United States Department of Agriculture

Forest Service

Southern Forest Experiment Station

New Orleans, Louisiana

Research Paper **SO-233**

A Comparison of Tree Volume Estimation Models for Forest Inventory

John F. Kelly and Roy C. Beltz

SUMMARY

Eight different linear regression models were tested for ability to predict timber inventory for four tree species in Northeast Texas. Sample trees were selected for the Forest Survey by the variable plot (prism) method. Each model was tested using two weighting schemes for weighted least squares regression-probability weights and optimal heteroscedasticity-correcting weights. In general, the probability weights performed best for inventory prediction. Spurr's combined formula model, incorporating D^2H, fitted with probability weights made the best predictions.

May 1987

A Comparison of Tree Volume Estimation Models for Forest Inventory

John F. Kelly and Roy C. Beltz

INTRODUCTION

Developing forest inventory estimates often involves predicting tree volumes from only diameter at breast height (d.b.h.) and/or merchantable height. Prediction equations based on these two factors from a small number of intensively measured trees can lead to significantly different inventory estimates. The nationwide Forest Survey is particularly susceptible to these effects because of the interval between successive surveys and the importance of estimating inventory trends.

The principal goal of the Forest Survey at its inception in 1928 by the McSweeney-McNary Act was to estimate the total volume and area of the forest resource. As user demands become more intense, concerns center more on inventory components such as species and d.b.h. classes, and estimates of change, i.e., growth and removals. But these estimates are not nearly as reliable as those for the total inventory since they deal with smaller numbers of trees or relatively small changes in volume. Estimates of change deal with differences in volumes of those surviving trees which were measured in previous surveys and the addition or removal of trees from the inventory. The increasing concerns for reliable estimates of inventory components and volume changes requires that Forest Survey volume estimation techniques be as accurate as possible.

Techniques may differ enough between successive surveys to cloud the change statistics. While procedures are standard, slight variations creep in as survey analysts continually try to improve their methods. Volume equations, for example, may change slightly due to a change in the model or estimation methods. Past volumes are usually recalculated to new standards (equations) for comparison, However, the subtle effects of these volume equations can affect evidence of inventory trends.

Data from the 1985 Forest Survey of east Texas were used to investigate eight linear models and two weighting schemes in predicting timber inventory volumes. Specific objectives of the study were to: (1) evaluate linear models incorporating d.b.h. and merchantable height as explanatory components; (2) evaluate linear models incorporating d.b.h. only; and (3) evaluate two alternative weighting schemes for estimation of linear regression parameters: (a) probability-based weights, which emphasize compatibility with sample selection probabilities, and (b) optimal heteroscedasticity-correcting weights, which emphasizes conformance to a key assumption of the general linear regression model. The basis for evaluations were comparisons of timber inventory predictions with estimates developed from Smalian formula tree volumes and sample probabilities for numbers of trees per acre.

LINEAR REGRESSIONS AND TIMBER INVENTORY SAMPLES

The linear regression model, with parameters estimated by ordinary least squares (OLS) techniques, is often used as a convenient means of estimating tree and inventory volumes. This classical linear regression model involves specific assumptions, some of which may be violated by application to timber inventory situations. Furthermore, effective use depends on the correct specification of the models; i.e., a linear relationship should exist between predicted variables (such as tree volume) and independent variables (d.b.h., height, etc.).

The general linear regression model has the form:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \epsilon_i$$

where Y = dependent variable;
X = independent variable;
β = regression parameters;
ϵ = variation not explained by the regression model (error term).

The six basic assumptions regarding application of the general linear regression model are: (1) the mean of the error term (ϵ) is zero; (2) the variance of the error term is constant; i.e., the data are homoscedastic; (3) the error terms for different observations are uncorrelated; (4) independent variables are non-

John F. Kelly is research forester and Roy C. Beltz is project leader for FIA, Southern Forest Experiment Station, Starkville, MS.

stochastic; (5) no perfect collinearity exists between independent variables (in models containing two or more independent variables); and (6) the number of observations is greater than the number of independent variables.

