Optimization of the forest volume estimation with plot-double sampling when there is a walking cost of the second-phase sample

GEORGE STAMATELLOS (*) - VAIOS BLIOUMIS (**)

Abstract

The Lagrange multiplier method was applied for estimating the forest volume at minimum cost by plot-double sampling. A composite cost function was used which has also included the walking cost of the second-phase sample. Compared to the plot sampling the error was increased per 10,45% and the cost decreased per 7,35%. When the volume measurement cost is increased a significant decrease of the total cost of the plot-double against plot sampling is observed

Résumé

La méthode multiplicatrice de Lagrange a été appliquée pour estimer le volume des forêts au moindre coût par l'échantillonnage "plot-double". Une fonction de coût composée a inclus aussi le coût de marche de l'échantillon de la deuxième phase. Par rapport à l'échantillonnage "plot", l'erreur a augmenté de 10,45% et le coût a baissé de 7,35%. Lorsque le coût de la mesure de volume augmente, le coût total du "plot-double" diminue d'une manière significative par rapport à l'échantillonnage "plot".

The cost is generally an important factor in the management of natural resources while its effect on the forest inventory is determinant. In extensive inventories the selection between alternative designs can be made by taking into account a number of factors. A factor of prime importance is the cost-effectiveness, which is measured as the relative efficiency between design alternatives (Scott and Kohl, 1993). The optimization in sampling is attained either by minimizing the variance with fixed cost or by minimizing the cost with fixed variance. In the second case the complexity of the problem solution depends partly on the form of the cost function. In two-phase sampling are mostly used simple cost functions; therefore it is possible to have types of the optimum size of two-phase samples. However, in forest inventories the access cost for locating the sampling plots (cost of walking) is determinant for the general planning and selection of inventory. Consequently, this should be included into the cost function resulting to be more complex. An application of such a cost function for cluster sampling (plot and per Bitterlich points) was done by O'Regan and Arvanitis (1966). In the plot-double sampling it is possible, with a correct initial planning, after one visit to the forest to have the required information for both sampling phases. However, sometimes is necessary to make a second visit to the forest (e.g. for reasons of statistical investigation, stratification, etc.) so that in the cost function also is included the cost of the second walking for locating the sampling plots of the second phase. The paper aims at finding the cost efficiency for the estimation of forest volume with a such cost function, which will also include the cost of walking of sampling units (plots) of the second phase. Comparison of results with the respective plot sampling will provide useful conclusions for the practice.

(*) Lecturer in Forest Biometrics - Dept. of Forestry and Natural Environment, Aristotle University of Thessaloniki, Greece.
(**) Lecturer in Forest Economics - Dept. of Forestry and Natural Environment, Aristotle University of Thessaloniki, Greece.

Methods

Assume we have a forest area of A ha. The plot-double sampling for the estimation of its total volume may be applied as follows: a simple random sample is taken by replacing of n' plots (1st phase) in which the basal area per ha is measured and in a subset of it of n plots (2nd phase) the volume per ha is measured. The regression estimator of volume per ha ($\bar{Y}_{lr}$) and its estimated variance var($\bar{Y}_{lr}$) (Cochran 1977, De Vries 1986, Matis 1992, Schreuder et al., 1994) are:

$$\bar{Y}_{lr} = \bar{y} + b(\bar{x}' - \bar{x}) \quad (1)$$

$$\mathrm{var}(\bar{Y}_{lr}) = s_y^2\left(\frac{1}{n} - \frac{n'-n}{nn'} \cdot \rho^2\right) \quad (2)$$

where $\bar{y}$ and $\bar{x}$ are the estimations of volume and basal area per ha calculated by the n sampling plots (small sample), $\bar{x}'$ is the estimation of basal area per ha from the n' plots (large sample), b the estimated regression coefficient, ρ the estimated correlation coefficient between the x and y and s_y^2 the variance of y. The b, ρ and s_y^2 are estimated by the n plots. Taking as base the cost function of O'Regan and Arvanitis (1966), a composite cost function for the plot-double sampling having in mind also the walking cost of the second-phase sample, is:

$$C = k_1'\sqrt{n'} + k_2'n' + k_1\sqrt{n} + k_2 n \quad (3)$$

where C the total inventory cost (although the cost is an economic term, here is used by the meaning of time consumption), n' the size of sample of the first and n of

