Evaluation of probability proportional to predictions estimators of total stem volume

Steen Magnussen

Abstract: A plethora of probability proportional to predictions (PPP) estimators makes it hard for a user to decide which one to use. This study demonstrates the need for an extensive screening procedure by example of four PPP estimators of total stem volume and five estimators of sampling error. Bias, absolute bias, root mean square error, sample-based estimators of sampling error, and achieved significance levels of confidence intervals with a nominal significance level were compared across 832 distinct settings. Population size, sample size, the variance and skewness of the volume predictors, and the strength of the correlation and the slope between predicted and actual stem volume varied between settings. Estimators converged in performance as sample sizes increased but were otherwise sensitive to actual settings. Of the tested estimators, Brewer's "cosmetically calibrated" estimator was consistently the best in terms of mean absolute relative bias and generally favored in an overall assessment of five performance criteria. Grosenbaugh's adjusted estimator was a close second and was often ranked first in overall performance when $n > 0.15N$.

Résumé : Étant donné la grande diversité d'estimateurs de probabilté proportionnelle aux prédictions (PPP) il est difficile pour l'utilisateur de décider lequel utiliser. Cette étude démontre la nécessité d'une procédure intensive de sélection en prenant pour exemple quatre estimateurs PPP du volume total de la tige et de cinq estimateurs de l'erreur d'échantillonnage. Le biais, le biais absolu, l'erreur moyenne quadratique, les estimateurs de l'erreur d'échantillonnage basée sur l'échantillon et les seuils de signification des intervalles de confiance obtenus avec une valeur nominale de signification sont comparés à l'aide de 832 groupes de données distincts. La taille de la population, la taille de l'échantillon, la variance et l'asymétrie des prédicteurs du volume, le degré de corrélation et la pente du volume prédit sur le volume réel de la tige varient selon le groupe de données. La performance des estimateurs converge avec l'augmentation de la taille de l'échantillon mais ceux-ci sont sensibles aux groupes réels. Parmi les estimateurs testés, l'estimateur artificiellement calibré de Brewer est constamment le meilleur en terme de biais moyen relatif absolu et est généralement favori pour l'évaluation de l'ensemble des cinq critères de performance. L'estimateur ajusté de Grosenbaugh suit de très près au deuxième rang mais il occupe souvent le premier rang pour sa performance générale lorsque $n > 0{,}15N$.

[Traduit par la Rédaction]

Introduction

Extraction of tree volume from a forest stand is often preceded by a volume inventory to determine the amount, composition, and type of wood products in the stand. Development of fast, cost-efficient, and accurate methods of volume estimation has a long tradition in forestry (Avery and Burkhart 1983; Clutter et al. 1983). Fortuitous relationships among various tree attributes, such as diameter, height, and taper, are exploited to make predictions of wood volume from a simple and fast to measure attribute such as basal area or diameter at breast height. These predictions can often be improved significantly by measuring one or a few additional attributes, such as upper stem diameter(s) or the height to various predefined positions along the bole on a subset of trees (Lynch 1995; Turnblom and Burk 1996; Van Deusen 1987; Yamamoto 1994; Zakrzewski and Termikaelian 1994). Sampling the additional attributes with probability proportional to predictions (PPP) is economically and statistically attractive and has a proven track record in forestry (Schreuder et al. 1993; Shiver and Borders 1996).

Received April 12, 2001. Accepted August 30, 2001. Published on the NRC Research Press Web site at http://cjfr.nrc.ca on December 20, 2001.

S. Magnussen. Natural Resources Canada, Canadian Forest Service, 506 West Burnside Road, Victoria, BC V8Z 1M5, Canada (e-mail: smagnuss@pfc.forestry.ca).

Total volume TY of a set of trees in a stand is the attribute of interest. The set includes either all trees of interest or a sample representing a known area. The number of trees N is known either prior to, or after, sampling (deVries 1986). The volume of an individual tree Y_i (i = 1, 2, ..., N) is only determined on a subset of trees. A reasonable predictor of volume X_i is available for each of the N trees. Although the total TX of the predictors X_i is an estimator of TY, it is decided that improvement is warranted. A subset of trees with a nominal size of n is now selected with PPP sampling, and their actual volume Y_i is determined. The actual volume may be an improved estimate considered as error free, for example, a volume estimate obtained after a few additional measurements of height and diameter (Lynch 1995). The number of trees actually selected is n_s, where subscript s denotes the sample ($0 \le n_s \le N$, $E(n_s) = n$).

For a given target sample size n, the inclusion probability p_i of tree i (i = 1, 2, ..., N) is calculated as

$$[1] \qquad p_i = n\left(\frac{X_i}{\mathrm{TX}}\right)$$

 DOI: 10.1139/X01-173

where X_i is the available predictor of the tree's actual volume Y_i, and TX is the sum of the N predictors. In the event $p > 1$, the ith tree would be assured inclusion, and its inclusion probability would be set to 1. The inclusion probabilities of the remaining trees were then recalculated as above with n replaced by $n - 1$, and so on until no inclusion probability was larger than 1. A decision to select a tree for volume estimation was determined by N random draws from a uniformly distributed variable u_i on the interval [0, 1]. If $u_i \leq p_i$, then tree i would be selected for estimation of volume (Y_i), otherwise not ($i = 1, 2, ..., N$). Sampling is thus without replacement (a tree can enter only once in the sample).

Empty samples ($n_s = 0$) were rejected, and the selection process was repeated until at least one tree was sampled for volume estimation. Inclusion probabilities were adjusted to reflect the rejection of zero samples $\tilde{p}_i = p_i/(1 - P_0)$, where P_0 is the probability of obtaining a zero sample, $P_0 = \prod_{i=1}^{N}(1 - p_i)$. Note that P_0 was negligible in all cases ($P_0 <$ 0.0023, $\overline{P}_0 = 0.0003$). The set of trees selected for volume estimation is denoted by s, and the number of trees in s by n_s. The expected sample size

$$E[n_s] = \sum_{i=1}^{N} \frac{1}{\tilde{p}_i}$$

which is exactly n when all inclusion probabilities p_i are less than, or equal to, 1.

Selection of trees with PPP has the obvious advantage of concentrating sampling efforts to trees making a larger than average contribution to the stand volume (Brewer and Hanif 1983; Cochran 1977; deVries 1986; Thompson 1992). Unequal probability sampling is mainly a tool to reduce the sampling variance of an estimate or, conversely, a tool to reduce sample sizes for a fixed accuracy. Foresters were therefore quick to explore the potential of unequal probability sampling in forest inventory procedures (Bonnor 1972; Grosenbaugh 1965; Hartman 1967; Schreuder et al. 1968, 1971; Space and Turman 1977).

In PPP sampling the realized sample size n_s is random with a mean equal to a predefined target n (Brewer and Hanif 1983). Estimates of PPP sampling errors are therefore quite variable; a definite drawback from a user's perspective, since the success of a volume inventory depends on its accuracy (Bonnor 1972; deVries 1986; Magnussen 2000; Schreuder et al. 1968, 1971). To combat the high variability of PPP estimates a variety of alternative estimators have been coined (Brewer 1999; Brewer et al. 2000; Brewer and Hanif 1983; Furnival et al. 1987; Gregoire and Valentine 1999; Särndal 1996).