Assumption 2 is commonly violated when using linear regressions to predict tree volume. The absence of homoscedasticity, i.e., heteroscedasticity, occurs when samples include a range of tree sizes, from small trees to large ones. Since the height and form of large diameter trees may vary considerably, due to various stand and site factors influencing the growth, the volume will likewise vary. Small diameter trees that are shorter and contain less total volume will usually vary less in volume. Tree samples selected to estimate timber inventory for a variety of forest conditions thus may not be homoscedastic because of volume variance attributable to factors not explicit in regression models. Figure 1 illustrates the variance of total tree volume by diameter class of sample trees in this study.

If heteroscedasticity is, in fact, present in a sample of trees, there are some consequences from using OLS techniques to estimate linear regression parameters. Fortunately, the regression coefficients estimated under these conditions are not biased, and their estimates tend to converge on the true parameters as the sample size increases. The variances of the coefficient estimates will, however, be larger than those estimated for homoscedastic samples. Furthermore, the variances for the coefficients calculated using standard OLS methods may be biased, appearing to be smaller than their actual value. The extreme consequence posed by this situation is the rejection of the hypothesis that regression coefficients are equal to zero when the hypothesis is actually true. Barring this potential problem, however, the presence of heteroscedasticity will still allow unbiased estimates of regression coefficients, but these estimates will not have minimum variance.

The use of weighted least squares is a common remedy for heteroscedasticity. This technique involves multiplying the regression equation by a factor (say, Z_i) which equalizes the variance of the error term for the sample data. For instance, if it is determined that the variance of the error term is proportional to the square of an independent variable, then the appropriate weighting factor would be the inverse of the squared variable. A technique for determining correct weighting factors will be discussed later.

Weighting techniques may alter regression results when sample characteristics vary according to the weighting factors. This may occur when the inverse of tree diameter (or some transformation such as diameter squared) is used as a weighting factor. In this case, the influence of large diameter trees in calculating regression parameters will be discounted, while the influence of small diameter trees will be amplified. Since the relationship between volume and predictive variables (diameter, height) may vary depending on tree diameter, the use of weights related to diameter may skew results.

While the use of weights related to the inverse of tree diameter may potentially skew regression results, the actual outcome will depend on the specific relationship between the sample and weights selected. The selection of weights is particularly important for probability-related samples where the number of large trees selected is greater than that obtained from simple random samples. Because of the

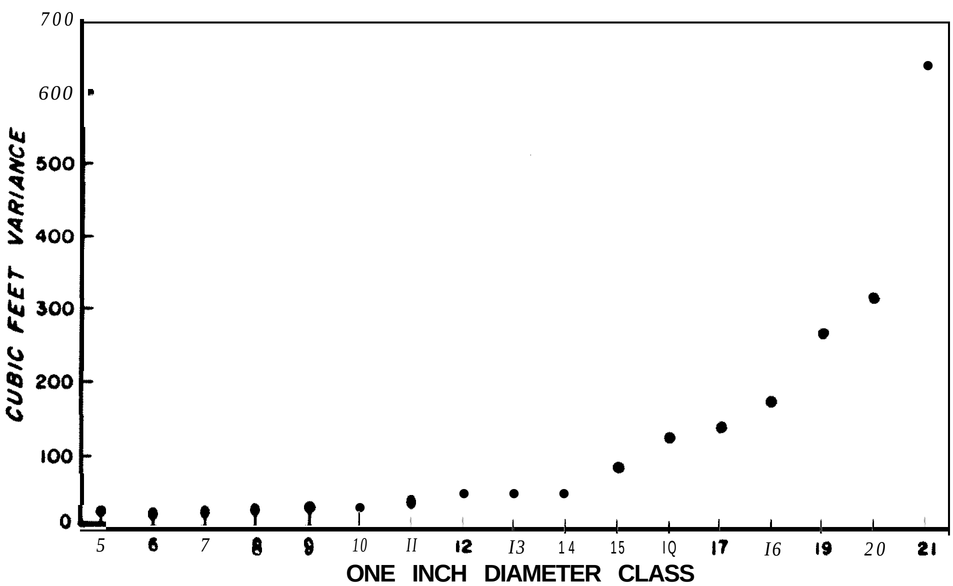

Figure L-Comparison *of tree volume variance. Average variance* ***by diameter class, regression sample.***

emphasis on large trees in a prism sample, the use of compatible weights is necessary to fully realize the advantages of the sample.