the second phase, $k_1' = c_w' \sqrt{A}$ where c_w' the cost of walking a unit distance and A is the inventory area corresponding to c_w' square units, $k_2' = c_1' + c_2'$ where c_1' and c_2' are the costs of establishing and callipering a plot respectively, $k_1 = c_w \sqrt{A}$ where c_w the cost of walking a unit distance for finding out the n plots and k_2 the volume measurement cost of a plot.
Suppose that $\bar{Y}_{lr}$ is approximately distributed normally, we require a double sample of which the half 95% confidence interval will not exceed a value $[E \cdot \bar{Y}_{lr}]$, where E is a fraction of $\bar{Y}_{lr}$:

$$1/2 \text{ (confidence interval)} = Z_{0,975} \cdot \sqrt{\text{var}(\bar{Y}_{lr})} = E \cdot \bar{Y}_{lr}$$

$$\Rightarrow \text{var}(\bar{Y}_{lr}) = \frac{E^2 \cdot \bar{Y}_{lr}^2}{Z_{0,975}^2} \qquad (4)$$

The optimization for the forest volume estimation was carried out by minimizing the cost function (formula 3) assuming the variance as fixed (formula 4). The n' and n are determined by applying the Lagrange multiplier method (De Vries, 1986). By this method, finally, the couple of optimum values is chosen by a number of couples (n', n), the amount of which depends on the degree of the under study equation (the course of the method applied is shown analytically in the Annex).

RESULTS - DISCUSSION

The research was conducted at the University forest of Pertouli, in the block "Vathy" 192,25 ha. It is almost covered entirely with *Abies borisii regis Matf.* The trees had a diameter of 13-87 cm and the calculation of their volume over bark (main trunk) was done by Huber's type. The sampling plots had an area of 0,05 ha and the necessary information was taken by preliminary samples of random plots 30 (large) and 20 (small). The cost coefficients of the (3) were taken by previous research (Stamatellos, 1991), and all costs refer to three-member crew minutes.
If $E = 0,10$ from equation 8 (see Annex) and by the help of MATLAB package (Moler, C. et al., 1987) the couple of optimum values was chosen $(n'_{opt}, n_{opt}) = (101, 42)$. The selection from more value couples was based on the following two requirements: 1) $n' > n > 30$ and 2) smaller cost. In **table 1** the volume per ha, the 95% confidence interval, the percentage sampling error and the cost of the plot-double and the respective plot sampling as well are shown.
The deviation of volume per ha is 3,33%, the sampling error increased per 10,45% (from 6,22 increased to 6,87) and the cost reduced per 7,35%.
In forest inventories the volume is measured with a relative difficulty and requires some time while the basal area is measured easily and quickly; consequently, the

usage of double sampling against other simple sampling designs would require an adequately smaller cost (Matis, 1992). An investigation of the total cost of the plot-double and the plot sampling if the volume measurement cost increases is given by **table 2**.
As shown by above table the decrease of cost in the plot-double against plot sampling becomes 20,85, 28,71 and 33,85% when the volume measurement cost is doubled, tripled or quadrupled in relation to the cost if the volume measurement is done by Huber's type (the calculation cost of trees volume by Huber's Type is one of the smallest costs given than only one tree diameter is used for its calculation). The decrease rate of total cost is greater in the case of doubling the volume measurement cost.
A combined investigation of error and cost changes may be conducted by the sensitivity analysis. The sensitivity analysis in sampling can be understood as a procedure for generating and interpreting different design alternatives (Scott and Kohl, 1993). In the sensitivity analysis which was conducted, the ratio $\xi = n'/n$ and the volume measurement cost were used as input variables and more specifically, the ξ changed per ±10 and ±20% from the ξ_{opt} and the volume measurement cost was increased per 2,3 and 4 times of the cost if volume measurement is done by Huber's type (**table 3**).
From Table 3 is shown that the per cent increase of percentage error of plot-double in relation to the respective plot sampling is becoming smaller: 1) if $n'_{opt} = 101$ and $n > n_{opt} = 42$ and 2) if $n_{opt} = 42$ and $n' < n'_{opt} = 101$, which is justified by the variance formula (2). The smaller increase is taking place in the couple of samples $n = n_{opt} = 42$ and $n' = 81$, a fact of special interest if during the selection of alternative designs more attention is given to the sampling error. In respect to the cost, it is most interesting its greater decrease which takes place

in the following two cases: 1) $n' = 101$ and $n = 35$ and 2) $n = 42$ and $n' = 121$. In the first case, the decrease of cost is taking place only on the plot-double sampling by reducing the n (the greater decrease is obviously taking place in the smaller n). In the second case, as long as the n' is increasing an increase of cost is taking place in both samplings with a greater increase in plot sampling. Subsequently, the bigger cost reduction takes

Table 1 ***Volume, 95% confidence interval, percentage sampling error and the cost of plot-double and plot sampling.***

