

The Weighted Least Squares Method Affects the Economic Rotation Age of Loblolly Pine - Two Planting Density Scenarios

Curtis L. VanderSchaaf^{*1}, Yanshu Li² and David B. South³

¹Forest Modeler, Resource Assessment, Minnesota Department of Natural Resources, Grand Rapids, MN, 55744, USA

²Forest Economist, Texas Forest Service, 301 Tarrow, Suite 364, College Station, TX 77840-7896, USA

³School of Forestry and Wildlife Sciences, Auburn University, AL 36849-5418, USA

Abstract: Managers are typically interested in using equations that provide accurate volumes for large, merchantable trees while modelers are often more interested in equations that meet statistical assumptions. Therefore, modelers often use weighted (WLS) rather than ordinary least squares (OLS) to fit individual tree volume equations since the random error variance is not constant. Since OLS and WLS produce different parameter estimates, perhaps the method chosen will impact the predicted optimum economic rotation age. To address this concern, individual tree volumes were obtained from plots established in an unthinned loblolly pine plantation in southeastern Arkansas. Parameters of the combined-variable function were then estimated using OLS and WLS for the same dataset. Stand-level projections and predicted diameter distributions for two planting densities (748 and 2,692 seedlings per hectare) were then obtained from a growth and yield model (PTAEDA 3.1) and individual tree volumes were predicted using the two parameter estimation approaches. When conducting economic analyses, we found the optimum rotation age between OLS and WLS can differ by as much as 11 yrs (i.e. 10 yr clearcut vs 21 yr clearcut).

Keywords: Models, ordinary least squares, *Pinus taeda* L., plantations.

INTRODUCTION

When developing equations to be used by foresters, the statistical method of weighted least squares (WLS) is commonly used to account for non-constant variance. Some examples include understory vegetation across a range of stand densities [1] or individual tree volumes as a function of diameter and height [2-4]. WLS models are typically used because parameter estimates have lower standard errors than estimates generated using ordinary least squares (OLS). From a statistical perspective, a low standard error is preferred.

From a theoretical aspect, WLS is always used to estimate parameters. OLS (where the weight is constant for all observations) is simply a special case of WLS. When someone decides to use OLS, they assume the error covariance-variance structure is $\sigma^2\mathbf{I}$ (i.e. the variance is assumed constant). Conceptually, a weight is applied to all observations, but when the weight is constant for all data points, in practice it can be ignored.

When estimating parameters using OLS, equal weight is applied to all observations because the random error variance is assumed constant across the domain of the regressor(s). From a statistical theory perspective, if this assumption is assumed but is not true, using OLS to estimate parameters produces unbiased parameter estimates but the estimates are inefficient [5, pgs. 96-97]. Inefficiency results because all

observations should not equally influence the parameter estimates [6; pg. 279]. When the variance is non-constant, from a statistical theory perspective, WLS should be used to obtain more efficient parameter estimates. In this context, the word "efficient" refers to the fact that across repeated random samplings of the same size, WLS parameter estimates will, on average, be closer to the population parameter than the OLS parameter estimates [7; pgs. 366-368, pg. 417, 8]. It should be understood though that for any one observed dataset from the population, the OLS estimate for any parameter may in fact be closer to the parameter value than the WLS estimate.

Parameter Estimation Equations

When estimating parameters of a linear regression model using OLS (variance is assumed constant) and matrices, the following equation is used:

$$\hat{\underline{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \underline{\mathbf{y}} \quad (1)$$

where:

$\hat{\underline{\beta}}$ -- are parameter estimates ($k \times 1$), where k is the number of regressors in the model plus one for the intercept,

$\underline{\mathbf{y}}$ -- are observed values of the dependent variable ($n \times 1$), where n is the number of observations, and

\mathbf{X} -- regressor matrix ($n \times k$) consisting of a column of 1s and the observed values of the regressor(s).

Address correspondence to this author at the Resource Assessment, Department of Natural Resources, 413 SE 13th Street, Grand Rapids, MN 55744, USA; Tel: (218) 327-4449, Ext. 247; E-mail: curtis.vanderschaaf@state.mn.us

characteristics). Individual tree outside-bark total volumes were obtained using the Grosenbaugh height accumulation method [10] – thus volumes were obtained without harvesting the trees. By this method, volumes are estimated from measurements of heights to the occurrence of diameters in a diminishing arithmetic progression up the stem. Volume is computed using these data and published coefficients.

At the time of study establishment (at age 27 yr), there were 156 trees. A total of 431 individual tree volume observations were obtained across all measurement ages. Site index (base age 25 yr) was determined to be near 18.9 m.