In the special case where the weighting variable Z_i is the number of trees per unit area represented by a sample tree, the product of the weight and tree volume (Y_iZ_i) would be the volume per unit area represented by that tree. These weights would then be compatible with the probability of tree selection. For prism samples, the number of trees represented by each sample tree is a linear function of the inverse of the squared tree diameter.

Because of the possible consequences posed by correcting for heteroscedasticity in a prism sample, it is not clear if the procedure is desirable. The weights for correcting heteroscedasticity will likely be quite different from probability weights. Researchers using probability-based homoscedastic samples have not agreed on the desirability of using probability weights. DuMouchel and Duncan (1983) indicated that probability weights may not be desirable in all situations. Holt, Smith and Winter (1980) found that probability weighted regressions provided good results as long as selection probabilities were related to an independent variable in the model. The current problem, however, deals with probability-based heteroscedastic sample data and with procedures desirable for timber inventory estimation.

METHODS

Periodic forest surveys are conducted by the Forest Inventory and Analysis (FIA) unit of the Southern Forest Experiment Station, USDA-Forest Service, as a part of a nationwide forest survey. Plot locations are at the intersections of a three-mile grid. Information was collected for Northeast Texas according to standard procedures (USDA Forest Service 1985). Plots consist of 10 point-samples with point centers systematically placed on a grid 66 feet apart. Sample trees are selected using a 37.5 basal area factor prism. For each tree, diameter is recorded at l-foot stump, d.b.h., mid-saw-log height, saw-log height, mid-bole height, and merchantable height (fig. 2). Merchantable limits are 5 inches d.b.h., and 4 inches top diameter outside bark (d.o.b.) for poletimber. Softwood sawtimber limits are 9 inches d.b.h. and 7 inches top d.o.b.; hardwood sawtimber limits are 11 inches d.b.h. and 9 inches top d.o.b. Top merchantability limits may also be reached where deformities or other defects are present. Bark thickness at d.b.h. is measured and deducted to get d.b.h. inside bark. Upper stem bark values are derived from the ratio of d.b.h.i.b./d.b.h.o.b. according to procedures in STX (Grosenbaugh 1964). Cubic foot volume for each section in each tree is computed using the Smalian formula. These sectional volumes are then summed for each tree to obtain total tree volume; these calculated volumes are referred to as "actual" volumes throughout this paper.

