) and Van Deusen (1987) note that $\hat{Y}_a$ can be rewritten as

$$\hat{Y}_a=(\hat{Y}_u / \hat{X}_u)X, \qquad [6]$$

where $\hat{X}_u$ is defined analogously to $\hat{Y}_u$ in [1] and X is the sum of the covariate values. Hence, the comparison in efficiency of the ratio-of-means estimator $\hat{Y}_a$ and the unadjusted estimator of the population total $\hat{Y}_u$ is applicable. Thus, the relative efficiency of $\hat{Y}_a$ to $\hat{Y}_u$ is

$$RE=\frac{V(\hat{Y}_u)}{V(\hat{Y}_a)}\approx\left(1+R^2\frac{V(\hat{X}_u)}{V(\hat{Y}_u)}-2\rho R\frac{\sqrt{V(\hat{X}_u)}}{\sqrt{V(\hat{Y}_u)}}\right)^{-1} \qquad [7]$$

where $\rho=\frac{Cov(\hat{X}_u,\hat{Y}_u)}{\sqrt{V(\hat{X}_u)V(\hat{Y}_u)}}$, $R=\frac{Y}{X}$, and Y and X are the population totals for the variable of interest and covariate respectively. Then, $\hat{Y}_u$ is more efficient than $\hat{Y}_a$ provided $\rho<\frac{1}{2}\left(\frac{CV(\hat{X}_u)}{CV(\hat{Y}_u)}\right)$ (Cochran 1977).

These results seem straightforward but do not explain why $\hat{Y}_u$ is more efficient than $\hat{Y}_a$ for populations like the one described later. It can be shown that if the following minimum list of assumptions is met then the adjusted estimator will be at best as efficient as the unadjusted estimator.

1. $n_e^2 \geq N$, where n_e is the expected sample size and N is the total number of trees in the population.
2. There exists at least one point $y_k \geq \overline{X}+e_k$ with βx_k small relative to y_k for at least one k.
3. The coefficient β is such that $\beta^2 \leq \frac{X}{N^2(n_2-1)}$.
4. $X-1 \approx X$.
5. $(\beta(X-1)+\overline{X})^2 \approx \beta^2 X^2$.

The proof is given in the appendix.

Alternative estimators exist for Poisson sampling. Ouyang and Schreuder (1993) describe an unbiased Srivastava estimator for Poisson sampling in a forestry setting. The Srivastava estimator

$$\hat{Y}_{Sra1} = \frac{1}{|s|}\sum_{i \in s} \frac{y_i}{\pi_i^{'}} \text{ with } \pi_i^{'} = \sum_{i \in s} \frac{\rho(s)}{|s|} \qquad [8]$$

has zero variance if $\pi_i \, \alpha \; y_i$. If pps sampling is used, the Srivastava estimator can be approximated by the following ratio estimator

$$\hat{Y}_{Sra2} = \frac{\sum_{i=1}^{n} \frac{y_i}{\pi_i}}{\sum_{i=1}^{n} \frac{y_i^*}{\pi_i}} Y^* \qquad [9]$$

(Ouyang et al. 1992) where $y_i^* = \alpha + \beta x_i$ and $Y^* = \sum_{i=1}^{N} y_i^*$. This ratio estimator is not an unbiased estimator, but Ouyang and Schreuder (1993) found the bias to be small provided a linear relationship exists between y and x.

Särndal (1980,1982) gives the asymptotically unbiased generalized regression estimator

$$\hat{Y}_{GR} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} + \alpha(N - \sum_{i=1}^{n} \frac{1}{\pi_i}) + \beta(X - \sum_{i=1}^{n} \frac{x_i}{\pi_i}). \qquad [10]$$

Van Deusen (1987) shows that $\hat{Y}_a$ and $\hat{Y}_u$ are special cases of the $\hat{Y}_{GR}$, where $\hat{Y}_u$ uses no model and $\hat{Y}_a$ uses a ratio model with variance proportional to x_i^2.

This paper includes these alternative estimators to demonstrate which of these estimators might be a good choice in similar survey situations.