Table 1. Summary Statistics of Observations from Trees Used to Estimate Parameters of Equations [4] and [5], Number of Trees at the Initial Measurement Period at Age 27 yr was 156. Where: n is the Number of Individual Tree Observations, Min is the Minimum, Mean is the Arithmetic Mean, and Max is the Maximum Value

	n	Min	Mean	Max
dbh (cm)	431	7.9	26.7	46.7
Ht (m)	431	5.9	19.7	29.1
Volume (m ³)	431	0.03	0.68	3.09

Model Development and Parameter Estimation

The combined-variable function can be expressed as:

$$Vol_i = \beta_0 + \beta_1 D_i^2 Ht_i + \varepsilon_i \tag{4}$$

where:

- Vol_i -- individual tree total cubic foot outside-bark volume,
- D_i -- individual tree outside-bark diameter at breast-height (in.),
- Ht_i -- individual total tree height (ft),
- ε_i -- random error for an individual tree, assumed to follow a Gaussian distribution (0, σ_j²), where σ_j² is the variance of Vol at a particular D²Ht, and
- β₀, β₁ -- parameters to be estimated.

When using OLS to estimate parameters, σ_j² is assumed constant across all observations. With WLS, σ_j² is assumed to vary systematically with changes in the regressor, denoted by the letter *j*. To estimate parameters using WLS, a variance stabilizing transformation can be used to obtain a constant variance across the domain of the regressor [5; pgs. 226-228]. After the variance stabilizing transformation, OLS can be used to estimate parameters. It is assumed that the variance of total cubic foot volume can be described by a power function of D²Ht to the second degree (i.e. 2, 3, 4, 9, pgs. 21-24, [11]), or:

$$Var(Vol) \propto \sigma^2 (D^2Ht)^2$$

Therefore, when dividing both sides of equation (4) by D²Ht, the variance of volume can be made constant:

$$Var\left(\frac{Vol}{D^2Ht}\right) = \frac{1}{(D^2Ht)^2} Var(Vol) = \frac{1}{(D^2Ht)^2} \sigma^2 (D^2Ht)^2 = \sigma^2$$

In this case, $\frac{1}{(D^2Ht)}$ is the variance stabilizing transformation. To obtain WLS parameter estimates, equation (4) was modified by dividing both sides by D²Ht producing:

$$\frac{Vol_i}{D_i^2 Ht_i} = \beta_0 \frac{1}{(D_i^2 Ht_i)} + \beta_1 + \varepsilon_i \tag{5}$$

where:

All variables as previously defined.

For each parameter estimation method, Proc MODEL of the SAS Institute [12] was used to obtain parameter estimates.

Economic Analyses

To determine the potential economic impacts of using WLS to estimate tree volumes, a distance-dependent individual tree model was used to estimate stand development [13]. Output from PTAEDA 3.1 (e.g. number of trees by diameter class and average height of a particular diameter class for a certain age) was used to conduct economic analyses based on the OLS and WLS estimated combined-variable individual tree volume equations. Input variables included a site in the Coastal Plain physiographic region with a site index of 21.3 m (base age 25 yr); site preparation was chop, burn, and bed, with a first-yr herbaceous weed control treatment. The drainage class selected was well-drained and no fertilization treatments were conducted. This regeneration scenario is representative of commonly used scenarios but results will vary depending upon input variables. Two planting densities were simulated: 748 seedlings per hectare (SPH) with 4.9 m rows (2.7 m within rows) and 2,692 SPH using 3.0 m rows (1.2 m within rows).

To obtain merchantable estimates of individual tree volume, we combined the OLS and WLS estimated individual tree total cubic-foot volume equations with a merchantable volume ratio equation presented by [14]. The following inside-bark equation was used:

$$Vol_{Merch} = Vol e^{-0.91505 \left[\frac{d^{4.93352}}{D^{4.60614}} \right]} \tag{6}$$

where:

- d* -- specified upper stem outside-bark diameter limit (in.),
- Vol_{Merch} -- inside bark cubic foot volume to *d*, and all other variables as previously defined.

Equations (4) and (5) were used to estimate total cubic foot volume while equation (6) was used to estimate merchantable volumes to upper stem diameters. For this

analysis, we assume all trees that meet minimum D specifications were merchantable. Minimum D for pulpwood, chip-n-saw and sawtimber were 10.2, 22.9 and 30.5 cm, respectively. For pulpwood, an upper stem outside-bark diameter limit of 5.1 cm was employed; for chip-n-saw, the upper stem outside-bark diameter limit was 10.2 cm; and for sawtimber, the upper stem outside-bark diameter limit was 20.3 cm. For chip-n-saw and sawtimber size trees, all volume above either a 10.2 cm or an 20.3 cm top, respectively, was classified as pulpwood (to a 5.1 cm top). It should be noted in growth and yield modeling that due to the costs and time involved, it is common to utilize volume equations developed using data from vastly different populations than the data used in fitting tree or stand development equations.

