Eur J Forest Res (2010) 129:847–861
DOI 10.1007/s10342-009-0277-3

ORIGINAL PAPER

Statistical analysis of ratio estimators and their estimators of variances when the auxiliary variate is measured with error

Christian Salas · Timothy G. Gregoire

Received: 14 May 2008 / Revised: 2 December 2008 / Accepted: 23 December 2008 / Published online: 3 June 2009

Abstract Forest inventory relies heavily on sampling strategies. Ratio estimators use information of an auxiliary variable (x) to improve the estimation of a parameter of a target variable (y). We evaluated the effect of measurement error (ME) in the auxiliary variate on the statistical performance of three ratio estimators of the target parameter total τ_y. The analyzed estimators are: the ratio-of-means, mean-of-ratios, and an unbiased ratio estimator. Monte Carlo simulations were conducted over a population of more than 14,000 loblolly pine (*Pinus taeda* L.) trees, using tree volume (v) and diameter at breast height (d) as the target and auxiliary variables, respectively. In each simulation three different sample sizes were randomly selected. Based on the simulations, the effect of different types (systematic and random) and levels (low to high) of MEs in x on the bias, variance, and mean square error of three ratio estimators was assessed. We also assessed the estimators of the variance of the ratio estimators. The ratio-of-means estimator had the smallest root mean square error. The mean-of-ratios estimator was found quite biased (20%). When the MEs are random, neither the accuracy (i.e. bias) of any of the ratio estimators is greatly affected by type and level of ME nor its precision (i.e. variance). Positive systematic MEs decrease the bias but increase the variance of all the ratio estimators. Only the variance estimator of the ratio-of-means estimator is biased, being especially large for the smallest sample size, and larger for negative MEs, mainly if they are systematic.

Keywords Sampling · Forest inventory · Design-based inference · Variance estimators · Bias

Communicated by T. Knoke.

This article belongs to the special issue "Linking Forest Inventory and Optimisation".

C. Salas (✉) · T. G. Gregoire
School of Forestry and Environmental Studies, Yale University, 360 Prospect Street, New Haven, CT 06511-2104, USA
e-mail: christian.salas@yale.edu

C. Salas
Departamento de Ciencias Forestales, Universidad de La Frontera, Temuco, Chile

Introduction

Sampling methods are important for assessing natural resource abundance. Natural populations in ecology (forestry, fisheries, and wildlife) are extremely large; consequently, sampling techniques have to be conducted for characterizing those populations. Sampling allows us, based on a very small portion of the population, to extend the sample results to the population level through the use of statistical inference. There are three key components to be defined for any sampling task: sample design, estimator, and inferential procedure. The sample design elucidates how to draw the sample, while the estimator is the statistic that estimates a parameter of interest of the population, and the inferential procedure determines the reliability of the estimator. The combination of a particular design and estimator defines a sampling strategy in the sense of Gregoire and Valentine (2008). Here we stay within the design-based framework of statistical inference (*sensu* Gregoire 1998; Gregoire and Valentine 2008), where the population of interest is regarded as a fixed—not a random—quantity, and statistical inference is based on the distribution of all estimates possible under the given sampling design.

Following a simple random or systematic sampling, the Horvitz-Thompson (expansion) estimator of the population mean (μ) of total (τ) is common. Nevertheless, more efficient estimators have been developed for the same sampling designs. Among them are those that use an auxiliary variable x that is correlated with the variable of interest y. As pointed out by Robinson et al. (1999), the opportunities to integrate auxiliary information into forest stand inventory are considerable, and the potential benefits are very attractive. An example of the use of auxiliary information in a forestry context is two-phase sampling, where in the first phase auxiliary data are obtained for all sampling units from "large area measurements" techniques such as aerial photographs, remote sensing (e.g., Landsat), and laser scanning (e.g., LiDAR) and in the second phase a portion of them are measured in the field. This has been used in Finnish forest inventory for almost 30 years (Poso et al. 1999), and since 1950's also in most states in The United States (Frayer and Furnival 1999). In most surveys, data are collected on many items beyond the one variable of primary interest; making the most use of the additional information collected is an issue of both practical and theoretical interest (Dryver and Chao 2007). Examples of estimators that use auxiliary information are the Grosenbaugh's (1964) adjusted estimator based on probability proportional to prediction (3P) sampling, regression estimators, and ratio estimators. The latter is particularly interesting because the sampling variability may be quite smaller than other estimators (Cochran 1977; Gregoire and Valentine 2008), providing a more reliable estimate of the population parameter than the comparable estimate based on the simple arithmetic mean (Sukhatme and Sukhatme 1970). In general, ratio estimators are the simplest estimators that incorporate related information (Mickey 1959).

ME is present in most sampling. The quality of an estimator is a function of both sampling and nonsampling errors (Scali et al. 2005). Sampling errors arise due to drawing a probability sample rather than conducting a census (Stage and Wykoff 1998). Non-sampling errors are due to data collection and processing procedures. ME arises when a given measurement differs from the true value of a variable of interest. MEs depend on the measuring instruments and the way in which each particular field technician uses these instruments (Cunia 1965; Gertner 1990). ME is also called the "observational error" or the "response error" (Hansen et al. 1951). Customarily, it is assumed that the data collected on the units in the sample are the actual values of the characteristics observed, and that the estimates of the population values obtained are uniquely subject to errors solely due to sampling (Sukhatme and Sukhatme 1970). MEs are unavoidable, yet increasing the sample size is typically not a viable method for reducing their effects (Canavan and Hann 2004). An easy way to deal with MEs is to pretend they do not exist, or if they do, assume that their effect is negligible (Chandhok 1988). However, MEs might affect the accuracy (i.e., bias) and precision (i.e., variability) of some estimators. The cumulative effect of the various errors on the estimate is not always negligible, since errors from different sources may not cancel out one another (Sukhatme and Sukhatme 1970).

The effect on inference of ME has not been widely studied in a design-based inference framework. The effect of MEs when fitting parametric models, e.g. regression analysis, has been widely studied not only in the statistical literature (e.g. Fuller 1987; Myers 1990; Bay and Stefanski 2000) but also in the forestry literature (e.g. Gertner 1988, 1990; Kangas 1996, 1998; Stage and Wykoff 1998; Kangas and Kangas 1999; Canavan and Hann 2004; Hordo et al. 2008). In sampling, MEs have been mostly studied in a model-based inference setting or using a mathematical model for the errors of measurement or observational errors (e.g., Cochran 1977, p. 37; Sukhatme and Sukhatme 1970, p. 390).

The effect of ME in the auxiliary variate on the performance of the ratio estimator only recently has been studied. Although some studies have assessed the performance of ratio estimators in sampling without MEs (Tin 1965; Ek 1971; Hutchison 1971; Royall and Cumberland 1981), only a recent theoretical study conducted by Gregoire and Salas (2009) has examined the performance of ratio estimators under MEs. They assessed the effects of having systematic and random MEs in the auxiliary variate (x) used in three ratio estimators. They provided mathematical expressions both to determine how the bias of the ratio estimators change due to systematic ME in x and to compute the variance of the ratio estimators with ME in x. In order to assess the effect of random ME of the ratio estimators these authors conducted simulations over a population of 501 *Eucalyptus nitens* leaves. Gregoire and Salas (2009) neither assess the effect of MEs on the estimates of the variance of the estimators nor of using a larger population. For practical purposes the estimates of the variance estimators are crucial for computing confidence intervals of parameters. Furthermore, higher variability in y and/or x might enhance the performance of one ratio estimator over the others. In the present study, our objective is to assess the statistical performance of three ratio estimators under various forms and magnitudes of ME in the auxiliary variate in a design-based inference framework, and of the estimates of the variance of those ratio estimators, using a large tree population.

Materials and methods

Population

Our population data consists of $N = 14{,}387$ loblolly pine (*Pinus taeda* L.) trees collected in southern USA. The data were provided by the U.S. Forest Service. For each tree of this population, the following variables were measured: crown class, diameter at breast height (d), total height (h), and total volume (v), which was computed based on multiple measurement points along the standing stem. Trees from all crown classes are represented, except open-grown trees. The same data set was used by Gregoire and Williams (1992) and Magnussen (2001) in a volume equations and a 3P sampling study, respectively. For these data volume has the highest variability and skewness, followed by basal area, diameter, and height (Table 1).

Volume and diameter have a linear correlation coefficient of $r = 0.91$ (Fig. 1b). In the context of our study we prefer to use d, instead of basal area (g), because it is the variable that is directly measured in the field (i.e., g is only a function of d, and therefore fully depends on it) and a better understanding of the ME effect can be achieved using it instead of g. Although the relationship between v and d is not linear for the entire range of the data, it is linear across most of the range. Therefore, we chose diameter to be the auxiliary variable for the ratio estimators.