Users of PPP volume estimators have a plethora of choices (Brewer and Hanif 1983; Gregoire and Valentine 1999; Särndal et al. 1992; Schreuder et al. 1992, 1993). Which one to use? A review of the literature would only produce partial clues, as most estimators have only been used or tested in a narrow range of conditions. It is known that the performance of an estimator can vary in response to population attributes such as the distribution and variance–covariance structure of the predictor and the target variable (Magnussen 2000; Schreuder et al. 1968, 1971; Williams and Schreuder 1998). In forestry, where data structures varie widely, an estimator that performs well across a wide spectrum of data structures would be preferable to an estimator with fluctuating performance.

The objectives of this study are to demonstrate the need for an extensive testing scheme for PPP estimators and to provide a suitable testing procedure. An exhaustive testing of all PPP estimators is neither intended nor desired. Instead a select group of four PPP estimators and five variance estimators are compared. The selected estimators represent a mix of estimators that are well known to forestry and new developments. Also, as sample sizes increases their performance converges.

Simulated sampling

Nominal sample sizes of n = 6, 8, 10, 11, and 13 were taken from a population of 80 trees, and sample sizes of n = 6, 13, 19, 26, 32, 38, 45, and 51 were taken from a population of 320 trees. A sample size of six is a priori considered as the smallest sample size that in 19 of 20 cases provides an estimate of TY within 25% of the actual value. The maximum sample sizes are regarded as the limit for achieving a sampling error around 5%. Sample sizes less than six were deemed of low practical value. Estimation of reliable sampling errors for small sizes appears to require special approaches (Magnussen 2001) beyond the scope of this study.

For each of the 832 distinct data settings, a PPP sample was obtained via the protocol described above. Estimates of totals and sampling errors were then obtained with estimators listed in Tables 1 and 2. This process was repeated 27 000 times for a total of 22.464×10^6 samples. With 27 000 replicates of each sample, the mean estimate of a total obtained would have a relative standard error of less than approximately 0.05%. Means of sampling errors obtained with a specific estimator would have relative standard errors of approximately 0.1%.

Estimators of total volume (TY)

Four PPP estimators of TY, two ratio estimators and two calibration estimators, were chosen to demonstrate the need for an extensive evaluation (Brewer et al. 2000; Gregoire and Valentine 1999; Magnussen 2000). Recall that it is possible for some estimators not included in this study to outperform one or all of the included estimators. Table 1 lists details of the estimators. The first is an adjusted ratio estimator by

Table 1. Estimators of total volume (TY).

Estimator	Estimation	Reference
Adjusted ratio	$\hat{\mathrm{TY}}_{\mathrm{GR}} = \left(\frac{E[n_s]}{n_s}\right)\left(\sum_{i\in s}\frac{Y_i}{\tilde{p}_i}\right)$	Grosenbaugh 1965
Ratio	$\hat{\mathrm{TY}}_{\mathrm{R}} = \hat{R}\times \mathrm{TX}$ $\hat{R} = \left(\sum_{i\in s}\frac{Y_i}{\tilde{p}_i}\right)\left(\sum_{i\in s}\frac{X_i}{\tilde{p}_i}\right)^{-1}$	Särndal et al. 1992
Calibration	$\hat{\mathrm{TY}}_{\mathrm{C}} = \sum_{i\in s}\hat{w}_i Y_i$ $N = \sum_{i\in s}\hat{w}_i \wedge \mathrm{TX} = \sum_{i\in s}\hat{w}_i X_i$	Särndal 1996
Calibration 2	$\hat{\mathrm{TY}}_{\mathrm{C2}} = \hat{\alpha}_{\mathrm{C2}} + \hat{\beta}_{\mathrm{C2}}\left(\mathrm{TX} - \sum_{i\in s}\frac{X_i}{\breve{p}_i}\right)$	Brewer 1999

Note: Contributions from trees with an inclusion probability of 1 are not included. See text for definitions of symbols.

Table 2. Sample-based estimators of sampling variance.

Estimator	$\hat{\mathrm{TY}}$	Reference
(1) $\hat{\sigma}^2_{\mathrm{GR}} = \left(\frac{1-P_0}{n_s \max(1, n_s - 1)}\right)\left(\frac{n_s}{E[n_s]}\right)\sum_{i\in s}\left(\frac{Y_i \mathrm{TX}}{X_i} - \hat{\mathrm{TY}}_{\mathrm{GR}}\right)^2\left(1 - \frac{E[n_s]}{N}\right)$	$\hat{\mathrm{TY}}_{\mathrm{GR}}$	Grosenbaugh 1976
(2) $\hat{\sigma}^2_{\mathrm{R}} = \left(\frac{\mathrm{TX}}{\hat{\mathrm{TX}}}\right)^2\sum_{i\in s}\frac{1-\tilde{p}_i}{\tilde{p}_i^2}(Y_i - \hat{\mathrm{R}}X_i)^2$	$\hat{\mathrm{TY}}_{\mathrm{R}}$	Särndal et al. 1992
(3) $\hat{\sigma}^2_{\mathrm{C}} = \sum_{i\in s}\left(1 - \frac{1}{\tilde{p}_i}\right)(\hat{\varepsilon}_i \hat{w}_i)^2$	$\hat{\mathrm{TY}}_{\mathrm{C}}$	Särndal 1996
(4) $\hat{\sigma}^2_{\mathrm{C2}} = \sum_{i\in s}\frac{(c^{-1} - \breve{p}_i)\left(Y_i - \frac{\hat{\alpha}_{\mathrm{C2}}}{Nn_s} - \hat{\beta}_{\mathrm{C2}}X_i\right)^2}{\breve{p}_i^2},\ c = \frac{n_s - 1}{\left(n_s - \frac{1}{n_s}\sum_{i\in s}\breve{p}_i^2\right)}$	$\hat{\mathrm{TY}}_{\mathrm{C2}}$	Brewer 1999
(5) $\hat{\sigma}^2_{\hat{\mathrm{TY}}_{\mathrm{R}}\mid\mathrm{TX}} = \hat{\sigma}^2_{\hat{\mathrm{TY}}_{\mathrm{R}}}(1 - \hat{\rho}^2_{XY})\left(1 + \frac{3\sigma^2_{n_s}}{n_s^2} - \frac{4\sigma^3_{n_s}}{n_s^3} + \frac{5\sigma^4_{n_s}}{n_s^4}\right)\left(1 - \frac{n_s}{N}\right)$	$\hat{\mathrm{TY}}_{\mathrm{R}}$	

Note: The $\hat{\mathrm{TY}}$ column gives the corresponding estimator of totals as per Table 1.