For the economic analyses, an interest rate of 6 percent was used and rotation length varied from 10 to 35 yr. For this analysis, we assumed costs of \$315.99 per hectare for site preparation and \$103.17 per hectare for herbaceous weed control. Establishment costs were obtained from [15] for the Southern Coastal Plain. It was assumed that the diameter of all planted seedlings was 5 mm (at root-collar). Seedling cost was set at \$45 per thousand and each seedling would cost \$0.10 to plant by hand. The annual management cost was \$4.94 per hectare and the annual tax rate was \$4.94 per hectare. Since timber is commonly bought and sold using green weight rather than volume, we converted the estimates of merchantable volume (inside bark) to green weight assuming each cubic foot weighs 65 lbs. Since equation (6) was presented in English units [14], all economic analyses were initially calculated on a per acre basis and then converted to a per hectare value.

As part of a sensitivity analysis, four sets of stumpage values were examined. The base-case included region-wide stumpage values similar to that for November 2009 (<http://www.tmart-south.com/tmart/prices.html>); \$7.75 per ton of pulpwood, \$15 per ton of chip-n-saw, and \$27 per ton of sawtimber. The second set assumed reasonable prices (\$10 - \$19.50 - \$35) and the third set involved optimistic prices (\$20 - \$30 - \$44). The fourth set (\$7.75 - \$19.50 - \$44) checks to see the impacts on rotation age when sawtimber is more than five times that for pulpwood. For example, in some locations there might be a short distance to the nearest sawmill but a long distance to the nearest pulpmill.

RESULTS AND DISCUSSION

The Total Sum of Squares for the error term was lower for the OLS fitted equation (Table 2). This is likely a result

of WLS under-predicting volumes of larger trees (Fig. 1) since the WLS method applied less weight to trees with large volumes. [8] showed the variance stabilizing transformation

$$\left(\frac{1}{D^2Ht}\right)$$

helped to stabilize the variance. Predictions of total cubic meter volume per hectare were also affected by the parameter estimation techniques (Fig. 2). The predicted difference was greater when seedlings were planted 1.2 m apart (2,692 SPH). With the wider-spacing (748 SPH) the volume difference between the OLS and WLS methods was less drastic. At early ages, the WLS volume estimates were greater, at age 25 they were nearly equal, and at older ages the OLS volume estimates were greater. This cross-over is also observed in the wide-spacing BLV plots (Fig. 3).

Since the D²Ht values for the 748 SPH case are generally within the range of 5,000 to 9,000 (Fig. 1), the particular parameter estimation procedure had no effect on the optimum economic rotation age. For both estimation methods the optimum rotation was essentially the same for all four sets of revenues (Fig. 3).

However, for the 2,692 SPH case, the age for the optimum BLV differed for some price scenarios. The biggest difference (i.e. 11 yr) was observed for the third set of revenues (\$20 - \$30 - \$44) where optimum rotation ages were 10 yr (WLS) and 21 yr (OLS). At age 10 yr (where all merchantable trees have D²Ht values smaller than 5,000) the WLS volume estimate was 83% greater than the OLS estimate (Fig. 2). At age 21 yr, the volume estimates for OLS and WLS were more similar. At that age, about 18% of merchantable trees have D²Ht between 5,000 and 9,000, and these larger trees make up 35 to 40% of the total volume estimate (for both methods). However, due to the high pulpwood revenues, and the large yield at age 10 yr for the WLS method, the increase in yield from ages 10 to 21 yr is not great enough to offset the loss in revenue due to discounting (Fig. 3). This occurs despite the additional 159 merchantable cubic meters at age 21 yr (or 90% more than the merchantable volume at age 10 yr).

Since the OLS volume per hectare at age 10 yr is substantially less than the WLS method, the growth per hectare from ages 10 to 21 yr is enough to offset the loss in revenue due to discounting and therefore the optimum rotation age occurs later. For OLS, the increase in merchantable cubic meter per hectare volume from age 10 to 21 yr is 182 cubic meters (i.e. 185% more). For the WLS method, there is a slight peak in BLV at about age 17. This peak results because of greater volumes in the chip-n-saw product class which has greater revenue relative to the

Table 2. Parameter Estimates, Standard Errors, and Sum of Squares Error (SSE) when the Model is Fitted Using Ordinary Least squares and Weighted Least Squares (Number of Volume Observations (n) = 431 and Number of Trees Equals 156). Statistics are Calculated Using English Units