Description of estimators

We consider the following estimators of $\tau_y = \sum_{k=1}^{N} y_k$ based on data from a simple random sample without replacement:

$$\text{``ratio-of-means''} \longrightarrow \hat{\tau}_{y1} = \hat{R}\tau_x = \frac{\bar{y}}{\bar{x}}\tau_x, \qquad (1)$$

where $\tau_x = \sum_{k=1}^{N} x_k$. In (1), $\hat{R}$ is an estimator of $R = \tau_y/\tau_x = \mu_y/\mu_x, \tau_y = \sum_{k=1}^{N} y_k$, as in Gregoire and Valentine (2008), and $\bar{y}$ and $\bar{x}$ are the sample means for the y and x variables, respectively. Notice that μ_y and μ_x are the population average of y and x, and are computed as $\mu_y = \tau_y/N$ and $\mu_x = \tau_x/N$, respectively.

Also,

$$\text{``mean-of-ratios''} \longrightarrow \hat{\tau}_{y2} = \bar{r}\tau_x, \qquad (2)$$

where $\bar{r}$ is the average ratio of $r_k = y_k/x_k$ of those units in the sample. The population average ratio is denoted by μ_r, and computed as $\mu_r = \frac{1}{N}\sum_{k=1}^{N} r_k$.

Also,

$$\text{``unbiased ratio estimator''} \longrightarrow$$
$$\hat{\tau}_{y3} = \hat{\tau}_{y2} + \frac{N-1}{N}\frac{n}{n-1}\left(\hat{\tau}_{y\pi} - \bar{r}\,\hat{\tau}_{x\pi}\right), \qquad (3)$$

where $\hat{\tau}_{y\pi}$ and $\hat{\tau}_{x\pi}$ are the Horvitz-Thompson (HT) estimators (Horvitz and Thompson 1952) of τ_y and τ_x, respectively, as follows

$$\hat{\tau}_{y\pi} = N\bar{y}, \quad \hat{\tau}_{x\pi} = N\bar{x}. \qquad (4)$$

The estimator in (1) is the usual ratio-of-means estimator of τ_y. The estimator $\hat{\tau}_{y2}$ in (2) is sometimes called the mean-of-ratios estimator. It is well known that the ratio-of-means and the mean-of-ratios are biased estimators of τ_y. The estimator $\hat{\tau}_{y3}$ in (3) is the unbiased ratio-type estimator introduced by Hartley and Ross (1954) and further developed by Goodman and Hartley (1958).

The usual approximation to the bias of $\hat{\tau}_{y1}$ may be deduced from (6.34) in Cochran (1977) as

$$B\left[\hat{\tau}_{y1} : \tau_y\right] = \left(\frac{1}{n} - \frac{1}{N}\right)\left(\gamma_x^2 - \rho\gamma_x\gamma_y\right)\tau_y, \qquad (5)$$

where γ_x and γ_y are the coefficients of variation (expressed in relative units) of x and y, respectively, and ρ is the correlation coefficient between y and x in the population.

The bias of $\hat{\tau}_{y2}$ is exactly

$$B\left[\hat{\tau}_{y2} : \tau_y\right] = \sum_{k=1}^{N} r_k(\mu_x - x_k). \qquad (6)$$

The bias of $\hat{\tau}_{y3}$ is zero (Hartley and Ross 1954).

The usual approximation of the variance of $\hat{\tau}_{y1}$ under simple random sampling without replacement (i.e.

Table 1 Descriptive parameters of the loblolly pine (*Pinus taeda* L.) trees population for different variables ($N = 14{,}387$)

Parameter	Variable			
	Diameter (d)	Height (h)	Basal area (g)	Volume (v)
Minimum	12.7	4.3	126.7	0.01
Maximum	87.4	43.9	5,996.2	7.79
Mean (μ)	28.2	19.9	736.6	0.62
Variance (σ^2)	140.3	41.6	424,706.0	0.56
Total (τ)	406,330.7	285,916.7	10,598,091.1	8,932.42
Coefficient of variation (γ) (in %)	41.9	32.4	88.5	120.4
Coefficient of skewness	0.9	0.3	2.1	2.5
Kurtosis	0.8	−0.4	6.8	9.7

Diameter in cm, height in m, basal area in cm^2, and volume in m^3

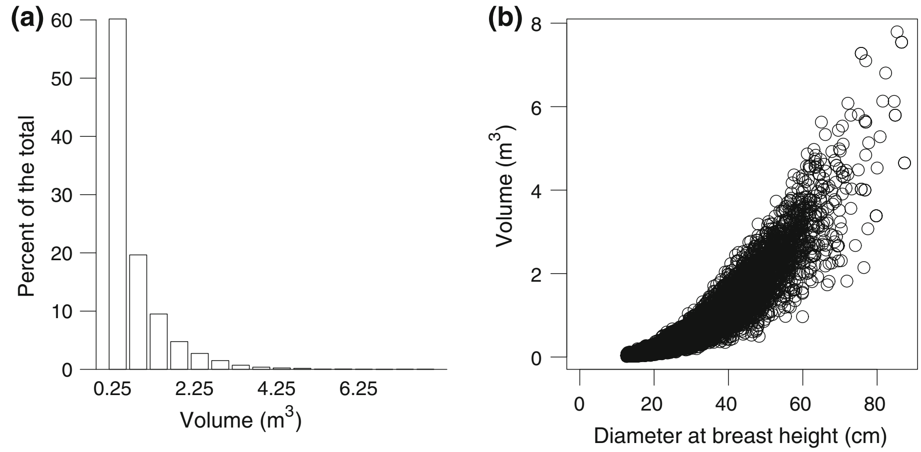

Fig. 1 Histogram of volume (**a**) and scatterplot between volume and diameter at breast height (**b**) for 14,387 loblolly pine trees

SRSwoR) is given as (6.16) in Gregoire and Valentine (2008) as

$$V\left[\widehat{\tau}_{y1}\right] = N^2\left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2_{rm}, \tag{7}$$

where $\sigma^2_{rm} = \frac{1}{N-1}\sum_{k=1}^{N}(y_k - Rx_k)^2$.

The variance of $\widehat{\tau}_{y2}$ is

$$V\left[\widehat{\tau}_{y2}\right] = \left(\frac{1}{n} - \frac{1}{N}\right)\tau^2_x\sigma^2_r, \tag{8}$$

where $\sigma^2_r = \frac{1}{N-1}\sum_{k=1}^{N}(r_k - \mu_r)^2$, as shown in Goodman and Hartley (1958), Eq. (8).

Goodman and Hartley (1958) also derive a variance approximation for $\widehat{\tau}_{y3}$ in their Eq. (6), which is

$$V\left[\widehat{\tau}_{y3}\right] = \left(\frac{1}{n} - \frac{1}{N}\right)\tau^2_y\left[\gamma^2_y + \gamma^2_x - \frac{2C(x,y)}{\mu_x\mu_y}\right], \tag{9}$$

where $C(x, y)$ is the covariance between y and x. Finally, we computed the mean square error (MSE) of each estimator as the sum of its bias square plus its variance, and for interpretation the square root of the MSE (or RMSE), was used (Table 2).

Measurement error processes

As mentioned by Rice (1988), a distinction is usually made between random and systematic ME. Random MEs vary among units of the population. On the other hand, systematic MEs, have the same effect on every measurement. Following Gregoire and Salas (2009) we used 25% of μ_x, which given our population data is equal to 7 cm in diameter, as the maximum ME to be tested.

Systematic measurement error in x We suppose that x_k cannot be measured without a systematic error in measurement denoted by δ^s_k. The magnitude of δ^s_k may be due to a miscalibrated instrument used in the measurement process. The measurement of x_k contaminated with systematic ME is denoted by

$$x^*_k = x_k + \delta^s_k, \tag{10}$$

and likewise $\tau^*_x = \sum_{k=1}^{N} x^*_k$. That is to say, for each level of systematic ME, $\delta^s_k = \delta^s$, then a constant level of ME was added to d_k, the d for the kth element of the population. Thus, (1) computed with ME is $\widehat{\tau}_{y1} = \frac{\bar{y}}{\bar{x}^*}\tau^*_x$; likewise, (2) becomes $\widehat{\tau}_{y2} = \bar{r}^*\tau^*_x$; and (3) is calculated similarly. We use a range of values of δ^s, from -7 to 7 in evenly spaced increments in order to have a total of 11 classes (5 with positive MEs, 5 with negative MEs, and 0 ME).

Random measurement error in x Suppose that the error in the measurement of x is random, rather than systematic, such that the value that is measured is not x_k but

$$x^{**}_k = x_k + \delta^r_k, \tag{11}$$

which implies $\tau^{**}_x = \sum_{k=1}^{N} x^{**}_k$. In (11), δ^r_k varies among the x_k, $k = 1,..., N$. We assume that, on average, the magnitude of δ^r_k is close to zero, yet in any particular sample of n elements, its average is not identically zero, viz.,

$$\bar{\delta}^r_x = \frac{1}{n}\sum_{k=1}^{n}\delta^r_k \neq 0. \tag{12}$$

Let the variance of δ^r_k be denoted by σ^2_δ. In summary, when we are considering systematic MEs, we have $E[\delta^s] = \delta^s$ and $V[\delta^s] = 0$, and $E[\delta^r] = 0$ and $V[\delta^r] = \sigma^2_\delta$, when considering random MEs.