Grosenbaugh (1965), known in forestry as the 3P7 estimator. The second is a ratio estimator ($\hat{\mathrm{TY}}_{\mathrm{R}}$) developed by Särndal et al. (1992). The third choice is a calibration estimator ($\hat{\mathrm{TY}}_{\mathrm{C}}$) that through a reweighting of the inclusion probabilities ensures that the sample estimate of the total of the predictors:

$$\hat{\mathrm{TX}} = \sum_{i\in s}\frac{X_i}{\tilde{p}_i} = \frac{n_s}{E[n_s]}\mathrm{TX}$$

and the sample estimate of population size:

$$\hat{N} = \sum_{i\in s}\frac{1}{\tilde{p}_i}$$

match the known values of TX and N, respectively. Table 1 details $\hat{\mathrm{TY}}_{\mathrm{C}}$. Calibration weights w_i that minimize the difference between the original inclusion probabilities and the new weights were obtained from Särndal (1996):

$$[2]\qquad \hat{w}_i = \frac{1}{\tilde{p}_i}\left(\frac{1+\lambda X_i}{c_i}\right),\qquad i\in s$$

with

$$[3]\qquad c_i = \frac{\tilde{p}_i}{1-\tilde{p}_i}$$

$$\lambda = (\mathrm{TX} - \hat{\mathrm{TX}})\mathrm{TS}^{-1}$$

and

$$\mathrm{TS} = \sum_{i\in s} \frac{X_i^2(1-\tilde{p}_i)}{\tilde{p}_i^2}$$

The fourth choice, $\hat{\mathrm{TY}}_{\mathrm{C2}}$, is also a calibration estimator derived from first-order inclusion probabilities only and is said to be interpretable in terms of design-based and model-based inference simultaneously (Brewer 1999). It was coined "cosmetically calibrated" to reflect that the estimator retains its "natural" variance without attempts to manipulate joint inclusion probabilities (the probability that both tree i and j are in the sample). It introduces scaled (generalized) versions of the inclusion probabilities:

$$\breve{p}_i = \frac{n_s \tilde{p}_i}{E[n_s]}$$

to dampen fluctuations of estimated totals caused by samples larger and smaller than n. Estimation (see Table 1) requires an estimation of

$$[4] \qquad \hat{\alpha}_{\mathrm{C2}} = \sum_{i\in s} \frac{Y_i}{\tilde{p}_i}, \qquad \hat{\beta}_{\mathrm{C2}} = \frac{\sum_{i\in s}\left(\frac{1}{\tilde{p}_i}-1\right)Y_i}{\sum_{i\in s}\left(\frac{1}{\tilde{p}_i}-1\right)X_i}$$

Compared with the first calibration estimator, $\hat{\mathrm{TY}}_{\mathrm{C2}}$ uses an additive adjustment rather than a multiplicative one.

Sample-based estimators of sampling error

Five alternative sample-based estimators of sampling variance are listed in Table 2; the last is a new development. Various studies indicate that these estimators exhibit desirable features (relative to other estimators) such as closeness to the actual sampling variance in simulated sampling, low root mean square errors, and low variability in replicated sampling (Brewer et al. 2000; Gregoire and Valentine 1999; Magnussen 2000). The first estimator, named *vS* by Grosenbaugh (1976), is an improvement over a previous adjusted variance estimator named 3P9$_a$. The ratio variance estimator $\hat{\sigma}^2_{\hat{R}}$ in Table 2 uses the sample based prediction $\hat{\mathrm{TX}}$ of TX to adjust the squared sum of the probability weighted residuals between the actual volume and the volume predicted by the ratio estimator $\hat{\mathrm{R}}\mathrm{X}_i$. The calibration estimator $\hat{\sigma}^2_{\mathrm{C}}$ by Särndal (1996) is derived from a first order Taylor series of probability expanded and reweighted ($\hat{w}_i$, see Table 1) residuals $\hat{\varepsilon}_i$, where $\hat{\varepsilon}_i = Y_i - \hat{\beta}_c X_i$ and

$$[5] \qquad \hat{\beta}_{\mathrm{c}} = \left(\sum_{i\in s}\frac{X_i^2}{\tilde{p}_i c_i}\right)^{-1}\left(\sum_{i\in s}\frac{X_i Y_i}{\tilde{p}_i c_i}\right)$$

where c_i is defined in [3]. The fourth variance estimator (Table 2) was developed by Brewer (1999) and is, as the first estimator, conditional on the actually achieved sample size (n_s). After limited testing it was accredited with desirable properties (Brewer et al. 2000).

The last estimator, $\hat{\sigma}^2_{\hat{\mathrm{TY}}_{\mathrm{R}}|\mathrm{TX}}$, is proposed as an improvement to $\hat{\sigma}^2_{\hat{R}}$. It exploits the known sampling distribution of sample sizes $\pi(n_s)$. The first step computes the variance of sample estimates of the total of the known predictor $\hat{\mathrm{TX}}$, which, of course, turns out to be a simple function of n_s. Second, a scaling to $\hat{\mathrm{TY}}_{\mathrm{R}}$ via the estimated volume ratio $\hat{R}$ of actual to predicted volume leads to an estimate of the total unconditional variance of the estimated totals. Because X is known, only the conditional variance of $\hat{\mathrm{TY}}$ given $\hat{\mathrm{TX}}$ contributes to the sampling variance of totals. By conditioning on $\hat{\mathrm{TX}}$ a final estimate of the sampling variance is obtained. Hence the variance of $\hat{\mathrm{TY}}_{\mathrm{R}}$ is

$$[6] \qquad \hat{\sigma}^2_{\hat{\mathrm{TY}}_{\mathrm{R}}} = \hat{R}^2\sigma^2_{\hat{\mathrm{TX}}} + E[\hat{\mathrm{TX}}]^2\hat{\sigma}^2_{\hat{R}}$$

where expectation is over all possible sample sizes since $\hat{\mathrm{TX}} = (n_s/n)\mathrm{TX}$. The probability distribution of sample sizes $\pi(n_s)$, $1 \le \pi(n_s) \le N$ was recovered by standard techniques (Johnson et al. 1992) from the probability generating function (pgf) of sample sizes

$$[7] \qquad \mathrm{pgf}_{n_s}(z) = \prod_{i=1}^{N}(1-\tilde{p}_i+\tilde{p}_i z)$$

Magnussen (2001) provides details. $\sigma^2_{\hat{\mathrm{TX}}}$ was obtained from

$$[8] \qquad E[(\hat{\mathrm{TX}}-\mathrm{TX})^2] = \mathrm{TX}^2 E\left[\left(\frac{n_s}{n}-1\right)^2\right]$$

To compute the desired conditional variance $\hat{\sigma}^2_{\hat{\mathrm{TY}}_{\mathrm{R}}|\mathrm{TX}}$ an estimate of the variance in Y_i explained by X_i is needed. The square of the correlation coefficient ρ_{XY} between X_i and Y_i estimates this quantity $\hat{\rho}_{XY} = \hat{R}\sigma_{\hat{\mathrm{TX}}}\hat{\sigma}^{-1}_{\hat{\mathrm{TY}}}$ (Draper and Smith 1981). From this follows the estimator given in the last row of Table 2. Two additional multipliers were added, the first:

$$[9] \qquad \left(1+\frac{3\sigma^2_{n_s}}{n_s^2}-\frac{4\sigma^3_{n_s}}{n_s^3}+\frac{5\sigma^4_{n_s}}{n_s^4}\right)$$

provides a five-term Taylor series approximation to the expected value of $\hat{R}^2$, and the second:

$$\left(1-\frac{n_s}{N}\right)$$

provides a correction for the sampled fraction of a finite population (Thompson 1992).

All sampling error estimates in excess of twice the estimated total were considered as extreme. Extreme sampling errors were censored to a value of $\hat{\mathrm{TY}}\sqrt{2}$. The interval $\hat{\mathrm{TY}} \pm \hat{\mathrm{TY}}$ would then have a probability of containing the true value of at least 88% if $\hat{\mathrm{TY}}$ has a normal distribution

(Wall et al. 2001). Censoring was a rare event; it was only applied at a rate of about 1:4500. Censored cases were predominantly associated with the first and last variance estimator, with the calibrated estimator a distant third. However, the rates were considered statistically equal ($\hat{P}$ = 0.17).