Data Description and Results

For a data set described by Johnson and Hartman (1972), ocular estimates of net volume $(= x_i)$ and the net volume obtained by destructive sampling $(= y_i)$ are given. The size of the data set was $N = 131$. For the simple linear regression model $R^2 = 0.9505$ and $y = 1.0469x$. Figure 1 shows the net volume y versus the ocularly estimated net volume x. This data set contains one outlying point where x_i is very small in relation to y_i. Given the high correlation of the model and the results of Schreuder et al. (1968), the adjusted estimator was expected to produce standard deviations which were substantially smaller than the unadjusted estimator. The first line of table 1 contains the true standard deviation as a percentage of the total net volume for the population of $N = 131$ units.

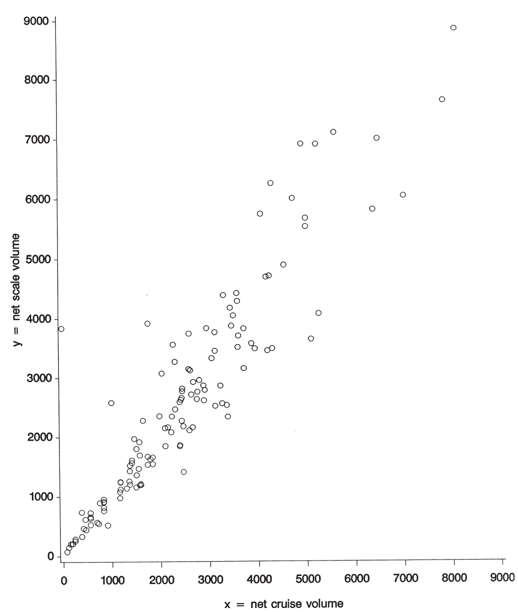

Figure 1. Net scale volume – Net cruse volume relationship

Table 1. Population standard deviations as a percentage of Y.

Population	$\hat{Y}_a$	$\hat{Y}_u$
$N = 131$	108.832	108.699
$N = 130$	3.840	13.641

The true variance for the population, $N = 131$, is slightly larger for the adjusted estimator than for the unadjusted estimator. The cause of this is the one outlying point where x_i is very small in relation to y_i. This produced a very small probability of selection, π_i, in conjunction with an average value of y_i. By deleting this point from the data set, the standard deviation values in the second line of the table are produced ($N = 130$). These values fall in line with the traditional belief that the adjusted estimator is more efficient than the unadjusted estimator. To better understand the cause of this situation, the variance equations for the adjusted and unadjusted estimator need to be studied. For the unadjusted estimator the variance is given as

$$V(\hat{Y}_u) \div \sum_{i=1}^{N} \frac{y_i^{\ 2}}{\pi_i} - \sum_{i=1}^{N} y_i^{\ 2} = (T_1 - T_2), \qquad [11]$$

where $T_1 = \sum_{i=1}^{N} y_i^2 / \pi_i$ and $T_2 = \sum_{i=1}^{N} y_i^{\ 2}$. For the adjusted estimator the approximate variance is given by

$$V(\hat{Y}_a) \doteq [\sum_{i=1}^{N} \frac{y_i^{\ 2}}{\pi_i} - \frac{Y^2}{n_e}](1 + \frac{V(n)}{n_e^{\ 2}}) = (T_3 - T_4)T_5, \qquad [12]$$

where $T_3 = T_1$, $T_4 = Y^2 / n_e$, and $T_5 = 1 + V(n) / n_e^2$. The magnitudes of each term of equations [11] and [12] are listed in table 2 for the data set with and without the outlier data point included. The first line of the table indicates that terms T_1 and T_3 dominate the variance equations when the outlier data point is included. When T_3 is multiplied by the variable sample size correction term, T_5, the adjusted estimator can be less efficient than the unadjusted estimator.

Table 3 contains iterated standard deviations as a percentage of the total net volume after 200,000

Table 2. Magnitudes of terms in equations [5] and [6].

Population	T_1	T_2	T_3	T_4	T_5
$N = 131$	1.381×10^{11}	1.304×10^{9}	1.381×10^{11}	3.308×10^{9}	1.017
$N = 130$	3.396×10^{9}	1.289×10^{9}	3.396×10^{9}	3.235×10^{9}	1.017

simulations using the two data sets previously mentioned. As the results indicate, $\hat{Y}_{Sra2}$ is substantially more efficient than either $\hat{Y}_u$ or $\hat{Y}_a$ for $N = 131$. The reason $\hat{Y}_{Sra2}$ is more efficient is that when the outlier data point is selected y_k and y_k^* divided by π_k produces very large overestimates of

Table 3. Iterated standard deviations as a percentage of $\sum_{i=1}^{N} y_i$.