We examined three probability density functions (pdf) to characterize the distribution of the random errors. We used a uniform, Gaussian, and beta pdf as a way to mimic uniformly, symmetrically, and asymmetrically distributed random MEs. We scaled the random MEs in such a way that the maximum (and minimum) δ_k would be close to the maximum and minimum systematic error also tested.

Table 2 Concurrence of simulation moments to the exact or approximate moments of ratio estimators

n	Estimator	Bias (%)		SE (%)		RMSE (%)	
		Theoretical	Empirical	Theoretical	Empirical	Theoretical	Empirical
7	$\widehat{\tau}_{y1}$	−4.09	−4.08	31.61	30.23	31.88	30.51
	$\widehat{\tau}_{y2}$	−23.05	−23.13	22.63	22.65	32.31	32.37
	$\widehat{\tau}_{y3}$	0.00	−0.10	35.06	35.03	35.06	35.03
15	$\widehat{\tau}_{y1}$	−1.91	−1.78	21.59	21.15	21.67	21.22
	$\widehat{\tau}_{y2}$	−23.05	−22.96	15.46	15.43	27.75	27.67
	$\widehat{\tau}_{y3}$	0.00	0.12	23.79	23.77	23.79	23.77
37	$\widehat{\tau}_{y1}$	−0.77	−0.76	13.74	13.58	13.76	13.60
	$\widehat{\tau}_{y2}$	−23.05	−23.07	9.83	9.82	25.06	25.07
	$\widehat{\tau}_{y3}$	0.00	0.00	15.09	15.04	15.09	15.04

100,000 simulations of each size n were conducted

Uniform

Let $\delta_k^r \sim f \times 7 \times U[-1, 1]$, where U is a random number from a uniform distribution, and f is some fraction of the maximum ME to be tested, and 7 is the maximum ME in x to be tested. We use a range of values of f, from 0 to 1 in increments of 0.1, establishing 11 different levels (the same number of levels used for systematic MEs) of random uniformly distributed MEs.

Normal

Let $\delta_k^r \sim f \times \sigma_\delta \epsilon$, and $\epsilon \sim N(0, 1)$, $\sigma_\delta = 0.02\sigma_x$, and f is a fraction of the random ME to be tested. We use a range of values of f, from 0 to 1 in increments of 0.1, establishing 11 different levels of random normally distributed MEs.

Beta

We wished to examine the performance of the estimators under skew ME too. We used the Beta distribution. Specifically, we let $\delta_k^r \sim \beta[a,b] \times f \times 7 / \max(\beta[a,b])$, where a and b are parameters of the distribution, β is a random number from a Beta distribution, f is some fraction of the maximum ME to be tested, 7 is the maximum ME in x to be tested, and $\max(\beta[a,b])$ is the maximum random number from a Beta distribution (obtained when setting the random number seed). We used a range of values of f, from 0 to 1 in increments of 1, establishing 11 different levels of random beta distributed MEs. We fixed the parameters of the Beta pdf to be $a = 2$ and $b = 10$, positive skewed (right-skewed) shape distribution.

Monte Carlo simulation study

Statistical properties of estimators can be assessed using computational re-sampling techniques. We can approximate expected values of estimators by computing the arithmetic average for a large number of simulated samples, and also approximate the distribution of the estimator for these several samples (i.e., empirical sampling distribution). We conducted simulations (each simulation corresponds to an independent random sample) for each combination of ME type and level with samples of sizes (n) 7, 15, and 37. These sample sizes correspond to sampling intensities of 0.05%, 0.10%, and 0.25%, respectively. We conducted 100,000 simulations, and all the analysis were programmed using the free statistical software R (R Development Core Team 2007). The number of simulations was chosen based on a prior analysis for this population in order to make the sampling error of the simulation itself negligibly small. A similar analysis to determine or justify the number of simulation has been conducted by Gregoire and Schabenberger (1999).

Based on the simulations, we computed the empirical estimates of the bias (B), standard error (SE), and root mean square error (RMSE) of each estimator studied. The bias of an estimator relates to the accuracy of it, while the variance of an estimator relates to the precision of it. An estimator should be judged for both accuracy and precision, hence the use of the RMSE of an estimator is more suitable since takes it into account both features. All these statistics were expressed in percentage terms, after dividing them by τ_y.

Assessing variance estimators of the ratio estimators

We also examine the behavior of the estimators of variance for the ratio estimators. The precision of the ratio estimators is judged through an approximate expression for its variance. Therefore it is important to examine the accuracy of this approximation (Raj 1964). Therefore, we compute the empirical bias of the estimates of variance for the ratio estimators. The following variance estimators were used.

For $\widehat{\tau}_{y1}$, we used (6.31) of Gregoire and Valentine (2008), as follows

$$\widehat{v}\left[\widehat{\tau}_{y1}\right] = \frac{N^2\mu_x^2}{\overline{x}^2}\left(\frac{1}{n} - \frac{1}{N}\right)s_{rm}^2, \tag{13}$$

where s_{rm}^2 is an estimator of σ_{rm}^2 of (7):

$$s_{rm}^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(y_k - \widehat{R}x_k\right)^2, \tag{14}$$

and $\widehat{R} = \overline{y}/\overline{x}$.

For $\widehat{\tau}_{y2}$, we used the unbiased estimator of $V\left[\widehat{\tau}_{y2}\right]$ namely

$$\widehat{v}\left[\widehat{\tau}_{y2}\right] = \left(\frac{1}{n} - \frac{1}{N}\right)\tau_x^2 s_r^2, \tag{15}$$

where s_r^2 is the estimator of σ_r^2 of (8):

$$s_r^2 = \frac{1}{n-1}\sum_{k=1}^{n}(r_k - \overline{r})^2, \tag{16}$$

For $\widehat{\tau}_{y3}$, we used the unbiased estimator presented by Goodman and Hartley (1958, Eq. 35). For this estimator, the statistics k_{22}, c, and c' (see Appendix for formulas) must be computed first, followed by the variance estimator. We adjusted[1] the variance estimator of $\widehat{\tau}_{y3}$ presented in (35) of Goodman and Hartley (1958), and the correction of Goodman and Hartley (1969), to

$$\widehat{v}\left[\widehat{\tau}_{y3}\right] = \left(\frac{\tau_x^2 s_r^2}{n} + \frac{2\tau_x c'}{n-2} + \frac{(n-1)s_r^2 s_x^2 + (n-3)c^2 + \left(1-\frac{2}{n}\right)(n-1)k_{22}}{n^2-n-2}\right) \times n \times \left(\frac{1}{n} - \frac{1}{N}\right)N^2, \tag{17}$$

where s_r^2 and s_x^2 are the sample variance of r and x, respectively.

We assessed these variance estimators (Eqs. 13, 15, and 17) using our simulations described in Section "Monte Carlo simulation study". The bias in percentage terms of the estimator of variance was obtained dividing the bias by the empirical variance of the corresponding ratio estimator.

Results

Without measurement error in x

Both theoretical and empirical results (i.e., bias, SE, and RMSE) are almost identical, with differences (for most cases) smaller than 0.1% (Table 2). Theoretical formulas for bias, SE, and RMSE for the ratio estimators are presented in Gregoire and Salas (2009). The mean-of-ratios estimator, $\widehat{\tau}_{y2}$, had the largest bias (underestimation) for all sample sizes: its magnitude is unacceptably large, and does not diminish with increasing the sample size (second row of Fig. 2). The ratio-of-means estimator, $\widehat{\tau}_{y1}$ had a bias less than that of $\widehat{\tau}_{y2}$ (smaller than −4.1% for the smallest sample size). Its bias decreases to less than −0.8% with increasing sample size. $\widehat{\tau}_{y2}$, although biased, had the best precision (smaller standard error values) for all sample sizes. The ratio-of-means estimator had the smallest RMSE, followed by $\widehat{\tau}_{y3}$, with small difference with increasing sample size (Table 2). $\widehat{\tau}_{y2}$ performs similar to $\widehat{\tau}_{y3}$ when $n = 7$, but the RMSE of mean-of-ratios does not decrease much with increasing sample size because its bias is invariant to sample size. The precision clearly increases when the sample size increases (Fig. 2).

With measurement error in x

We report the effect of ME on bias by considering the ratio of bias with ME to bias in the absence of ME, which we call relative bias henceforth. The extent to which this ratio is smaller (greater) than unity is a measure of the relative decrease (increase) in bias due to ME. In an analogous fashion, we report relative SE and relative RMSE.

Systematic measurement error in x Positive MEs tend to decrease the absolute bias of the ratio estimators, especially for the mean-of-ratios estimator (first panel-row inner plots of Fig. 3). Estimator 3 remained unbiased under ME in x, as also noticed by Gregoire and Salas (2009). On the other hand, if we consider the ME effect in the relative bias (i.e., B^*/B), positive MEs tend to reduce the B^*/B of both the ratio-of-means and the mean-of-ratios estimator by approximately 10% (first panel-row of Fig. 3).