Sampling	Volume		95% confidence interval	Sampling error		Cost	
	m³/ha	Difference %		%	Difference %	Crew minutes	Difference %
Plot-double	231,19		200,06-262,32	6,87		4912,48	
		3,33			+10,45		–7,35
Plot	223,75		196,47-251,03	6,22		5302,18	

Table 2 ***The total costs of plot-double and plot sampling when the volume measurement cost increases.***

Sampling	Total costs							
	Volume measurement by Huber's type		Double volume measurement cost[a]		Triple volume measurement cost[a]		Quadruple volume measurement cost[a]	
	Crew minutes	Difference %	Crew minutes	Difference %	Crew minutes	Difference %	Crew minutes	Difference %
Plot-double	4912,48		5705,02		6497,56		7290,10	
		–7,35		–20,85		–28,71		–33,85
Plot	5302,18		7208,04		9113,91		11019,78	

a: In relation to the cost if volume measurement is done by Huber's type.

Table 3 ***Percentage errors (% SE) and total costs (C1) of plot-double and plot sampling for various combinations of sample sizes and volume measurements costs.***

Error (%), cost and % difference of two samplings		Sizes of samples							
		n' = 101				n = 42			
		n:				n':			
		35	38	47	53	81	91	111	121
%SE	Plot-double	6,94	7,02	6,70	6,80	7,29	7,04	6,61	6,48
	Plot	6,22	6,22	6,22	6,22	7,12	6,57	5,97	5,85
	Difference %	11,58	12,86	7,72	9,32	2,39	7,15	10,72	10,77
$C_1^{a,j}$	Plot-double	4753,59	4821,99	5024,6	5157,97	4387,21	4653,37	5165,54	5413,37
	Plot	5302,17	5302,17	5302,17	5302,17	4481,91	4895,57	5702,74	6098,07
	Difference %	–10,35	–9,06	–5,23	–2,72	–2,11	–4,95	–9,42	–11,23
$C_2^{b,j}$	Plot-double	5414,04	5539,05	5911,5	6158,08	5179,75	5445,91	5958,08	6205,91
	Plot	7208,04	7208,04	7208,04	7208,04	6010,38	6612,74	7797,31	8381,34
	Difference %	–24,89	–23,15	–17,99	–14,57	–13,82	–17,65	–23,59	–25,96
$C_3^{d,j}$	Plot-double	6074,49	6256,11	6798,39	7158,19	5972,29	6238,45	6750,62	6998,45
	Plot	9113,91	9113,91	9113,91	9113,91	7538,85	8329,91	9891,88	10664,60
	Difference %	–33,35	–31,36	–25,41	–21,46	–20,78	–25,11	–31,76	–34,38
$C_4^{e,j}$	Plot-double	6734,94	6973,17	7685,28	8158,30	6764,83	7030,99	7543,16	7790,99
	Plot	11019,78	11019,78	11019,78	11019,78	9067,32	10047,08	11986,45	12947,88
	Difference %	–38,88	–36,72	–36,26	–25,97	–25,39	–30,02	–37,07	–39,83

a. The volume measurement was conducted by Huber's formula.
b. The volume measurement cost is double compared to that of Huber's.
d. The volume measurement cost is triple compared to that of Huber's.
e. The volume measurement cost is quadruple compared to that of Huber's.
j. In crew minutes.

place in the bigger n'. In both cases the couple of samples $n' = 121$ and $n = 42$ seems to provide the bigger cost reductions at various levels of volume measurement cost. By **table 3** results that it has also an advantage on the sampling error (10,77 against 11,58%).

CONCLUSIONS

The forest volume was estimated by plot-double sampling. The estimation was done by minimizing a composite cost function, which also included the walking cost of second phase plots. The optimum sizes of samples were found to be $n'_{opt} = 101$ and $n_{opt} = 42$. In comparison to the respective plot sampling there has been a deviation of volume per ha 3,33%, an increase of % error per 10,45% and a decrease of cost per 7,35%. When volume measurement cost increased a significant reduction of the total cost of plot-double sampling against plot sampling took place. ●

REFERENCES

Cochran W.G. (1977) - Sampling techniques. 3rd ed. John Wiley & Son, New York, 428 pp.

De Vries P.G. (1986) - Sampling for Forest Inventory. Springer, Berlin, 399 pp.

Matis K.G. (1992) - Forest Sampling. Aristotle University of Thessaloniki Press, Thessaloniki, 253 pp (in Greece).

Moler C., Little J. and Bangent S. (1987) - PC-MATLAB, version 3,0-PC. The MathWorks, Inc., MA.

O'Regan W.G. and Arvanitis L.G. (1966) - Cost effectiveness in forest sampling, For. Sci., 12 (4): 406-414.