Performance criteria

Unbiased, low mean departure from the true total and low variability are the desired attributes for the estimator of total volume (Shiver and Borders 1996). Accordingly, the following three performance indicators were chosen

$$\text{(C1)} \quad \hat{E}[\hat{\text{TY}} - \text{TY}] \times \frac{100\%}{\text{TY}}$$

expected percent bias

$$\text{(C2)} \quad \hat{E}\left[\left|\hat{\text{TY}} - \text{TY}\right|\right] \times \frac{100\%}{\text{TY}}$$

expected percent absolute bias

$$\text{(C3)} \quad \hat{E}[(\hat{\text{TY}} - \text{TY})^2]^{0.5} \times \frac{100\%}{\text{TY}}$$

expected percent root mean square error

where approximate expectations ($\hat{E}$) are the mean of the 27 000 repeated samples, $\hat{\text{TY}}$ is an estimator of total volume, and TY is the actual total.

Attractive sample-based estimates of sampling error of an estimated total are relatively small and positive; they match, in expectation, the root mean square error of the estimated total; and they achieve significance levels of estimated confidence intervals that are close to the stated confidence level. These desirable properties prompted two additional criteria (C4 and C5). C4 is the ratio of sample-based root mean square error to the root mean square error of the estimated totals. Specifically

$$\text{(C4)} \quad \left(\frac{\hat{E}[\hat{\sigma}^2_{\hat{\text{TY}}}] + \hat{E}[(\hat{\text{TY}} - \text{TY})]^2}{\hat{E}[(\hat{\text{TY}} - \text{TY})^2]} \right)^{0.5}$$

where $\hat{E}$ denotes the expectation (mean) over the 27 000 repeated samples. For estimated sampling variances ($\hat{\sigma}^2_{\hat{\text{TY}}}$) close to their actual values, the ratio should be close to 1 when $\hat{\text{TY}} \approx \text{TY}$. A ratio less than 1 generates an optimistic (liberal) estimate of the sampling variance and vice versa (conservative).

C5 measures departures in the achieved significance levels ($\hat{\alpha}$) of a sample-based confidence interval with a nominal chance of $1 - \alpha$ of containing the true mean. The achieved significance level was computed from the indicator variable δ_α, where

$$[10] \quad \hat{\delta}_\alpha = \begin{cases} 1 \text{ if } \text{TY} \in [\hat{\text{TY}} - z_{\alpha/2}\hat{\sigma}_{\hat{\text{TY}}}, \hat{\text{TY}} + z_{1-\alpha/2}\hat{\sigma}_{\hat{\text{TY}}}] \\ 0, \text{ otherwise} \end{cases}$$

where z_α is the $100(1 - \alpha)$% quantile of a standard normal distribution. For each of the five estimators and 832 data settings, the mean $\bar{\delta}_\alpha$ over the 27 000 replicated samples was compared with the nominal levels of $1 - \alpha$ = 0.80, 0.90, 0.95, and 0.99. C5 was then computed as

$$\text{(C5)} \quad \left|1 - \alpha - \bar{\delta}_\alpha\right|$$

Deviations in achieved significance level of ±0.17% would be declared significant at the 95% level of significance under the null hypothesis of nominal significance level (χ^2 test, Miller 1980)

Estimators of totals were ranked against C1–C3, and estimators of sample errors (and bias), against C4–C5. Recommendations for practical use were based on an overall performance index (PI) calculated from the five criteria as outlined below.

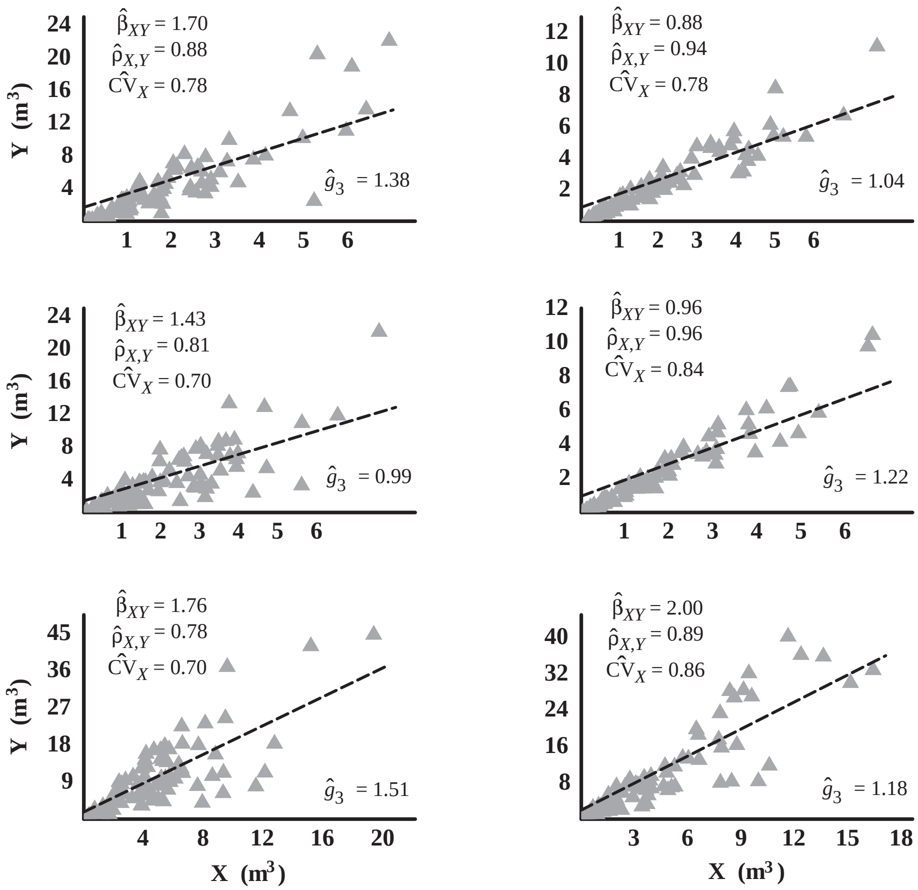
$$[11] \quad \hat{\text{PI}} = 12 - \text{Rank(C1)} - \text{Rank(C2)} - \text{Rank(C3)} - 10 \times |1 - \text{C4}| - \lceil 20 \times |0.95 - \bar{\delta}_{0.05}| \rceil - \lceil 10 \times |0.80 - \bar{\delta}_{0.20}| \rceil$$

where $|x|$ is the absolute value of x, and $\lceil x \rceil$ is the smallest integer larger than x. Consequently, an estimator with a ranking of 1 for C1–C3, a ratio of 1 between expected and observed root mean square error, and a perfect probability coverage of stated confidence intervals would have $\hat{\text{PI}}$ = 9. Conversely, a really poor estimator could end up with a performance index of –16. The overall ranking of methods based on the performance index was rather insensitive to the exact weighting of criteria C4 and C5.

Sample data

A wide spectrum of data structures were generated to allow for a comprehensive testing of the PPP estimators. Paired values of the predictor X and the actual volume Y of an individual tree were generated for populations of sizes of 80 and 320. Populations are trees sampled from a forest stand with either fixed-area plots or variable-radius plots (deVries 1986; Schreuder et al. 1993). It is assumed that 80 (and 320) trees provide a good representation of X. To mimic a series of realistic distributions for X and Y and stochastic linear relationships between X and Y with variances of Y given X increasing in X (Gregoire and Dyer 1989; Williams 1997), 64 distinct bivariate distribution were generated for each combination of n and N. A step-by-step protocol for the data generation is listed below. It can be skipped with impunity, since there is no transparent relationship between the bivariate distribution of X and Y and the parameters of the protocol. The protocol only documents the actual data generation.