The mean-of-ratios estimator had better precision than the other ratio estimators under systematic MEs for all the sample sizes (second panel-row inner plots of Fig. 3). If we only consider the effect of ME, it is possible to infer that positive systematic ME on x decreases the precision of all the ratio estimators tested in our study. Conversely, negative ME produces better precision (second panel-row of Fig. 3). Gregoire and Salas (2009) found the opposite trend in a similar study but with a different distribution of both target and auxiliary variables. The effect of ME in the precision is reduced with increasing sample size. Overall, even though large values of MEs were added to x, this alters the precision of the ratio estimators comparatively

[1] The formula given by these authors is for an estimator of the population mean and assuming infinite populations; but here we are dealing with population total and finite populations.

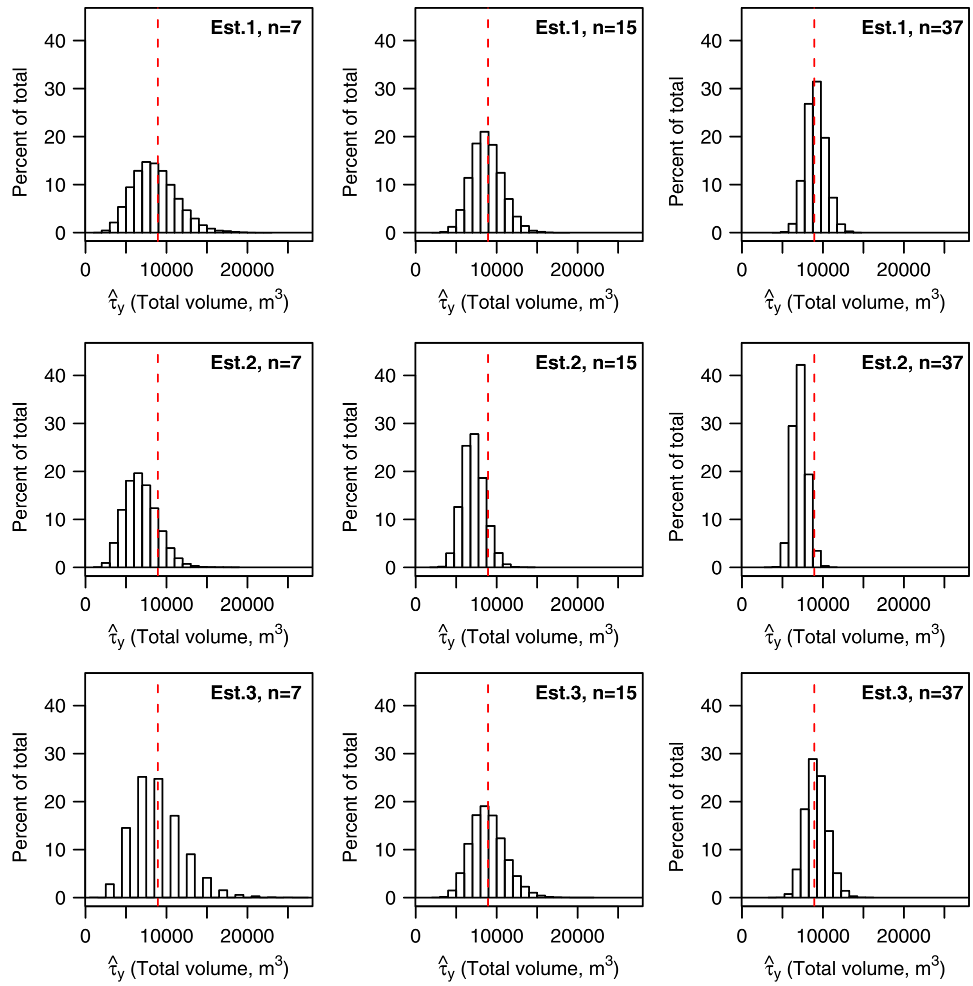

Fig. 2 Empirical distribution of three different ratio estimators and three sample sizes (100,000 simulations were conducted) in predicting total volume of loblolly pine. The *dashed vertical line* represents the value of the target parameter τ_y, Est.1 is $\hat{\tau}_{y1}$, Est.2 is $\hat{\tau}_{y2}$, and Est.3 is $\hat{\tau}_{y3}$

little (<10%) compared to their precision in the absence of ME.

Root mean squares errors (RMSE%) slightly increase for positive systematic MEs, except for the mean-of-ratios estimator. According to this statistic, ratio-of-means performs the best for all conditions, with $\hat{\tau}_{y3}$ next best (third panel-row inner plots of Fig. 3). Positive ME increases the RMSE compared to the RMSE from x with no ME, for both ratio-of-means and estimator 3, but decreased for estimator 2 (third panel-row of Fig. 3). $\hat{\tau}_{y1}$, even though slightly biased, performed best among all the ratio estimators tested.

Random measurement error in x The effect of random ME in x on the ratio estimators is smaller than the effect of systematic ME. Neither the accuracy nor the precision of the ratio estimators are very affected by uniform, Gaussian, and beta (inner plots of Figs. 4, 5, and 6, respectively) distributed MEs. Only a minor change in the bias of each estimator with and without ME was observed (Figs. 4, 5, and 6 for uniform, Gaussian, and beta MEs, respectively). The effect of random ME is more notable when these errors are uniformly distributed than when they are either Gaussian or beta distributed.

Assessing estimators of variance of the ratio estimators

Let us explain our results for the following three scenarios: When x does not contain ME, the variance estimator of the ratio-of-means estimator (i.e., $\hat{v}[\hat{\tau}_{y1}]$) is highly biased for the smallest sample size (Table 3). Although $\hat{v}[\hat{\tau}_{y2}]$ and $\hat{v}[\hat{\tau}_{y3}]$ do not achieve a zero bias, the largest value is only 0.77%. With systematic MEs, both $\hat{v}[\hat{\tau}^*_{y2}]$ and $\hat{v}[\hat{\tau}^*_{y3}]$ retain their unbiasedness, while $\hat{v}[\hat{\tau}^*_{y1}]$ is more biased (underestimation) for negative MEs than positive MEs. With random MEs (Table 4), both $\hat{v}[\hat{\tau}^*_{y2}]$ and $\hat{v}[\hat{\tau}^*_{y3}]$ exhibit small bias. For $\hat{v}[\hat{\tau}^*_{y1}]$, the magnitude of its