Population	$\hat{Y}_a$	$\hat{Y}_u$	$\hat{Y}_{Sra1}$	$\hat{Y}_{Sra2}$	$\hat{Y}_{GR}$
$N = 131$	114.094	111.328	115.353	3.568	3.780
$N = 130$	3.151	13.835	3.164	3.479	3.708

the Y. Since these terms appear in the numerator and denominator, the effect of the overestimate is reduced. The generalized regression estimator, $\hat{Y}_{GR}$, also shows a considerable increase in efficiency over $\hat{Y}_a$ and $\hat{Y}_u$. The reason $\hat{Y}_{GR}$ is more efficient is because when the outlier point is selected y_k / π_k, α / π_k, and $\beta x_k / \pi_k$ all produce very large overestimates. When the terms $\sum_{i=1}^{n} \alpha / \pi_i$ and $\sum_{i=1}^{n} \beta x_i / \pi_i$ are subtracted from $\sum_{i=1}^{n} y_i / \pi_i$, the effect of the outlier point is reduced. Estimator $\hat{Y}_{Sra1}$ is as inefficient as $\hat{Y}_a$ and $\hat{Y}_u$ because there is no compensation for outlier points as occurred for $\hat{Y}_{Sra2}$ and $\hat{Y}_{GR}$. It should be noted that 200,000 simulations were not sufficient to fully stabilize the estimates of $\hat{Y}_a$ and $\hat{Y}_u$, but due to limited computing time no attempt was made to perform additional simulations.

For $N = 130, \hat{Y}_a, \hat{Y}_{Sra1}, \hat{Y}_{Sra2}$ and $\hat{Y}_{GR}$ are about equally efficient.

Conclusions and Recommendations

Since its introduction, $\hat{Y}_a$ has been considered more efficient than $\hat{Y}_u$. In situations where an outlier point x_k is small and βx_k much less than the corresponding y_k value, it is reasonable to expect $\hat{Y}_a$ to be less efficient than $\hat{Y}_u$. This occurs, for example, when ocular estimates of net volume are used to predict actual net volume. $\hat{Y}_{Sra2}$ and $\hat{Y}_{GR}$ perform quite well in the presence of these types of outlier observations. $\hat{Y}_{GR}$ would generally be a better choice than $\hat{Y}_{Sra2}$ since it has a reliable variance estimate (Schreuder and Ouyang 1992) and is asymptotically unbiased and efficient.

Literature Cited

Cochran, W.G. 1977. Sampling Techniques. Third ed. John Wiley and Sons, New York. 428 p.

Furnival, G.M., T.M. Gregoire, and L.R. Grosenbaugh. 1987. Adjusted inclusion probabilities with 3P sampling. For. Sci. 33:617-631.

Grosenbaugh, L.R. 1964. Some suggestions for better sample-tree measurement. Proc. Soc. Am. For., Oct. 1963, Boston, Massachusetts. Soc. Am. For. Bethesda, Maryland.

Hajek, J. 1958. Some contributions to the theory of probability sampling. Bull. Int. Stat. Inst. 36(3):127-134.

Hajek, J. 1964. Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Stat. 35:1491-1523.

Johnson, F.A., and G.B. Hartman, Jr. 1972. Fall, buck, and scale cruising. J. For. Sci. Sept. 1972. 566-568.

Ouyang, Z., and H.T. Schreuder. 1993. Approximate Srivastava estimation in forestry. For. Sci. 39:309-320

Ouyang, Z., J.N. Srivastava, and H.T. Schreuder. 1992. A general ratio estimator and its application to regression model sampling. Ann. Inst. Stat. Math. (in press).

Särndal, C.E. 1980. A two-way classification of regression estimation strategies in probability sampling. Can. J. Stat. 8:165-177.

Särndal, C.E. 1982. Implications of survey design for generalized regression estimation of linear functions. J. of Stat. Plan. Inf. 7:155-170.

Schreuder, H.T., J. Sedransk, and K.D. Ware. 1968. 3-P sampling and some alternatives, I. For. Sci. 14:429-454.

Schreuder, H.T. 1987. Design-dependent and model-dependent procedures in timber sale and updating surveys. Proc. Soc. of Am. For., National Communication, Birmingham, Alabama. Oct. 5-8, 1986. 74-76.

Schreuder, H.T., and Z. Ouyang. 1992. Optimal sampling strategies for weighted linear regression estimation. Can. J. For. Res. 22:239-247.