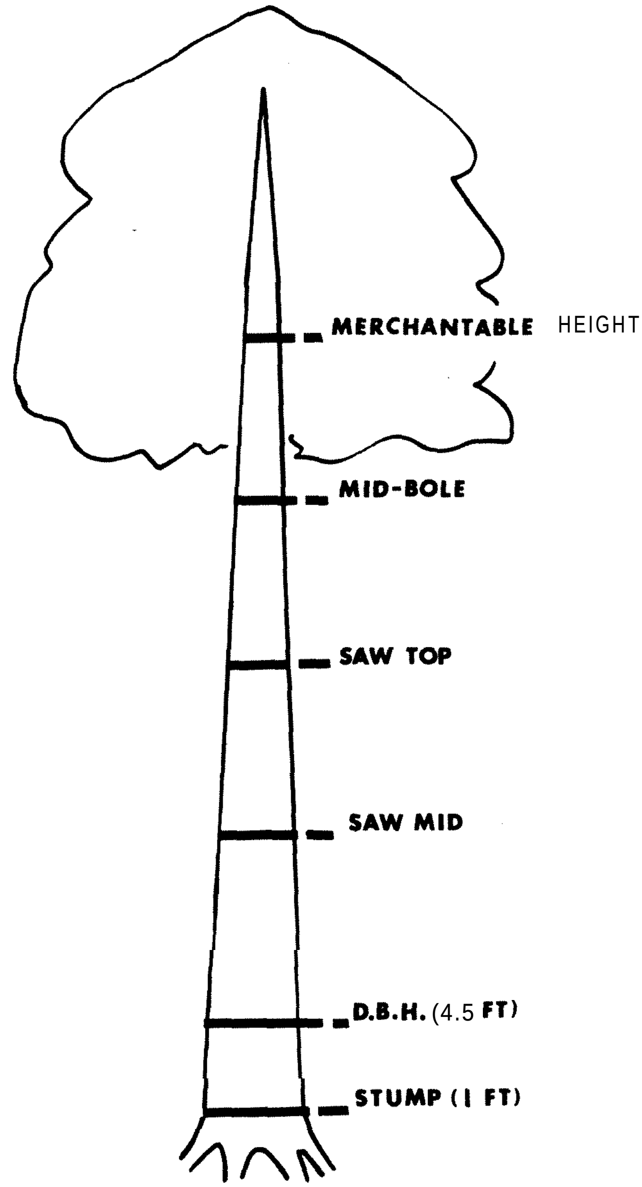

Figure *2.-Measurement points on FIA tally trees.*

Data from 676 forests plots in the Northeast Texas survey region (fig. 3) were screened for four common species—shortleaf pine (*Pinus echinata),* loblolly pine (*Pinus taeda*), sweetgum (*Liquidambar styraciflua*), and post oak (*Quercus stellata* var. *stellata*).

Each tree had to qualify as growing stock[1] and be at least 5 inches d.b.h. and no greater than 28 inches d.b.h. Plots were randomly allocated to one of two

[1]Growing stock is defined as trees that contain, or potentially contain, at least 1 12-foot saw log of minimum grade specification; rough culls, rotten culls, and dead trees do not qualify as growing stock.

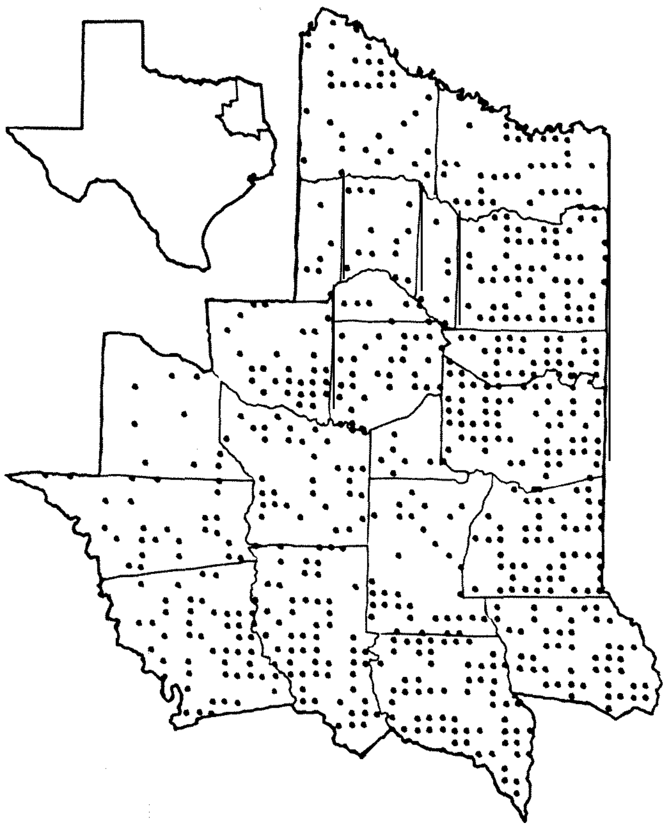

Figure 3.-Plot locations of sample trees, Northeast Texas survey region.

data sets. Qualifying trees in one data set were used to develop the equations. The other set was used to test them. Sample sizes for the two data sets are given in table 1.

Individual tree gross volumes were used to tit 8 linear regressions using (1) probability weights for individual tree selection, and (2) optimal weights for correcting heteroscedasticity. Regression models evaluated were:

1. $CF = \beta_0 + \beta_1 D + \beta_2 D^2 + \beta_3 H + \epsilon$
2. $CF = \beta_1 D + \beta_2 D^2 + \beta_3 H + \epsilon$
3. $CF = \beta_0 + \beta_1 D^2 H + \epsilon$
4. $CF = \beta_1 D^2 H + \epsilon$
5. $CF = \beta_0 + \beta_1 D + \beta_2 D^2 + \epsilon$
6. $CF = \beta_0 + \beta_1 D^3 + \epsilon$
7. $CF = \beta_0 + \beta_1 D^2 + \epsilon$
8. $CF = \beta_1 D^2 + \epsilon$

where

CF = gross cubic foot volume in the merchantable bole;
D = d.b.h. outside bark;
H = merchantable height;
β_i = coefficients to be estimated;
ϵ = error term.