Estimation Method	$\hat{\beta}_0$	Std. error	$\hat{\beta}_1$	Std. Error	Total SSE
Ordinary least squares	-1.68903	0.2757	0.003003	0.000026	4884.3
Weighted least squares	0.284409	0.0442	0.002703	0.000019	6542.8

Fig. (3). The effect of stumpage prices and planting density on bare land value (BLV) using PTAEDA 3.1. Parameters of the combined-variable volume equation were estimated using either weighted least squares (solid lines) or ordinary least squares (dashed lines). The numbers 748 and 2,692 correspond to the number of seedlings planted per hectare. BLV values were initially calculated using per acre values and then converted to per hectare values by multiplying by 2.47.

target tree size. Users must remember that growth and yield models do not provide the same predictions [16] because of factors such as sampling error and personal bias about how trees and stands grow but also because of disagreements about statistical methods thought to best estimate parameters; the structure of models can lead to vastly different “optimum” management scenarios.

For instances where non-constant variance exists and prediction is the overriding concern, OLS may be desirable when the economic or biologically most important observations are those with larger variances. However, when one is interested in significance levels of parameter estimates, WLS will be superior (5, pgs. 96-97). For individual tree volume, we know diameter and height influence volume and our greatest concern is the accurate prediction of volume for the merchantable and most valuable trees, not significance levels of the parameter estimates.

ABBREVIATIONS

BLV = Bare Land Value
 OLS = Ordinary Least Squares

SPH = Stems Per Hectare
 WLS = Weighted Least Squares

REFERENCES

- [1] Mitchell JE, Bartling PNS. Comparison of linear and nonlinear overstory-understory models for ponderosa pine. *For Ecol Manage* 1991; 42: 195-204.
- [2] Cunia T. Weighted least squares method and construction of volume tables. *For Sci* 1964; 10: 180-91.
- [3] Smalley GW, Bower, DR. Volume tables and point-sampling factors for loblolly pines in plantations on abandoned fields in Tennessee, Alabama, and Georgia Highlands. *USDA For Serv Res. Rep* 1968; SO-32.
- [4] Knoebel BR, Burkhardt HE, Beck DE. Stem volume and taper functions for yellow-poplar in the southern Appalachians. *South J Appl For* 1984; 8: 185-8.
- [5] Schabenberger O, Pierce, FJ. *Contemporary statistical models for the plant and soil sciences*. Boca Raton, FL: CRC Press 2002.
- [6] Myers RH. *Classical and modern regression with applications*. Pacific Grove, CA: Duxbury Classic Series 1990.
- [7] Wackerly DD, Mendenhall W III, Scheaffer RL. *Mathematical statistics with applications*. Pacific Grove, CA: Duxbury Advanced Series 2002.
- [8] VanderSchaaf CL, South D. Biological, economic, and management implications of weighted least squares. *South J Appl For* 2009; 33: 53-7.

- [9] Clutter JL, Fortson JC, Pienaar LV, Brister GH, Bailey RL. Timber management: a quantitative approach. Malabar, FL: Krieger Publishing Co 1983.
- [10] Grosenbaugh LR. New tree-measurement concepts: height accumulation, giant tree, taper and shape. USDA For. Serv. Occ. 1954; Paper SO-134.
- [11] Golden MS, Knowe, SA, Tuttle CL. Cubic-foot volume for yellow-poplar in the hilly coastal plain of Alabama. South J Appl For 1982; 6: 167-71.
- [12] SAS Institute. SAS/ETS user's guide, version 6. SAS Institute Inc., Cary, USA, 2003.
- [13] Burkhart HE, Amateis RL, Westfall JA, Daniels RF. PTAEDA 3.1: Simulation of individual tree growth, stand development and economic evaluation in loblolly pine plantations. School of Forest and Wild Research, Va. Polytech Inst. and State Univ 2004.
- [14] Tasissa G, Burkhart HE, Amateis RL. Volume and taper equations for thinned and unthinned loblolly pine trees in cutover, site-prepared plantations. South J Appl For 1997; 21: 146-52.
- [15] Barlow RJ, Smidt, MF, Morse, JZ, Dubois, MR. Cost and cost trends for forestry practices in the South. For Landowner 2009; 68(5): 5-12.
- [16] Buford MA. Performance of four yield models for predicting stand dynamics of a 30-yr-old loblolly pine (*Pinus taeda* L.) spacing study. For Ecol Manage 1991; 46: 23-38.

Received: May 21, 2010

Revised: October 26, 2010

Accepted: November 16, 2010

© VanderSchaaf *et al.*; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.