Fig. 1. Six scatterplots of simulated predicted volume (X) and actual volume (Y). $\hat{\beta}_{XY}$ is the slope of ordinary least squares linear regression of Y on X; $\hat{\rho}_{X,Y}$ is the Pearson moment correlation coefficient between X and Y; $\hat{cv}_X$ is the coefficient of variation of X; and $\hat{g}_3$ is the coefficient of skewness of X.



The data generation protocol produced coefficient of variation of X that would run from 0.48 to 1.28 (mean = 0.69). Skewness coefficient of X attained values from 0.4 to 3.5, with a mean of 1.3. Slopes and intercepts of the linear relationship between X and Y would run from 0.77 to 1.23 and −0.087 to 0.69, respectively. Pearson's product moment correlation coefficients between X and Y were always positive with a low of 0.63 and a high of 0.98 (mean = 0.88). Figure 1 provides a scatterplot of Y and X for six settings.

Protocol for data generation

(1) Generate X by random draw from a gamma distribution with parameters α and β with $(\alpha, \beta) \in \{1.5, 3.0\}$. Hence, a mean of X is $\alpha\beta$ and a variance of X is $(\alpha\beta)^2$ (Johnson et al. 1994).

(2) Generate constants xi = Min(X) and xx = Max(X).

(3) Determine the population regression line $Y = a + bX$ by solving $a + b \times xi = \eta \times xi \wedge a + b \times xx = \lambda \times xx$ for a and b for fixed values of η and λ. Values for η and λ were {0.7, 1.3} and {0.8, 1.2}, respectively.

(4) Generate stochastic realizations of $Y > 0$ given the mean population regression line as $Y = (a + bX)(r(B - A) + A)$, where r is a random draw from a beta distribution $B(2, 2)$, and A and B were assigned values of {0.3, 0.6} and {1.5, 3.4}, respectively.

Results

Criterion C1

Relative bias of all four estimators declined at about equal rates ($\hat{P}$ = 0.13) with increasing sample size but could otherwise not be shown to depend on any attribute of X or Y or the joint distribution of X and Y. The second calibration estimator ($\hat{TY}_{C2}$) produced in most (410 of 832) data settings the lowest relative bias of the four estimators, yet the adjusted ratio estimator ($\hat{TY}_{GR}$) was a close second and occasionally the best (in 227 of 832 cases). A sample of the results is presented in Table 3. For fixed N and increasing n, the second calibration estimator lost terrain to either the adjusted ratio or the ratio estimator. Generally, $\hat{TY}_{C2}$ was either the best or the third best (in 302 cases) estimator in terms of bias. Attempts to predict the ranking of $\hat{TY}_{C2}$ from design attributes (n and N), attributes of X, Y (coefficient of variation and skewness), and the relationship between X and Y (correlation, regression coefficient, and proportionality factor between regression residuals and X) by logistic regression

Table 3. Relative bias (%) of total volume estimates.

		N = 80		N = 320	
Estimator*		n = 6	n = 13	n = 6	n = 51
$\widehat{TY}_{GR}$	$\bar{x}$	0.010	–0.005	–0.001	0.002
	$s(x)$	0.100	0.084	0.089	0.025
	$\bar{r}$	1.6	2.0	1.6	2.1
	$s(r)$	0.6	0.8	0.6	0.7
$\widehat{TY}_{R}$	$\bar{x}$	–0.160	–0.006	–0.229	0.002
	$s(x)$	0.102	0.084	0.088	0.025
	$\bar{r}$	2.6	2.0	2.8	–0.190
	$s(r)$	0.8	0.7	0.5	0.471
$\widehat{TY}_{C}$	$\bar{x}$	–5.782	–2.636	–6.273	0.017
	$s(x)$	0.152	0.141	0.089	0.19
	$\bar{r}$	4.0	4.0	4.0	3.8
	$s(r)$	0.0	0.0	0.0	0.7
$\widehat{TY}_{C2}$	$\bar{x}$	0.010	–0.006	–0.001	0.000
	$s(x)$	0.103	0.075	0.089	0.003
	$\bar{r}$	1.8	2.0	1.6	1.8
	$s(r)$	0.6	1.0	0.5	1.0

Note: $\bar{x}$, mean across 64 data settings; $s(x)$, standard deviation of x (relative bias); $\bar{r}$, mean rank of relative bias; $s(r)$, standard deviation of r (rank).

*See Table 1 for details on estimators.

with multiple responses (McCullagh and Nelder 1989) only identified the slope as a statistically significant ($\hat{P}$ = 0.04) predictor. On average, the rank of $\widehat{TY}_{C2}$ would increase (worsen) by 0.37 ± 0.18 (mean ± SE) for every unit increase in the regression slope between X and Y. However, 98% of the regression slopes were in the interval from 0.6 to 2.7, a range barely wide enough to trigger a rank change. Ranks of $\widehat{TY}_{GR}$ were more stable, the estimator came in second 404 times and with an equal split between first and third for the remainder. The influence of the regression slope between X and Y on the ranking of $\widehat{TY}_{GR}$ was an improvement (decrease) of 0.40 ± 0.18 for every unit increase in the regression slope. None of the differences between $\widehat{TY}_{GR}$ and $\widehat{TY}_{C2}$ were statistically significant (t test, all P values > 0.21). Relative bias of $\widehat{TY}_{GR}$ and $\widehat{TY}_{C2}$ were about 0.15% lower than the bias of the ratio estimator when $n \le 12$ ($\hat{P}$ = 0.05 or less) but within 0.001% for $n > 12$ ($\hat{P} > 0.30$). The rank of the ratio estimator $\widehat{TY}_{R}$ was somewhat sensitive to sample and population sizes. Ranks increased by an average 0.5 when N increased from 80 to 320 but improved (decreased) by 0.01 for every unit increase in n. The calibration estimator $\widehat{TY}_{C}$ trailed the other three by a wide margin (factor 10 or more, $\hat{P} < 0.001$). It was only best in four settings (N = 320, n = 50, and $\hat{R} > 1.5$).

Criterion C2

Across all data settings the mean relative absolute bias (±SD) was 4.5 ± 2.1% for $\widehat{TY}_{C2}$, 6.0 ± 2.8% for $\widehat{TY}_{GR}$, 7.8 ± 4.3% for $\widehat{TY}_{C}$, and finally 20.1 ± 7.3% for $\widehat{TY}_{R}$. Note the parallel increase in standard deviations as the expected absolute bias goes up. Absolute bias was strongly dependent on design factors, the coefficient of variation of X, and the strength of the correlation between X and Y, but the relationship depended on the estimator (Table 4). After adjusting for the linear effects of significant covariates, a test of equal mean relative bias led to the rejection of this null hypothesis (analysis of covariance $\hat{F}_{[3,3308]}$ = 205.4, $\hat{P}$ = 0.001).

Table 4. Estimated effect ($\hat{\beta}$) of estimator and estimator interactions with design variables and attributes of X on expected absolute bias (in percent of true value).