Schreuder H.T., Gregoire T.G. and Wood G.B. (1994) - Sampling methods for multiresource forest inventory. John Wiley & Son, New York.

Scott C.T and Kohl M. (1993) - A method of comparing sampling designs alternatives for extensive inventories. Mitteilungen der Eidgenossischen Forschungsanstalt fur Wald, Schnee und Landschaft 68, 1: 3-62.

Stamatellos G.S. (1991) - Research of forest volume estimation possibilities with two-stage sampling designs. Ph. D. thesis. Dept. of For. and Nat. Env., Thessaloniki, 108 pp. (unpubl. in Greece).

ANNEX

Must be minimized the equation

$C = k'_1 \sqrt{n'} + k'_2 n' + k_1 \sqrt{n} + k_2 n$ under the restriction that

$$\Phi\,(0) = s_y^2 \left(\frac{1}{n} - \frac{n' - n}{nn'} \rho^2 \right) - \frac{\bar{Y}_{lr}^2}{Z_{0,975}^2} \cdot E^2 = 0 \qquad (1)$$

The Lagrange function is:

$$G = C + \lambda\Phi = k'_1 \sqrt{n'} + k'_2 n' + k_1 \sqrt{n} + k_2 n + \lambda \left[s_y^2 \left(\frac{1}{n} - \frac{n' - n}{nn'} \rho^2 \right) - \frac{\bar{Y}_{lr}^2}{Z_{0,975}^2} \cdot E^2 \right] \qquad (2)$$

and

$$\left. \begin{aligned} \frac{\partial G}{\partial n'} &= \frac{k'_1}{2\sqrt{n'}} + k'_2 - \lambda s_y^2 \frac{\rho^2}{n'_2} = 0 \\ \frac{\partial G}{\partial n} &= \frac{k'_1}{2\sqrt{n}} + k_2 - \lambda s_y^2 \left(\frac{1 - \rho^2}{n^2} \right) = 0 \end{aligned} \right\} \qquad (3)$$

By (3) $\Rightarrow k'_1 (n')^{\frac{3}{2}} + 2\, k'_2\, n'_2 = [k_1 (n)^{\frac{3}{2}} + 2 k_2 n^2] \left(\frac{\rho^2}{1 - \rho^2} \right)$ (4)

It put $n' = \xi n$, $\alpha_1 = \dfrac{k'_1}{\rho_2 / (1 - \rho^2)}$, $\alpha_2 = \dfrac{2\, k'_2}{\rho^2 / (1 - \rho^2)}$,

$\alpha_3 = k_1$ and $\alpha_4 = 2k_2$

the (4) becomes: $(\alpha_1 \xi^{\frac{3}{2}} - \alpha_3) n^{\frac{3}{2}} + (\alpha_2 \xi^2 - \alpha_4) n^2 = 0$

from which it results the: $n = \dfrac{(\alpha_1 \xi^{\frac{3}{2}} - \alpha_3)^2}{(\alpha_2 \xi^2 - \alpha_4)^2}$ (5)

By putting $b = (\bar{Y}_{lr}^2 / Z_{0,975}^2) \cdot (E^2 / s_y^2)$ and $n' = \xi n$ the (1) gives

$$\xi = \frac{\rho^2}{bn - (1 - \rho^2)} \qquad (6)$$

and by replacing of (5) it results

$$- \alpha_2^2 (1 - \rho^2)\, \xi^5 + (\alpha_1^2 b - \alpha_2^2 \rho^2)\, \xi^4 + 2\alpha_2 \alpha_4 (1 - \rho^2)\, \xi^3 - 2\alpha_1 \alpha_3 b \xi^{\frac{5}{2}} + 2\alpha_2 \alpha_4 \rho^2 \xi^2 + [\alpha_3^2 b - \alpha_4^2 (1 - \rho^2)] \xi - \alpha_4^2 \rho^2 = 0 \qquad (7)$$

Putting $\alpha_1 = 35{,}9862$, $\alpha_2 = 3{,}9598$, $\alpha_3 = 47{,}4500$, $\alpha_4 = 37{,}7400$, $\rho = 0{,}9336$ and $b = 0{,}01237$ the (7) becomes

$$- 2{,}013\xi^5 + 2{,}353\xi^4 + 38{,}377\xi^3 - 42{,}245\xi^{\frac{5}{2}} + 260{,}509\xi^2 - 155{,}030\xi - 1241{,}427 = 0 \qquad (8)$$

The (8) is an equation of 5th degree which may be solved with arithmetic methods. In our case the mathematical package MATLAB (Moler et al., 1987) was used.
By the values of ξ and with the help of (5) and $n' = \xi \cdot n$ the n and n' are found.