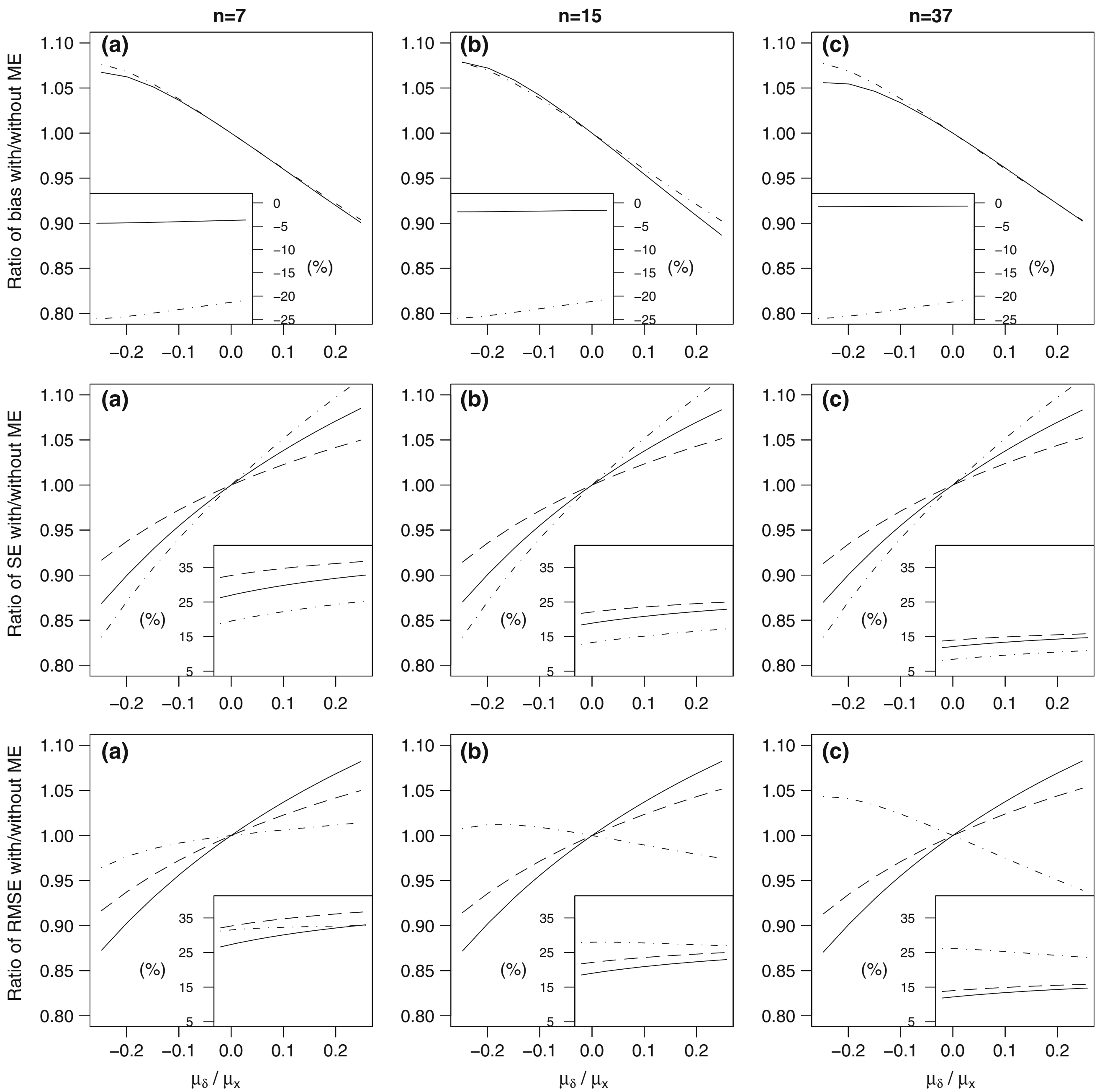

Fig. 3 Bias (*first panel-row*), standard error (*second panel-row*), and root mean square error (*third panel-row*) of $\widehat{\tau}^{*}_{y1}$ (*solid line*), $\widehat{\tau}^{*}_{y2}$ (*dot-dash line*), and $\widehat{\tau}^{*}_{y3}$ (*dashed line*) relative to that of $\widehat{\tau}_{y1}$, $\widehat{\tau}_{y2}$, and $\widehat{\tau}_{y3}$, respectively, with systematic measurement error in the auxiliary variate, and having samples of 7 (**a**), 15 (**b**), and 37 (**c**) trees. The inner plots represent bias (*first panel-row*), standard error (*second panel-row*), and root mean square error (*third panel-row*) expressed as a percentage of τ_y. The quotient μ_δ/μ_x represents the relative level of measurement of error with respect to the population mean of the auxiliary variate x. The horizontal axis of the inner plots span the same range as the axis of the larger plots

underestimation decreases slightly as σ_δ/σ_x increases. There are only slight differences in the estimators of the variance among the types of the distribution of the random ME, being the bias greater for the Gaussian, then the Beta, and finally the uniform.

Discussion

Our results show important differences in accuracy and precision for all the estimators evaluated when x is measured without error. While Gregoire and Salas (2009)

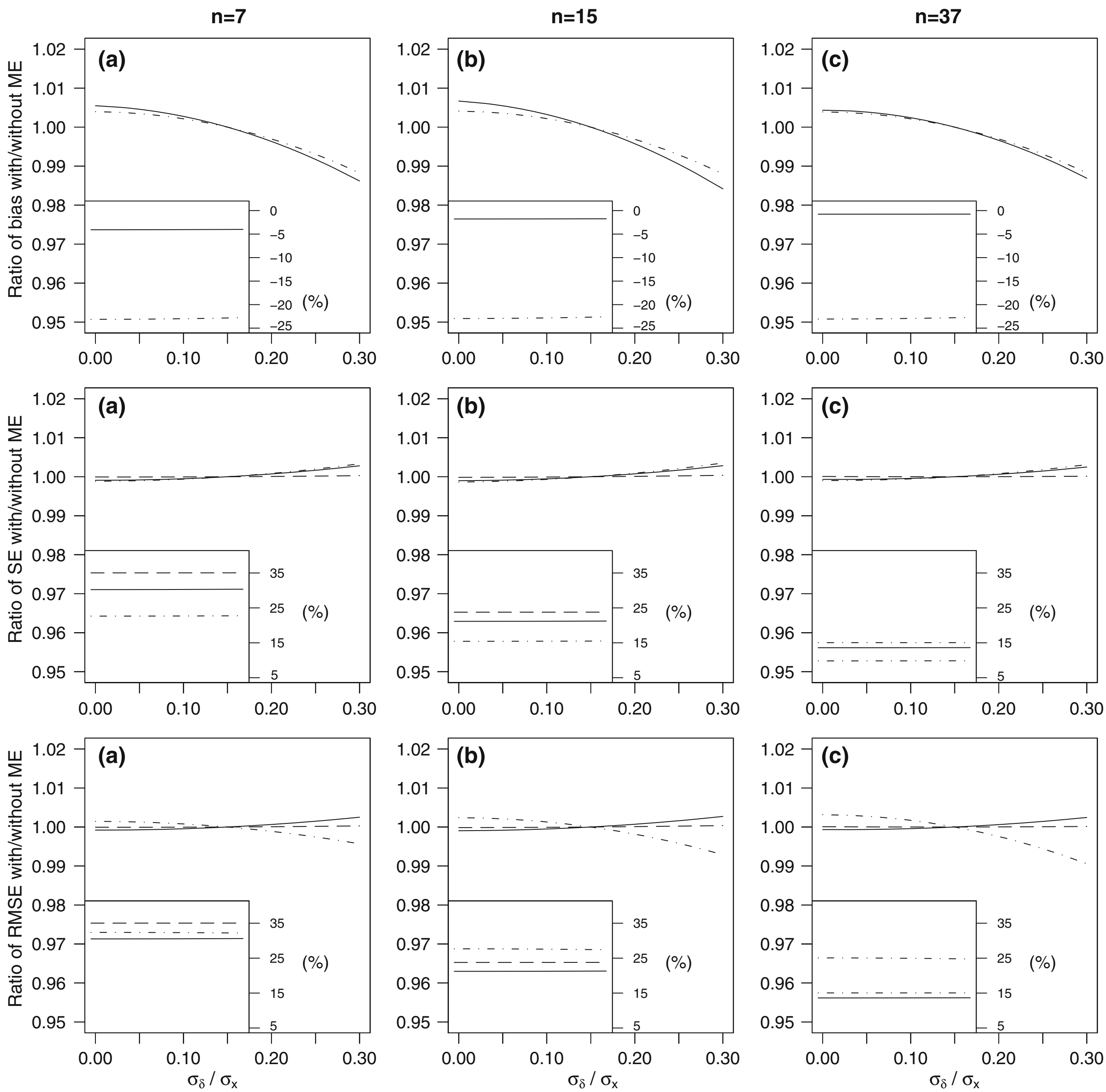

Fig. 4 Bias (*first panel-row*), standard error (*second panel-row*), and root mean square error (*third panel-row*) of $\widehat{\tau}_{y1}^{*}$ (*solid line*), $\widehat{\tau}_{y2}^{*}$ (*dot-dash line*), and $\widehat{\tau}_{y3}^{*}$ (*dashed line*) relative to that of $\widehat{\tau}_{y1}$, $\widehat{\tau}_{y2}$, and $\widehat{\tau}_{y3}$, respectively, with Uniform distributed measurement error in the auxiliary variate, and having samples of 7 (**a**), 15 (**b**), and 37 (**c**) trees. The inner plots represent bias (*first panel-row*), standard error (*second panel-row*), and root mean square error (*third panel-row*) expressed as a percentage of τ_y. The quotient σ_δ/σ_x represents the relative level of variation of the random measurement error with respect to the variation of the auxiliary variate x in the population. The horizontal axis of the inner plots span the same range as the axis of the larger plots

found that the mean-of-ratios estimator had a bias less than 2%, with the loblolly pine data $\widehat{\tau}_{y2}$ was very biased. We believe this is due to the fact that in our study the target variate has much greater variability, with a coefficient of variation of 120.4% versus 29.7% for the leaf area population studied by Gregoire and Salas (2009). Both studies reaffirm the recommendation of Ek (1971), who advocated against the use of the mean-of-ratios estimator because of its sometimes severe bias. On the other hand, the mean-of-ratios estimator always had better precision than the other two estimators. Even though $\widehat{\tau}_{y1}$ is slightly biased it performed better than the unbiased $\widehat{\tau}_{y3}$ in terms of RMSE.