Effect	$\hat{\beta}$	SE($\hat{\beta}$)
δ_{GR}	19.05	0.63
δ_{R}	6.4	0.63
δ_{C}	12.81	0.63
δ_{C2}	14.03	0.63
$\rho(X, Y)$	–7.55	0.45
$\delta_{GR}\sqrt{\left(1-\frac{n}{N}\right)/n}$	16.26	0.44
$\delta_{R}\sqrt{\left(1-\frac{n}{N}\right)/n}$	82.36	0.44
$\delta_{C}\sqrt{\left(1-\frac{n}{N}\right)/n}$	46.30	0.44
$\delta_{C2}\sqrt{\left(1-\frac{n}{N}\right)/n}$	20.46	0.44
$\delta_{GR}\hat{\rho}_{X,Y}$	–26.04	0.79
$\delta_{R}\hat{\rho}_{X,Y}$	–7.47	0.79
$\delta_{C}\hat{\rho}_{X,Y}$	–21.06	0.79
$\delta_{C2}\hat{\rho}_{X,Y}$	–19.01	0.79
$\delta_{GR}\widehat{cv}_{X}$	4.61	0.50
$\delta_{R}\widehat{cv}_{X}$	–0.26ns	0.50
$\delta_{C}\widehat{cv}_{X}$	3.05	0.50
$\delta_{C2}\widehat{cv}_{X}$	3.33	0.50

Note: Model R^2 = 0.993. Root mean square error of predictions = 0.62%. δ_M is the indicator variable for estimator M. δ_M = 1 if estimator is M and 0 otherwise. Estimators are listed in Table 1. $\hat{\rho}(X, Y)$ is the correlation coefficient between X and Y. $\widehat{cv}_X$ is coefficient of variation of X. All effects but one (ns) are significant at the 99.9% level.

In all 832 data settings, the $\widehat{TY}_{C2}$ estimator outperformed the others. Second place was in all but four cases (N = 320, n = 50) held by $\widehat{TY}_{GR}$. $\widehat{TY}_{C}$ consistently ranked third, except in the said four cases. $\widehat{TY}_{R}$ always produced the highest expected absolute bias. $\widehat{TY}_{GR}$ was significantly better than the third placed estimator in all 13 combinations of N and n (t > 27, $\hat{P}$ < 0.001).

Criterion C3

Relative root mean square errors of $\widehat{TY}_{GR}$, $\widehat{TY}_{R}$, and $\widehat{TY}_{C2}$ were virtually identical (mean 7.58%) with differences of less than 0.02%, which are of no practical significance. An analysis of variance confirmed the null hypothesis of no difference ($\hat{P}$ > 0.9) in each of the 13 tested combinations of N and n. In all 13 cases, $\widehat{TY}_{C}$ produced significantly ($\hat{P}$ <

Table 5. Relative root mean square errors (%) of total volume estimates.

Estimator*		N = 80		N = 320	
		n = 6	n = 13	n = 6	n = 51
$\hat{TY}_{GR}$	$\bar{x}$	12.6	7.4	13.3	3.6
	$s(x)$	3.0	1.9	3.1	0.8
	$\bar{r}$	2.4	2.3	2.5	1.7
	$s(r)$	0.5	0.4	0.5	0.6
$\hat{TY}_{R}$	$\bar{x}$	12.6	7.4	13.2	3.6
	$s(x)$	3.0	1.9	3.1	0.8
	$\bar{r}$	1.0	1.3	1.0	1.5
	$s(r)$	0.2	0.4	0.0	0.6
$\hat{TY}_{C}$	$\bar{x}$	21.5	10.4	22.8	3.8
	$s(x)$	1.4	1.3	1.3	0.8
	$\bar{r}$	4.0	4.0	4.0	4.0
	$s(r)$	0.0	0.0	0.0	0.0
$\hat{TY}_{C2}$	$\bar{x}$	12.6	7.4	13.3	3.7
	$s(x)$	3.0	1.9	3.1	0.3
	$\bar{r}$	2.6	2.5	2.5	2.8
	$s(r)$	0.6	0.9	0.5	0.6

Note: $\bar{x}$, mean across 64 data settings; $s(x)$, standard deviation of x; $\bar{r}$, mean rank of relative bias; $s(r)$, standard deviation of r.
*See Table 1 for details on estimators.

0.01) higher relative root mean square error than any other tested estimator. The performance of $\hat{TY}_C$ did, however, improve rapidly with increasing N and n (Table 5). The influence of the coefficient of variation and skewness of X and Y, the correlation and regression slope between X and Y, and the proportionality factor between regression residuals and X was, as expected, similar to their influence on the mean relative bias (C2).

Root mean square errors of $\hat{TY}_{GR}$, $\hat{TY}_C$, and $\hat{TY}_{C2}$ were, as expected, strongly correlated ($\hat{\rho} > 0.990$) with the expected absolute relative bias. For the fourth estimator, $\hat{TY}_R$, the correlation was only moderately strong (0.87). The expected absolute relative bias was about 0.7 times the relative root mean square error in the former group of three estimators versus 1.8 in $\hat{TY}_R$. In absence of bias the absolute mean deviation of a normally distributed variable is $(2/\pi)^{0.5} \approx 0.80$ times its standard deviation.

Criterion C4

Relative root mean square errors derived from the sample-based estimates of variance plus squared bias were, in four of five estimators, strongly correlated ($\hat{\rho} > 0.990$) with the root mean square errors obtained via resampling (Fig. 2). Variance estimates obtained with the adjusted ratio estimator were more erratic, especially as relative root mean square errors increased.

The mean ratio of expected to observed relative root mean square error varied from a low of 1.22 ± 0.10 for the calibration estimator to a high of 1.50 ± 0.10 for the conditional variance estimator $\hat{\sigma}^2_{TY|TX}$. Each estimator produced a mean ratio that was significantly different from the others ($\hat{P} < 0.001$). Root mean square error ratios obtained with the calibration estimator were more consistent across levels of relative root mean square errors without a tendency to increase for higher relative root mean square errors as evident in the others. Design parameters (N and n), the strength of correlation between X and Y, and the relative variation of X all had a significant impact on the ratio (robust regression coefficients (Staudte and Sheather 1990) were all significantly different from zero, $\hat{P} < 0.01$), albeit with distinct differences among estimators in sensitivity and response to a change in data settings.

Relative root mean square error of the conditional variance estimator was 1.15 (±0.01) times the relative root mean square error of the cosmetically calibrated estimator, and the correlation between them was almost perfect (0.996) and higher than between any other pair of estimators.