Van Deusen, P.C. 1987. 3-P sampling and design versus model-based estimates. Can. J. For. Res. 17:115-117.

Appendix

Schreuder et al. (1968) give the true variance $V(\hat{Y}_u)$ of $\hat{Y}_u$ and an approximate variance of $\hat{Y}_a$, $V(\hat{Y}_a)$ as

$$V(\hat{Y}_u) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \sum_{i=1}^{N} y_i^2 \tag{1}$$

and

$$V(\hat{Y}_a) \doteq [\sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \frac{Y^2}{n_e}](1 + \frac{V(n)}{n_e^2}) \tag{2}$$

where

$$V(n) = \sum_{i=1}^{N} \pi_i - \sum_{i=1}^{N} \pi_i^2. \tag{3}$$

We wish to show that under certain conditions $V(\hat{Y}_a) \geq V(\hat{Y}_u)$. For the adjusted estimator the variable sample size factor can be approximated by

$$(1 + \frac{V(n)}{n_e^2}) = (1 + \frac{\sum_{i=1}^{N} \pi_i - \sum_{i=1}^{N} \pi_i^2}{n_e^2}) = (1 + \frac{1}{n_e} - \frac{\sum_{i=1}^{N} x_i^2}{(\sum_{i=1}^{N} x_i)^2}) \approx (1 + \frac{1}{n_e}) \tag{4}$$

since for even moderately sized populations $V(n) \approx n_e$ because

$$\sum_{i=1}^{N} x_i^2 << (\sum_{i=1}^{N} x_i)^2. \tag{5}$$

Let $V(\hat{Y}_a) \approx V(\hat{Y}_a^*) = [\sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \frac{Y^2}{n_e}](1 + \frac{1}{n_e})$, then in order to have $V(\hat{Y}_a^*) \geq V(\hat{Y}_u)$, we want

$$V(\hat{Y}_a^*) - V(\hat{Y}_u) = \frac{1}{n_e} \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \frac{1}{n_e}(Y^2 + \frac{Y^2}{n_e}) + \sum_{i=1}^{N} y_i^2 \geq 0 \tag{6}$$

which implies

$$\frac{1}{n_e} \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} + \sum_{i=1}^{N} y_i^2 \geq \frac{Y^2}{n_e} + \frac{Y^2}{n_e^2}. \tag{7}$$

Now

$$\sum_{i=1}^{N} y_i^2 \geq \frac{Y^2}{n_e^2}, \text{ provided } n_e^2 \geq N. \tag{8}$$

If this condition holds, these two terms may be dropped from [7]. All that remains to be shown is that under the simplifying assumptions and a given population model

$$\sum_{i=1}^{N} \frac{y_i^2}{\pi_i} \geq Y^2. \qquad [9]$$

Some simplifying assumptions need to be made. A widely used model in forestry is

$$y_i = \beta x_i + e_i \qquad [10]$$

with $E[e_i] = 0,\ E[e_i^2] = \sigma^2 x_i^2$ with $E[e_i e_j] = 0$ for $i \neq j$. In order to formalize conditions under which $V(\hat{Y}_a^*) \geq V(\hat{Y}_u)$, consider the following alternative model

$$y_i = \beta x_i + e_i \qquad [11]$$

for $i = 1, 2, \ldots k-1, \ldots k+1, \ldots N$ with

$$y_k = \overline{X} \text{ and } x_k = 1 \qquad [12]$$

with $E[e_i] = 0,\ E[e_i^2] = \sigma^2 x_i^2$, and $E[e_i e_j] = 0$ for $i \neq j$ for all $i = 1, 2, \ldots N$. In [12] $\overline{X}$ is the mean of the x values in the population and $x_k = 1$ is used for simplicity. Any case where $x_k > 0$ and βx_k is sufficiently small compared to y_k will produce the same results. The model given in [11] and [12] is common in forestry when ocular estimation has to be used (D. Bruce - personal communication). The results which follow can be generalized to a population where more than one outlier exists, but for the sake of simplicity only one outlier will be considered. Using models [11] and [12] in conjunction with equation [9], we have

$$\frac{1}{n_e}\left(\sum_{i \neq k}^{N} \frac{X(\beta x_i + e_i)(\beta x_i + e_i)}{n_e x_i} + \frac{X(\overline{X} + e_k)(\overline{X} + e_k)}{n_e x_k}\right) \geq \frac{(\beta(X-1) + \overline{X})^2}{n_e}, \qquad [13]$$