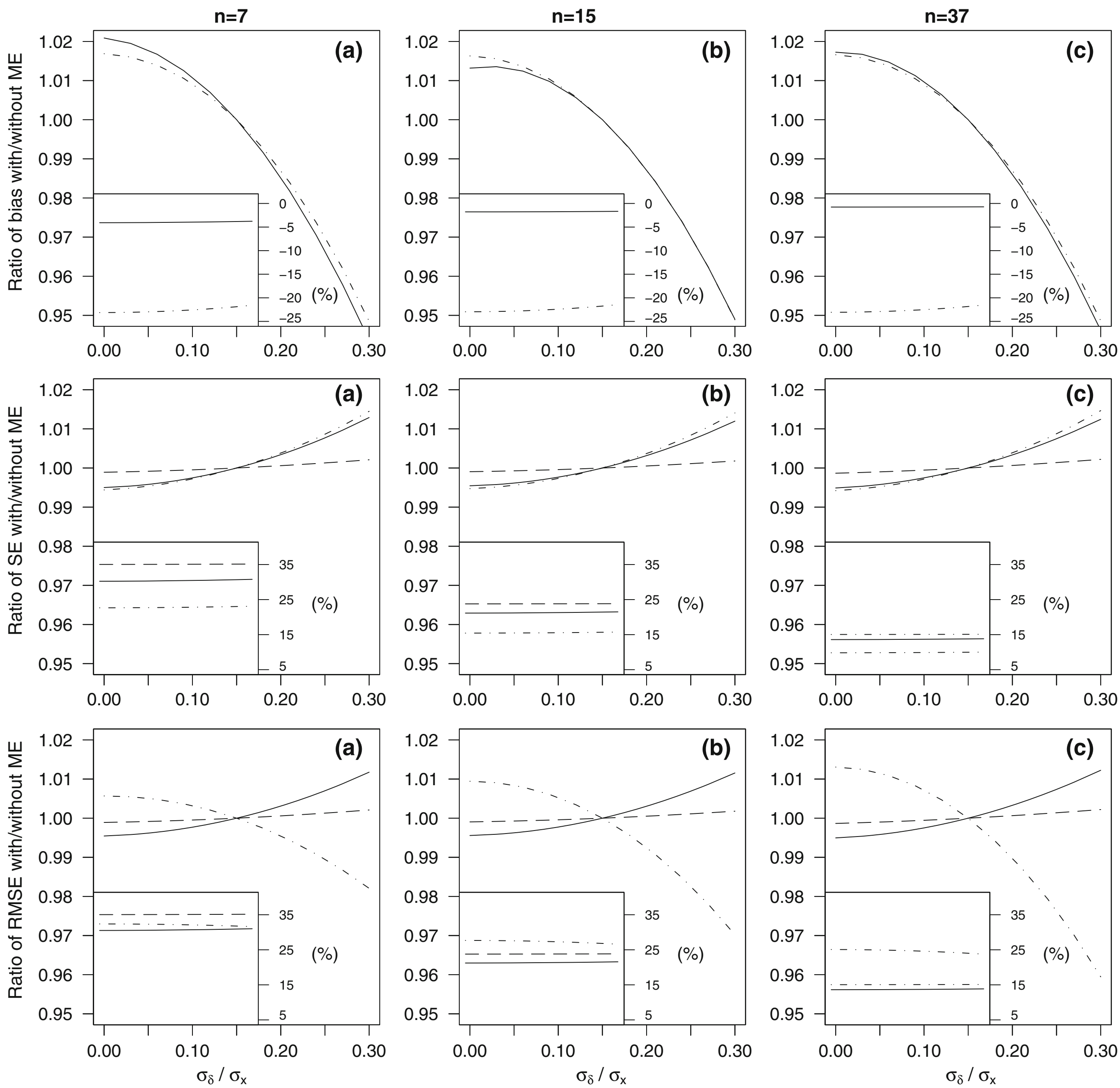

Fig. 5 Bias (*first panel-row*), standard error (*second panel-row*), and root mean square error (*third panel-row*) of $\widehat{\tau}^*_{y1}$ (*solid line*), $\widehat{\tau}^*_{y2}$ (*dot-dash line*), and $\widehat{\tau}^*_{y3}$ (*dashed line*) relative to that of $\widehat{\tau}_{y1}$, $\widehat{\tau}_{y2}$, and $\widehat{\tau}_{y3}$, respectively, with Gaussian distributed measurement error in the auxiliary variate, and having samples of 7 (**a**), 15 (**b**), and 37 (**c**) trees. The inner plots represent bias (*first panel-row*), standard error (*second panel-row*), and root mean square error (*third panel-row*) expressed as a percentage of τ_y. The quotient σ_δ/σ_x represents the relative level of variation of the random measurement error with respect to the variation of the auxiliary variate x in the population. The horizontal axis of the inner plots span the same range as the axis of the larger plots

Overall, we did not find any important advantage of using $\widehat{\tau}_{y3}$ (which adds a correction to $\widehat{\tau}_{y2}$ using the HT estimators) over the ratio-of-means estimator. The ratio-of-means estimator is easier to compute than $\widehat{\tau}_{y3}$, which might be important for practitioners.

Systematic MEs had a slight effect on the performance of the ratio estimators. There is only a slight effect of MEs on the bias of the ratio-of-means and the mean-of-ratios estimator (Fig. 3). *B**/*B* can also be checked without conducting any simulations using the formulas given by Gregoire and Salas (2009). On average, adding 7 to x for our population is equivalent to a 25% of the average value of d which is a large value of ME in diameter in the field. The *SE**/*SE* values lower than 1 for negative MEs of all the

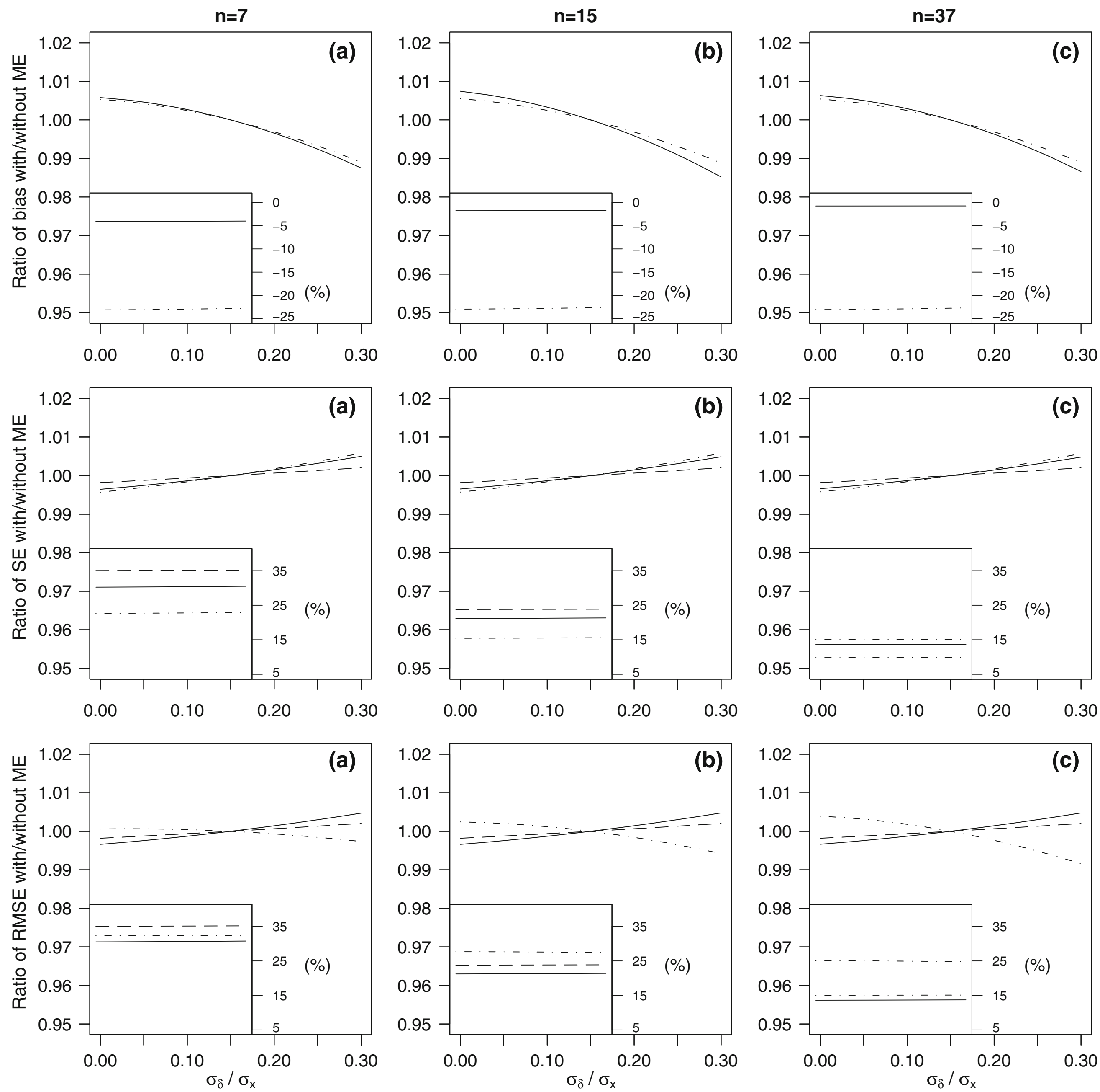

Fig. 6 Bias (*first panel-row*), standard error (*second panel-row*), and root mean square error (*third panel-row*) of $\widehat{\tau}^*_{y1}$ (*solid line*), $\widehat{\tau}^*_{y2}$ (*dot-dash line*), and $\widehat{\tau}^*_{y3}$ (*dashed line*) relative to that of $\widehat{\tau}_{y1}, \widehat{\tau}_{y2}$, and $\widehat{\tau}_{y3}$, respectively, with Beta distributed measurement error in the auxiliary variate, and having samples of 7 (**a**), 15 (**b**), and 37 (**c**) trees. The inner plots represent bias (*first panel-row*), standard error (*second panel-row*), and root mean square error (*third panel-row*) expressed as a percentage of τ_y. The quotient σ_δ/σ_x represents the relative level of variation of the random measurement error with respect to the variation of the auxiliary variate x in the population. The horizontal axis of the inner plots span the same range as the axis of the larger plots

estimators imply that the SE is decreased in comparison to the SE when x is measured without error. On the other hand, *SE*/SE* values greater than 1 for positive MEs imply an increasing of the SE in comparison to the SE when x in measured without error. Overall, systematic ME slightly increases the accuracy but decreases the precision of the ratio estimators.