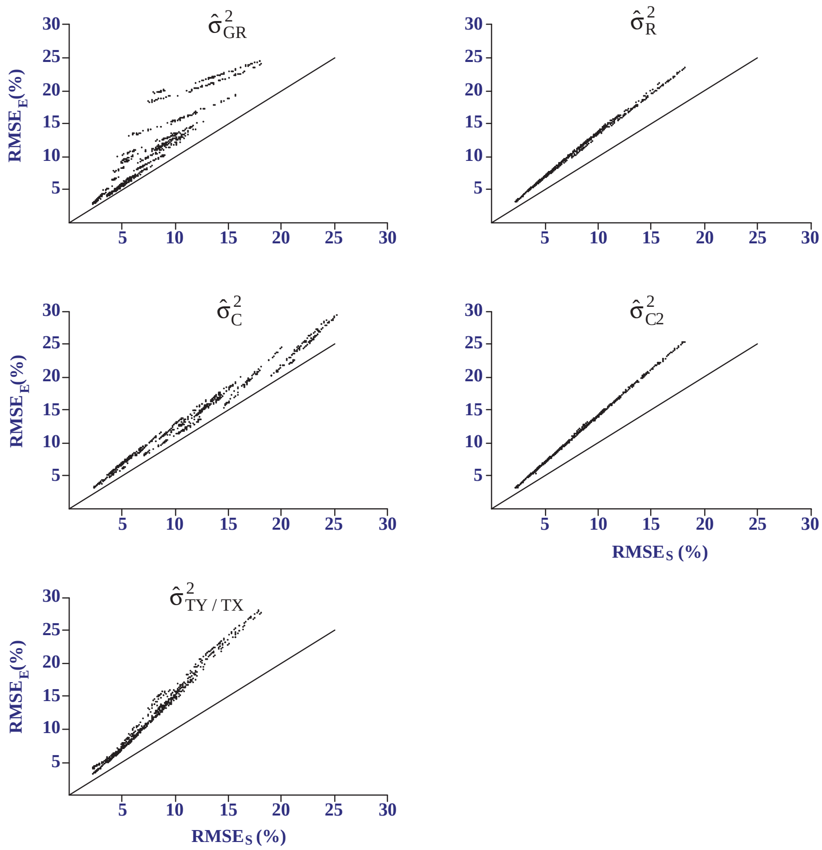
Criterion C5

Two variance estimators $\hat{\sigma}^2_{C2}$ and $\hat{\sigma}^2_{TY|TX}$ achieved nearly identical (difference less than 0.02) significance levels, which were also significantly closer to their nominal levels than were the other three estimators. At a nominal significance level of 0.80, their estimated confidence interval captured the true total with a mean probability of 0.83. At a nominal significance level of 0.95 they were within 1% of the target value. Figure 3 illustrates the results. Note that the best results were also associated with the smallest variation across the various data settings. Relative results from the 0.90 and 0.99 levels of significance were quite similar. Poor results (under-coverage) were obtained with the calibration estimator $\hat{\sigma}^2_C$ with deviations of –0.04 for the 0.99 level of significance and –0.08 for the 0.80 level of significance. The adjusted ratio estimator $\hat{\sigma}^2_{GR}$ on the other hand was overly conservative with achieved significance levels well above the nominal targets. Results with the ratio estimator $\hat{\sigma}^2_R$ were intermediate (about 0.02 below nominal level for $\alpha = \{0.2, 0.10, 0.05\}$). Design parameters ($N$ and n), correlation between X and Y, and the relative variance of X all exerted a significant impact on the achieved significance level. Coefficients in a logistic regression were significantly different from 0 (Collett 1991). Again, these effects varied significantly from one estimator to another.

Overall performance index (PI)

Two estimators, the cosmetically calibrated and the adjusted ratio estimator, achieved the best overall performance index of 2.7 ± 1.3 and 2.5 ± 3.2 (mean ± SE), respectively. Third place was earned by the ratio estimator (0.6 ± 1.0), and the last place was held by the calibration estimator (–1.8 ± 0.7). Of 832 cases, the cosmetically calibrated estimator was best 444 times (53%), but it clearly lost ground to

Fig. 2. Relative root mean square errors computed from sample-based estimates of sampling variance plus squared bias ($RMSE_E$ (%)) plotted against the root mean square error obtained from 27 000 repeat samples ($RMSE_S$ (%)). The variance estimator used in $RMSE_E$ is displayed in each scatterplot (see Table 2 for definitions). A 1:1 line is added for comparison. Each scatterplot contains 832 paired observations.

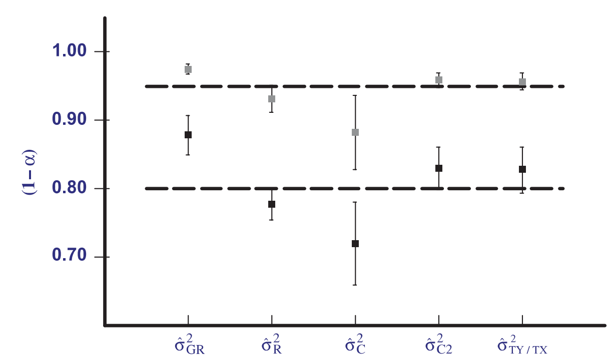


the adjusted ratio estimator as sample sizes increased for a given N. For example, with N = 320 the chance of a better result with the adjusted ratio estimator increased from 1:6 to 7:8 as n went from 6 to 51. Even the ratio estimator had a 2:8 chance of outperforming the cosmetically calibrated estimator. For the ratio estimator the odds of a first place improved as sample size increased. For the conditional estimator, the odds of a first place were about 1:8 in all data settings. Only the calibration estimator had virtually no chance of being the best within the tested settings.

Discussion

To combat the high variability inherent in PPP estimates of totals and sampling error, a phenomena linked to “too many small” or “too few large” (Brewer 2000; Brewer et al. 2000) observations in the sample, various modifications of the basic Horvitz–Thomson (Hanurav 1963) estimator have been successfully developed (Brewer 1999; Brewer and Hanif 1983; Grosenbaugh 1976; Särndal et al. 1992). Even optimal combinations of various estimators have been tried (Gregoire and Valentine 1999), but Leblanc and Tibshirani

Fig. 3. Achieved significance level of confidence interval of total volume with a nominal chance of containing the true total of 80% (solid squares) and 95% (gray squares). Variance estimators used for calculating confidence intervals are positioned along the horizontal axis of the graph. Vertical lines have a total length of two times the standard deviation of the achieved significance level in 832 data settings.



(1996) found that the conditions for a composite gain in efficiency by combining estimators may be relatively rare in most real life applications.

Highly variable results are anathema to practitioners, since they cannot afford to invalidate suspect results or the time to measure a “surplus” of trees (Bonnor 1972). Systematic and stratified volume sampling will remain attractive alternatives until the PPP variability problem has been completely resolved (Magnussen 2000; Schreuder 1975).

When PPP sampling is the chosen method, a careful choice of estimator(s) must be made because estimators do differ significantly in various aspects of performance. Yet, the choice could be difficult since no single estimator appears to be uniformly best throughout a set of applied criteria. Differences among methods also diminish as sample sizes and population size increase. In a typical forest enterprise, the application settings of a PPP estimator will vary greatly from one case to another. Only a comprehensive comparison of estimators across a wide range of realistic settings will produce a reliable assessment and quantify the odds that one estimator when compared with another will produce superior results. Comparisons restricted to a single data setting have a nonzero chance of inverting the expected performances of two estimators. The sensitivity of estimators to the strength of the correlation and slope between predicted and actual stem volume and to the variation and skewness of the predictors predicates testing across the range of values expected in practical applications. Outliers in either X, Y, or both can drastically change the performance of a PPP estimator (Willliams and Schreuder 1998). Williams and Schreuder (1998) found Grosenbaugh’s adjusted ratio estimator to be especially sensitive to outliers. The data used here were free of outliers; results obtained can therefore not be extended to situations where outliers must be expected.

The choice of estimator will also depend on the value attached to various performance criteria. In terms of absolute bias, Brewer’s cosmetically calibrated estimator was uniformly best, but in terms of root mean square error, the ratio estimator (Särndal et al. 1992) would in most cases be a better choice. Emphasis on relative bias would confirm the former but the odds that another estimator would be better in a single application were nevertheless slightly over 1.0. A desire to have sample-based estimates of sampling error matching the actual sampling error would clearly favor the calibration estimator (Särndal 1996), which has odds of 1.4 of producing the best match. The otherwise poor performance of the calibration estimator may be due to the relatively small population and sample sizes entertained in the evaluation scheme. For populations counted in the thousands and beyond and sample sizes in excess of 50, a calibration estimator should be expected to perform much better (Deville et al. 1993; Gregoire and Valentine 1999). However, it is difficult to imagine a practical forestry application of PPP volume sampling in such settings. Finally emphasis on achieved probability coverage of confidence intervals would consider the cosmetically calibrated and the conditional variance estimator as equally best. Combining the performance against various criteria into an overall index of performance is a pragmatic approach to finalize the choice. For use in a wide variety of settings, the cosmetically calibrated estimator appears to be the overall best. The margin is quite narrow, however, with odds ratio of obtaining the best results in a given data setting of no more than 1.16. Computing the mean of the adjusted and cosmetically calibrated estimator would produce a 5% drop in average performance ($\hat{P}$ = 0.11). A solution that should appeal to practice where buffering against adverse results is appropriate (Pope and Ziemer 1984).