which can be written as

$$\frac{1}{n_e}\sum_{i \neq k}^{N} \frac{X\beta^2 x_i^2}{n_e x_i} + \frac{2}{n_e}\sum_{i \neq k}^{N} \frac{X\beta x_i e_i}{n_e x_i} + \frac{1}{n_e}\sum_{i \neq k}^{N} \frac{X e_i^2}{n_e x_i} + \frac{X\overline{X}^2}{n_e^2} + \frac{2X\overline{X}e_k}{n_e^2} + \frac{X e_k^2}{n_e^2} \geq \frac{(\beta(X-1) + \overline{X})^2}{n_e}. \qquad [14]$$

Due to the assumed error structure, this reduces to

$$\frac{\beta^2 X(X-1)}{n_e^2} + \frac{\sigma^2 X(X-1)}{n_e^2} + \frac{X\overline{X}^2}{n_e^2} + \frac{\sigma^2 X}{n_e^2} \geq \frac{(\beta(X-1) + \overline{X})^2}{n_e}. \qquad [15]$$

Ignoring the error terms, it is sufficient to show

$$\frac{\beta^2 X(X-1)}{n_e^2} + \frac{X\overline{X}^2}{n_e^2} \geq \frac{(\beta(X-1) + \overline{X})^2}{n_e}. \qquad [16]$$

Equation [16] can be rewritten as

$$\frac{\beta^2 X(X-1)}{n_e^2} + \frac{X\overline{X}^2}{n_e^2} > \frac{\beta^2 X^2}{n_e} \qquad [17]$$

assuming $(\beta(X-1)+\overline{X})^2 \approx \beta^2 X^2$, which reduces to

$$\frac{X}{N^2} > (n_e - 1)\beta^2, \qquad [18]$$

assuming $X - 1 \approx X$. For the data set discussed in this paper, x_i and y_i are both estimates of net volume so $\beta \approx 1$, and $X >> N^2(n_e - 1)$. If the following minimum list of assumptions is met, then the adjusted estimator will be at best as efficient as the unadjusted estimator.

1. $n_e^2 \geq N$, where n_e is the expected sample size and N is the total number of trees in the population.
2. There exists at least one point $y_k \geq \overline{X} + e_k$ with βx_k small relative to y_k for at least one k.
3. The coefficient β is such that $\beta^2 \leq \dfrac{X}{N^2(n_e - 1)}$.
4. $X - 1 \approx X$.
5. $(\beta(X-1)+\overline{X})^2 \approx \beta^2 X^2$.

The Authors

Michael S. Williams is a mathematical statistician with the USDA Forest Service Rocky Mountain Research Station Fort Collins, Colorado.

Hans T. Schreuder is a mathematical statistician with the USDA Forest Service Rocky Mountain Research Station, Fort Collins, Colorado.

Gerardo H. Terrazas is a Ph.D. with the Department of Statistics, Colorado State University, Fort Collins, Colorado.

You may order additional copies of this publication by sending your mailing information in label form through one of the following media. Please send the publication title and number.

Telephone (970) 498-1719

E-mail rschneider/rmrs@fs.fed.us

FAX (970) 498-1660

Mailing Address Publications Distribution
Rocky Mountain Research Station
3825 E. Mulberry Street
Fort Collins, CO 80524-8597

ROCKY MOUNTAIN
Research Station

The Rocky Mountain Research Station develops scientific information and technology to improve management, protection, and use of forests and rangelands. Research is designed to meet the needs of National Forest managers, federal and state agencies, public and private organizations, academic institutions, industry, and individuals.

Studies accelerate solutions to problems involving ecosystems, range, forests, water, recreation, fire, resource inventory, land reclamation, community sustainability, forest engineering technology, multiple use economics, wildlife and fish habitat, and forest insects and diseases. Studies are conducted cooperatively, and applications can be found worldwide.

Research Locations

Flagstaff, Arizona
Fort Collins, Colorado*
Boise, Idaho
Moscow, Idaho
Bozeman, Montana
Missoula, Montana
Lincoln, Nebraska
Reno, Nevada
Albuquerque, New Mexico
Rapid City, South Dakota
Logan, Utah
Ogden, Utah
Provo, Utah
Laramie, Wyoming

* Station Headquarters, 240 West Prospect Road, Fort Collins, CO 80526