Uniform, Gaussian, and Beta distributed MEs in x degrade neither the accuracy nor the precision of these ratio estimators. Gregoire and Salas (2009) found that the

Table 3 Bias of the estimator of the variance, as a percentage of the Monte Carlo variance, for the ratio estimators under several levels of the quotient between the population mean of the systematic measurement error (μ_δ) in x and the population mean of x (μ_x), and different sample sizes

n	Estimator	μ_δ/μ_x										
		−0.25	−0.2	−0.15	−0.1	−0.05	0	0.05	0.1	0.15	0.2	0.25
7	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−24.59	−22.98	−21.55	−20.31	−19.22	−18.25	−17.40	−16.61	−15.91	−15.30	−14.70
	$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	−0.26	−0.19	−0.26	−0.26	−0.25	−0.23	−0.23	−0.21	−0.22	−0.21	−0.18
	$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.64	0.60	0.61	0.57	0.52	0.53	0.52	0.48	0.50	0.48	0.44
15	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−11.93	−11.16	−10.36	−9.80	−9.22	−8.73	−8.26	−7.87	−7.51	−7.15	−6.84
	$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	0.46	0.53	0.54	0.41	0.43	0.53	0.51	0.43	0.44	0.46	0.49
	$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.43	0.42	0.42	0.47	0.41	0.48	0.47	0.50	0.43	0.46	0.44
37	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−4.37	−3.94	−3.73	−3.50	−3.18	−2.99	−2.89	−2.69	−2.49	−2.40	−2.26
	$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	0.47	0.40	0.33	0.44	0.43	0.39	0.39	0.46	0.45	0.39	0.49
	$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.82	0.72	0.69	0.75	0.71	0.77	0.71	0.71	0.65	0.70	0.74

100,000 simulations of each size n were conducted

greater the variability of the random error distribution, the greater the bias, variance, and RMSE of the estimator. Nevertheless, the difference between B, SE, and RMSE with and without ME reported by them is smaller than 3%. There is no difference between a symmetric ME distribution (i.e. Gaussian and uniform) and a skewed ME distribution (i.e. Beta) on the performance of the ratio estimators.

Only the variance estimators of the ratio-of-means estimator is biased. The bias in $\hat{v}\left[\hat{\tau}_{y1}\right]$ is especially large for an extremely small sample size such as $n = 7$. Cochran (1977) pointed out that $\hat{v}\left[\hat{\tau}_{y1}\right]$ is based on large sample theory. Our results show underestimation of the variance of the ratio-of-means estimator, confirming Cochran's (1977, p. 162) assertion that the large sample approximation results in underestimation. We have also found similar results to those reported by Rao (1968) where the bias in the variance estimators, mainly for ratio-of-means, are more serious in small samples. Rao (1968) mention as well that these results are unsatisfactory at least up to $n = 12$, which is similar to our medium-size sample ($n = 15$). The usual approximation of the variance estimator of ratio-of-means would be adequate in large samples if the data follow a bivariate normal distribution (Sukhatme and Sukhatme 1970). The highest bias reported here is larger than the −9% mentioned by Cochran (1977), who used Sukhatme and Sukhatme's (1970) theoretical results, the estimator of variance of ratio-of-means, but smaller than the −25% of Koop (1968) for small population sizes. Neither systematic nor random MEs affect the unbiasedness of the variance estimators of $\hat{\tau}^*_{y2}$ and $\hat{\tau}^*_{y3}$. Only systematic MEs in x affect the bias of the variance estimates of the ratio-of-means estimator. Finally, the bias of $\hat{v}\left[\hat{\tau}_{y1}\right]$ raise an interesting point regarding its effect in statistical inference (e.g., in computing confidence intervals), where further research is needed. Furthermore, the precision of the variance estimators can be also assessed.

Concluding remarks

The statistical performance of ratio estimators is not very affected by the presence of either systematic or random ME in the auxiliary variate. Only some slight effect on bias was found when having systematic MEs. This resistance of the ratio estimators to ME revealed in this study confirms the results of Gregoire and Salas (2009). The ratio-of-means estimator performs the best in terms of RMSE. The unbiased estimator, $\hat{\tau}_{y3}$, does not provide precise enough estimation to perform better than the ratio-of-means estimator. The mean-of-ratios estimator is highly biased, yet always was more precise. The unacceptably large bias of $\hat{\tau}_{y2}$ observed in this study contrasts with the results of Gregoire and Salas (2009). We suspect that it is due to characteristics of the population being sampled that we have yet to identify. Neither systematic nor random ME affect the bias of the variance estimates of the ratio estimators. For small sample sizes, the estimator of the variance of $\hat{\tau}_{y1}$ has unacceptably large negative bias.

Table 4 Bias of the estimator of the variance of the ratio estimators, as a percentage of the Monte Carlo variance, of ratio estimators under several levels of the quotient between the standard deviation of three different random distributed measurement error (σ_δ) in x and the standard deviation of x (σ_x), and different sample sizes

Distribution	n	Estimator	σ_δ/σ_x										
			0	0.03	0.06	0.09	0.12	0.15	0.18	0.21	0.24	0.27	0.3
Uniform	7	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−18.25	−18.30	−18.28	−18.26	−18.23	−18.19	−18.15	−18.04	−17.99	−17.92	−17.84
		$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	−0.23	−0.24	−0.27	−0.24	−0.22	−0.23	−0.26	−0.30	−0.36	−0.34	−0.33
		$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.53	0.49	0.52	0.50	0.49	0.48	0.48	0.49	0.45	0.48	0.45
	15	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−8.73	−8.75	−8.71	−8.70	−8.63	−8.59	−8.57	−8.57	−8.52	−8.49	−8.41
		$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	0.53	0.48	0.49	0.44	0.46	0.42	0.44	0.41	0.45	0.44	0.51
		$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.48	0.41	0.44	0.48	0.43	0.40	0.46	0.43	0.42	0.41	0.41
	37	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−2.99	−3.07	−3.09	−3.05	−3.09	−3.08	−3.14	−3.15	−3.10	−3.13	−3.11
		$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	0.39	0.47	0.41	0.42	0.30	0.24	0.25	0.33	0.29	0.31	0.22
		$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.77	0.66	0.69	0.73	0.64	0.69	0.61	0.54	0.61	0.55	0.50
Gaussian	7	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−18.25	−18.30	−18.28	−18.25	−18.26	−18.25	−18.22	−18.23	−18.23	−18.22	−18.19
		$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	−0.23	−0.22	−0.19	−0.24	−0.18	−0.19	−0.18	−0.25	−0.21	−0.24	−0.25
		$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.53	0.52	0.52	0.52	0.53	0.53	0.54	0.49	0.50	0.51	0.52
	15	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−8.73	−8.72	−8.70	−8.75	−8.70	−8.72	−8.73	−8.72	−8.69	−8.65	−8.69
		$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	0.53	0.42	0.45	0.50	0.43	0.52	0.49	0.47	0.48	0.37	0.42
		$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.48	0.48	0.48	0.48	0.40	0.40	0.41	0.42	0.43	0.44	0.46
	37	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−2.99	−2.99	−2.97	−2.94	−3.04	−2.98	−3.04	−2.95	−2.99	−3.01	−3.02
		$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	0.39	0.40	0.43	0.48	0.54	0.42	0.51	0.42	0.55	0.50	0.46
		$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.77	0.77	0.77	0.77	0.77	0.77	0.78	0.79	0.79	0.80	0.82
Beta	7	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−18.25	−18.24	−18.22	−18.19	−18.20	−18.14	−18.13	−18.10	−18.07	−18.02	−17.97
		$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	−0.23	−0.17	−0.19	−0.20	−0.20	−0.27	−0.24	−0.20	−0.23	−0.26	−0.27
		$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.53	0.54	0.49	0.51	0.52	0.54	0.50	0.52	0.48	0.51	0.48
	15	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−8.73	−8.69	−8.65	−8.68	−8.61	−8.61	−8.61	−8.59	−8.57	−8.53	−8.56
		$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	0.53	0.55	0.45	0.49	0.55	0.48	0.43	0.52	0.50	0.48	0.48
		$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.48	0.46	0.45	0.44	0.43	0.43	0.42	0.50	0.50	0.50	0.50
	37	$\widehat{var}\left[\hat{\tau}^*_{y1}\right]$	−2.99	−3.02	−3.03	−3.02	−3.01	−2.99	−2.95	−2.91	−2.99	−2.93	−2.85
		$\widehat{var}\left[\hat{\tau}^*_{y2}\right]$	0.39	0.33	0.50	0.47	0.45	0.44	0.45	0.47	0.49	0.53	0.38
		$\widehat{var}\left[\hat{\tau}^*_{y3}\right]$	0.77	0.71	0.79	0.73	0.67	0.75	0.70	0.78	0.73	0.68	0.77

100,000 simulations of each size n were conducted

Acknowledgments We gratefully acknowledge Roy C. Beltz, U.S. Forest Service, Forestry Sciences Lab, Starkville, Mississippi for providing the population data used in our study.