Acknowledgement

Dr. S. Titus (University of Alberta) provided helpful comments on an earlier draft.

References

Avery, T.E., and Burkhart, H.E. 1983. Forest measurements. McGraw-Hill, New York.

Bonnor, G.M. 1972. A test of 3-P sampling in forest inventories. For. Sci. **18**: 198–202.

Brewer, K.R.W. 1999. Cosmetic calibration for unequal probability samples. Surv. Meth. **25**: 205–212.

Brewer, K.R.W. 2000. Deriving and estimating an approximate variance for the Horvitz-Thompson estimator using only first order inclusion probabilities. American Statistical Association, Buffalo, N.Y.

Brewer, K.R.W., and Hanif, M. 1983. Sampling with unequal probability. Lecture notes in statistics. Springer, New York.

Brewer, K.R.W., Gregoire, T.G., and Brewer, K. 2000. Estimators for use with Poisson sampling and related selection procedures. American Statistical Association, Buffalo, N.Y.

Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., and Bailey, R.L. 1983. Timber management: a quantitative approach. John Wiley and Sons, New York.

Cochran, W.G. 1977. Sampling techniques. John Wiley and Sons, New York.

Collett, D. 1991. Modelling binary data. Chapman & Hall, London.

Deville, J.C., Särndal, C.E., and Sautory, O. 1993. Generalized raking procedures in survey sampling. J. Am. Stat. Assoc. **88**: 1013–1020.

deVries, P.G. 1986. Sampling theory for forest inventory. A teach yourself course. Springer, Berlin.

Draper, N.R., and Smith, H. 1981. Applied regression analysis. John Wiley and Sons, New York.

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

Gregoire, T.G., and Dyer, M.E. 1989. Model fitting under patterned heterogeneity of variance. For. Sci. **35**: 105–125.

Gregoire, T.G., and Valentine, H.T. 1999. Composite and calibration estimation following 3P sampling. For. Sci. **45**: 179–185.

Grosenbaugh, L.R. 1965. Three-Pee sampling theory and program “THRP” for computer generation of selection criteria. USDA For. Serv. Res. Pap. PSW-21.

Grosenbaugh, L.R. 1976. Approximate sampling variance of adjusted 3P estimates. For. Sci. **22**: 173–176.

Hanurav, T.V. 1963. On “Horwitz and Thompson estimator”. Indian J. Stat. A, **24**: 429–436.

Hartman, G.B. 1967. Some practical experience with 3-P sampling and the Barr and Stround dendrometer in timber sales. Society of American Foresters. Seattle, Wash.

Johnson, N.L., Kotz, S., and Kemp, A.W. 1992. Univariate discrete distributions. John Wiley and Sons, New York.

Johnson, N.L., Kotz, S., and Balakrishnan, N. 1994. Continuous univariate distributions. Vol. 1. John Wiley and Sons, New York.

Leblanc, M., and Tibshirani, R. 1996. Combining estimates in regression and classification. J. Am. Stat. Assoc. **91**: 1641–1650.

Lynch, T.B. 1995. Use of a tree volume equation based on two lower-stem diameters to estimate forest volume from sample tree counts. Can. J. For. Res. **25**: 871–877.

Magnussen, S. 2000. Unequal probability sampling in fixed area plots of stem volume with and without prior inclusion probabilities. J. Appl. Stat. **27**: 975–990.

Magnussen, S. 2001. Saddlepoint approximations for statistical inference of *PPP* sample estimates. Scand. J. For. Res. **16**: 180–192.

McCullagh, P., and Nelder, F.A. 1989. Generalized linear models. Chapman and Hall, London.

Miller, J.G. 1980. Simultaneous statistical inference. 2nd ed. Springer, New York.

Pope, R.D., and Ziemer, R.F. 1984. Stochastic efficiency, normality, and sampling errors in agricultural risk analysis. Amer. J. Agr. Econ. **66**: 31–40.

Särndal, C.E. 1996. Efficient estimators with simple variance in unequal probability sampling. J. Am. Stat. Assoc. **91**: 1289–1300.

Särndal, C.-E., Swensson, B., and Wretman, J. 1992. Model assisted survey sampling. Springer, New York.

Schreuder, H.T. 1975. Two stratified sampling alternatives to 3-p sampling. *In* Monitoring Forest Environment Through Successive Sampling: Proceedings of the Symposium of the State University of New York College of Environmental Science and Forestry, 1974, Syracuse, N.Y. *Edited by* C. Cunia. College of Environmental Science and Forestry, Syracuse, N.Y. pp. 55–68.

Schreuder, H.T., Sedransk, J., and Ware, K.D. 1968. 3-P sampling and some alternatives, I. For. Sci. **14**: 429–453.

Schreuder, H.T., Sedransk, J., Ware, K.D., and Hamilton, D.A. 1971. 3-P sampling and some alternatives, II. For. Sci. **17**: 103–118.

Schreuder, H.T., Ouyang, Z., and Williams, M. 1992. Point-Poisson, point-pps, and modified point-pps sampling: efficiency and variance estimation. Can. J. For. Res. **22**: 1071–1078.

Schreuder, H.T., Gregoire, T.G., and Wood, G.B. 1993. Sampling methods for multiresource forest inventory. John Wiley and Sons, New York.

Shiver, B.D., and Borders, B.E. 1996. Sampling techniques for forest resource inventory. John Wiley and Sons, New York.

Space, J.C., and Turman, D. 1977. Three-P sampling: an efficient sampling technique for forest inventory. Ber. Eidg. Anstalt. Forstl. Vers. **171**: 15–22.

Staudte, R.G., and Sheather, S.J. 1990. Robust estimation and testing. John Wiley and Sons, New York.

Thompson, S.K. 1992. Sampling. John Wiley and Sons, New York.

Turnblom, E.C., and Burk, T.E. 1996. Adjusting volume table estimates using normal form quotient. Can. J. For. Res. **26**: 155–158.

Van Deusen, P.C. 1987. Combining taper functions and critical height sampling for unbiased stand volume estimation. Can. J. For. Res. **17**: 1416–1420.

Wall, M.M., Boen, J., and Tweedie, R. 2001. An effective confidence interval for the mean with samples of size one and two. Am. Stat. **55**: 102–105.

Williams, M.S. 1997. A regression technique accounting for heteroscedastic and asymmetric errors. J. Agric. Biol. Env. Stat. **2**: 108–129.

Williams, M.S., and Schreuder, H.T. 1998. Outlier-resistant estimators for Poisson sampling: a note. Can. J. For. Res. **28**: 794–797.

Yamamoto, K. 1994. A simple volume estimation system and its application to three coniferous species. Can. J. For. Res. **24**: 1289–1294.

Zakrzewski, W.T., and Termikaelian, M. 1994. New application of a polynomial curve to plot volume estimation. Can. J. For. Res. **24**: 1083–1088.