5-1-1958

Estimation of an Index of Site Productivity

George M. Furnival

Richard N. Rosett

COWLES FOUNDATION DISCUSSION PAPER NO. 52

Note: Cowles Foundation Discussion Papers are preliminary materials circulated privately to stimulate private discussion and critical comment. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

Estimation of an Index of Site Productivity

by

George M. Furnival and Richard N. Rosett

May 19, 1958

Estimation of an Index of Site Productivity

by

George M. Furnival and Richard N. Rosett

In forestry an index of the productive capacity of a site is constructed as follows: A relationship between the height of trees (H), age (A), and various soil factors is estimated. This is often of the form

$$(1)\quad \log H = a_o + a_1 A^{-1} + a_2 S_1 + \dots + a_{n+1} S_n + \epsilon_1$$

The index for a particular site is obtained by assigning an arbitrary value (usually 50) to A and inserting the values of the soil factors measured on the site. Thus

$$(2)\quad SI = a_o' + a_2 S_1 + \dots + a_{n+1} S_n$$

where SI is site index and

$$a_o' = a_o + .02a_1$$

Equation (1) sometimes includes a measure of stand density because tree height, for many species, is related to stand density.

$$(3)\quad \log H = b_o + b_1 A^{-1} + b_2 S_1 + \dots b_{n+1} S_n + b_{n+2} D + \epsilon_3$$

Density is not used to obtain the index for a particular site because it is not possible to predict exactly what density on the site will be, for instance, at the end of fifty years. The advantage of including density in estimating the relationship from which the index is obtained is that it reduces the unexplained variance and increases the sensitivity of the tests of soil factor coefficients. Unfortunately, the usual measures of stand density are often not independent of soil factors and age so that the inclusion of density in (3) and its exclusion from the predicting equation may result in biased estimates of site index. It is possible, however, to obtain both a sensitive test of the coefficients and unbiased site indices.

Assume that the relationship between density, age, and soil factors is

(4) $D = c_o + c_1 A^{-1} + c_2 S_1 + \dots + c_{n+1} S_n + \epsilon_4$

Substituting (4) in (3) we obtain

(5) $\log H = d_o + d_1 A^{-1} + d_2 S_1 + \dots + d_{n+1} S_n$

$$+ d_{n+2}(D - c_o + c_1 A^{-1} + c_2 S_1 + \dots + c_{n+1} S_n) + \epsilon_3$$

Estimates of the coefficients in the system consisting of (4) and (5) can be obtained by maximizing the function

(6) $\Phi(\hat{d}_o, \hat{d}_1, \dots, \hat{d}_{n+2}, \hat{c}_1, \dots, \hat{c}_{n+1}, \hat{\sigma}_3^2, \hat{\sigma}_4^2)$

$$= - \frac{N}{2} \log 2\pi - N \log \hat{\sigma}_3 \hat{\sigma}_4 - \frac{1}{2\sigma_3^2} \sum_{i=1}^{N} \epsilon_{3i}^2 - \frac{1}{2\sigma_4^2} \sum_{i=1}^{N} \epsilon_{4i}^2$$

where N is the number of observations.

Taking the partial derivatives with respect to the parameter estimates we obtain a system of $2n + 5$ equations.

(7)
$$\frac{\partial \Phi}{\partial \hat{d}_i} = 0 \qquad i = 0, 1, \dots, n+2$$
$$\frac{\partial \Phi}{\partial \hat{\sigma}_3} = 0$$

(8)
$$\frac{\partial \Phi}{\partial \hat{c}_i} = 0 \qquad i = 0, 1, \dots, n+1$$
$$\frac{\partial \Phi}{\partial \hat{\sigma}_4} = 0$$

Observing the zero restrictions in (7), (8) reduces to the system of equations from which the c's alone can be estimated. Similarly, (7) reduces to a system from which the d's alone can be estimated and the estimates which are obtained by solving (7) are exactly those which would be obtained from estimating (1). The estimate of σ_3^2 is identical with that which would have been obtained from estimating (3).

If one were to estimate (1) the coefficients might seem insignificant. If, however, it is possible to account for part of the unexplained variance by including in the relationship some variable uncorrelated with variables in the predicting equation (2), the estimated coefficients might actually be significantly different from zero. To the extent that D is correlated with age and soil factors, it is unsatisfactory for this purpose. But the variations in D due to other factors represented by ϵ_4 can be used profitably to obtain a better estimate of the worth of the predicting equation.

For the purpose of estimating the site index for a particular site, it would be useful to know what the value of ϵ_4 will be when the stand is fifty years old. Since this is not now observable, one would not use ϵ_4 in estimating the index. Since ϵ_4 is uncorrelated with soil factors and age, it can be dropped without introducing bias into the estimate of the index. It should be noted however that any variable uncorrelated with age and soil factors could be used in the same way. Suppose, for example, that some soil factor is known for an existing sample of sites, but is too expensive to be measured for the purpose of estimating the index for a particular site. Such a factor could be used just as D was used in the example. In predicting, this factor could be used or not, depending on whether the expense of measuring it was justified.

A Numerical Example

The example that follows is drawn from data collected by members of the Delta Research Center Southern Forest Experiment Station in the course of an investigation of soil factors influencing the site index of sweet gum (Liquidambar styraciflua). More than 40 soil variables were measured in the study, but for the purposes of the example only two will be considered here: the clay content and the level of potassium, both taken 36-48 inches below the surface.

An equation of the form specified by (1) was fitted to the data and the results were as follows:

(11) $\log H = 2.1409231 - 6.2587\ A^{-1} - .00202711\ C + .00014790\ K$

where H and A have been specified and C and K are percent clay and potassium in parts per million respectively. The analysis of variance appropriate for the equation is given below:

Source	Degrees of freedom	Sums of squares	Mean Square	F Ratio
A^{-1}	1	18055		
$C;A^{-1}$ fixed	1	44829		35.27
$K;C$ and A^{-1} fixed	1	7206		5.67
Residuals	58	73714	1271	

Upon adding basal area (a measure of stand density) to the model as in (3), the following equation was obtained:

(12) $\log H = 1.893325 - 5.8350\ A^{-1} - .00205917\ C + .00015315\ K + .10964\ B$

where $\log H$, A^{-1}, C and K have been defined and B is the cross-sectional area of the tree stems 4.5 feet above ground expressed in square feet per acre. The analysis for the expanded model is as follows:

Source	Degrees of freedom	Sums of squares	Mean square	F Ratio
A^{-1}	1	18055		
C; A^{-1} fixed	1	44829		40.39
K; A^{-1} and C fixed	1	7206		6.49
B; K, A^{-1} and C fixed	1	10454		
Residuals	57	63760	1110	

Following the procedure outlined in this paper, relationship (11) would be used to estimate site index but the analysis appropriate for testing the significance of the soil variables is that following equation (12).