Appendix: Expressions needed for computing the estimator of the variance of $\widehat{\tau}_{y3}$

We used the unbiased estimator presented by Goodman and Hartley (1958, Eq. 35). This estimator requires the computation of k_{22}, c, c'. These statistics are computed as follows,

* k_{22}

$$k_{22} = \frac{n}{(n-1)(n-2)(n-3)} \times \left\{ (n+1)S_{22} - \frac{2(n+1)}{n}(S_{21}S_{01} + S_{12}S_{10}) - \frac{(n-1)}{n}(S_{20}S_{02} + 2S_{11}^2) + \frac{2}{n}(S_{20}S_{01}^2 + S_{02}S_{10}^2 + 4S_{11}S_{10}S_{01}) - \frac{6}{n^2}(S_{10}^2 S_{01}^2) \right\}, \tag{18}$$

where

$$S_{tj} = \sum_{k=1}^{n} x_k^t r_k^j, \tag{19}$$

for example $S_{22} = \sum_{k=1}^{n} x_k^2 r_k^2 = \sum_{k=1}^{n} x_k^2 (y_k^2/x_k^2) = \sum_{k=1}^{n} y_k^2$.

Note that our expression for k_{22} (Eq. 18) has some algebraic manipulations compared that one gave by Goodman and Hartley (1958, Eq. 30), and also considering the corrections made by Goodman and Hartley (1969).

* c. From Goodman and Hartley (1958, Eq. 36, part a)

$$c = \frac{1}{n(n-1)} \left[n \sum_{k=1}^{n} y_k - \left(\sum_{k=1}^{n} x_k \sum_{k=1}^{n} r_k \right) \right], \tag{20}$$

which is the sample covariance between x and r as at the bottom of page 497 of Goodman and Hartley (1958), as follows

$$c = \left(\frac{1}{n-1} \right) \sum_{i=1}^{n} (x_k - \overline{x})(r_k - \overline{r}), \tag{21}$$

* c'. From Goodman and Hartley (1958, Eq. 32)

$$c' = \left(\frac{1}{n-1} \right) \sum_{k=1}^{n} (x_k - \overline{x})(r_k - \overline{r})^2. \tag{22}$$

References

Bay J, Stefanski LA (2000) Adjusting data for measurement error to reduce bias when estimating coefficients of a quadratic model. In: Proceedings of the Survey Research Methods Section, American Statistical Association, pp 731–733

Canavan SJ, Hann DW (2004) The two-stage method for measurement error characterization. For Sci 50(6):743–756

Chandhok PK (1988) Stratified sampling under measurement error. In: Proceedings of the Survey Research Methods Section, American Statistical Association, pp 508–510

Cochran WG (1977) Sampling techniques, 3rd edn. Wiley, New York, USA, 428 pp

Cunia T (1965) Some theories on reliability of volume estimates in a forest inventory sample. For Sci 11(1):115–127

Dryver AL, Chao CT (2007) Ratio estimators in adaptive cluster sampling. Environmetrics 18:607–620

Ek AR (1971) A comparison of some estimators in forest sampling. For Sci 17(1):2–13

Frayer WE, Furnival GM (1999) Forest survey sampling designs: a history. J For 97(12):4–10

Fuller WA (1987) Measurement error models. Wiley, USA, 464 pp

Gertner GZ (1988) Regressor variable errors and the estimation and prediction with linear and nonlinear models. In: Sloboda B (ed) Biometric models and simulation techniques for process of research and applications in forestry. Schriften aus der Forslichen Fakultat der Universität Göttingen, Göttingen, Germany, Band No. 160, pp 54–65

Gertner GZ (1990) The sensitivity of measurement error in stand volume estimation. Can J For Res 20(6):800–804

Goodman LA, Hartley HO (1958) The precision of unbiased ratio-type estimators. J Am Stat Assoc 53(282):491–508

Goodman LA, Hartley HO (1969) Corrigenda: the precision of unbiased ratio-type estimators. J Am Stat Assoc 64(328):1700

Gregoire TG (1998) Design-based and model-based inference in survey sampling: appreciating the difference. Can J For Res 28(10):1429–1447

Gregoire TG, Salas C (2009) Ratio estimation with measurement error in the auxiliary variate. Biometrics 62(2). doi:10.1111/j.1541-0420.2008.01110.x

Gregoire TG, Schabenberger O (1999) Sampling-skewed bilogical populations: behavior of confidence intervals for the population total. Ecology 80(3):1056–1065

Gregoire TG, Valentine HT (2008) Sampling strategies for natural resources and the environment. Chapman & Hall/CRC, New York, 474 pp

Gregoire TG, Williams M (1992) Identifying and evaluating the components of non-measurement error in the application of standard volume equations. Statistician 41(5):509–518

Grosenbaugh LR (1964) Some suggestions for better sample-tree measurement. In: Anon (ed) Proceedings. Society of American Foresters, Boston, MA, USA, pp 36–42

Hansen MH, Hurwitz WN, Marks ES, Mauldin WP (1951) Response errors in surveys. J Am Stat Assoc 46(254):147–190

Hartley HO, Ross A (1954) Unbiased ratio estimators. Nature 174(4423):270–271

Hordo M, Kiviste A, Sims A, Lang M (2008) Outliers and/or measurement errors on the permanent sample plot data. In: Reynolds KM (ed) Proceedings of the sustainable forestry in theory and practice: recent advances in inventory and monitoring, statistics and modeling, information and knowledge management, and policy science. USDA For Serv Gen Tech Rep, PNW-688. Portland, OR, USA, p 15

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47(260):663–685

Hutchison MC (1971) A monte carlo comparison of some ratio estimators. Biometrika 58(2):313–321

Kangas A (1996) On the bias and variance in tree volume predictions due to model and measurement errors. Scand J For Res 11:281–290

Kangas A (1998) Effects of errors-in-variables on coefficients of a growth model on prediction of growth. For Ecol Manag 102(2):203–212

Kangas AS, Kangas J (1999) Optimization bias in forest management planning solutions due to errors in forest variables. Silva Fenn 33(4):303–315

Koop JC (1968) An exercise in ratio estimation. Ann Math Stat 22(1):29–30

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

Mickey MR (1959) Some finite population unbiased ratio and regression estimators. J Am Stat Assoc 59(287):594–612

Myers RH (1990) Classical and modern regression with applications, 2nd edn. Duxbury, Pacific Grove, 488 pp

Poso S, Wang G, Tuominen S (1999) Weighting alternative estimates when using multi-source auxiliary data for forest inventory. Silva Fenn 33(1):41–50

R Development Core Team (2007) R: a language and environment for statistical computing. Available from http://www.R-project.org [version 2.5.0]. R Foundation for Statistical Computing, Vienna, Austria

Raj D (1964) A note on the variance of ratio estimate. J Am Stat Assoc 59(307):895–898

Rao JNK (1968) Some small sample results in ratio and regression estimation. J Ind Stat Assoc 6:160–168

Rice JA (1988) Mathematical statistics and data analysis. Wadsworth, Pacific Grove, 595 pp

Robinson AP, Hamlin DC, Fairweather SE (1999) Improving forest inventories: three ways to incorporate auxiliary information. J For 97(12):38–42

Royall RM, Cumberland WG (1981) An empirical study of the ratio estimator and estimators of its variance. J Am Stat Assoc 76(373):66–77

Scali J, Testa V, Kahr M, Strudler M (2005) Measuring nonsampling error in the statistics of income individual tax return study. In: Proceedings of the survey research methods section, American Statistical Association, pp 3520–3525

Stage AR, Wykoff WR (1998) Adapting distance-independent forest growth models to represent spatial variability: effects of sampling design on model coefficients. For Sci 44(2):224–238

Sukhatme PV, Sukhatme BV (1970) Sampling theory of surveys with applications, 2nd edn. Iowa State University Press, Ames, 452 pp

Tin M (1965) Comparison of some ratio estimators. J Am Stat Assoc 60(309):294–307