Equations 2, 4, and 8 are identical to 1, 3, and 7, respectively, except for the intercept (β_0) term. Equations l-4 include merchantable height and d.b.h.; equations 5-8 include only d.b.h. Equation 3 is Spurr's combined variable formula, and equation 4 is the constant form-factor equation (Spurr 1952). Equation 5 is a 2nd degree parabolic formula which was used by Cunia (1964).

Table 1.—*Number of trees in sample data sets, Northeast Texas*

	Regression sample	Test sample
Shortleaf pine	1,131	1,059
Loblolly pine	1,133	1,260
Sweetgum	580	618
Post Oak	371	390

Weighted least squares were used to estimate coefficients. Probability weights used were $1/D^2$; the number of trees per acre represented by each tree in a prism sample is a linear function of $1/D^2$. Optimal weights for correcting heteroscedasticity were determined using a maximum likelihood function as discussed below.

The test sample was used to simulate the effect of the equations on timber inventory estimation. Results were compared to Smalian formula volumes calculated from measurements by survey field crews.

DETERMINATION OF OPTIMUM WEIGHTS FOR CORRECTING HETEROSCEDACTICITY

Squares of non-weighted regression residuals plotted against predicted volumes indicated that the residuals were an approximate function of the squared volume. Plots of squared residuals against D^2 and H further indicated that variability was an approximate function of the square of each of these variables, i.e., D^4 and H^2. This would suggest that weights of the form $1/D^4H^2$ and $1/D^4$ would be appropriate for correcting heteroscedasticity.

Meng and Tsai (1986) reported that these weights ($1/D^4H^2$ and $1/D^4$) have been traditionally considered as appropriate by several sources, but the optimum exponent for D, say, A, will vary with different samples.

Following an approach derived by Meng and Tsai (1986) from Box and Cox (1964)[2], the optimum weights in the forms $(1/D^\lambda H)^2$ and $(1/D^\lambda)^2$ were determined for all species and all equations.

Optimum values for the exponent of D (λ) were determined by an iterative process which estimated the likelihood function for different A. Separate iterations were run for each equation; weights of the form $(1/D^\lambda H)^2$ were used for equations l-4, and weights of the form $(1/D^\lambda)^2$ were used for equations 5-8. The

[2]Kmenta (1971) presents a similar, though more general approach.

maximum value for the likelihood function indicated that an optimum A had been found. Values for λ were incremented by .025 until an optimum value was reached; i.e., likelihood functions were estimated for values such as 2.5, 2.525, 2.55, etc.

The log likelihood function for λ is

$$L_\lambda = \frac{-n}{2} SS_\lambda - \lambda \sum_{i=1}^{n} \ln(D_i) + \frac{n}{2}\ln(n) - \frac{n}{2} - \frac{n}{2}\ln 2\pi$$

where L_λ = the maximized log likelihood for A;
SS, = the sum of squares for the error term resulting from weighted regression;
n = the total number of trees in the sample;
ln = the natural logarithm for the variable indicated.

After finding the value for λ providing the maximum L, a confidence interval was established for λ at a probability level of 0.95 (α=.05) as follows (Meng and Tsai 1986):

$$A: L_{\tilde{\lambda}} - L_\lambda < 1/2\, \chi^2_{1df,\ .05}$$

where $L_{\tilde{\lambda}}$ = the maximized log function for the optimum λ;

$\chi^2_{1df,\ .05}$ = the chi-squared distribution for 1 degree of freedom and a = .05.

By following these procedures for all species and equations, the optimum values for A and confidence intervals were produced (table 2). For this analysis the optimum values of A were used for comparisons.

RESULTS

Regression Analysis

The F test and adjusted R^2 regression statistics indicate that all eight equations are effective predictors of volume. The F test indicated that all models were significant at the 95 percent probability level; adjusted R^2 values were at least .88.

Plots of regression residuals (on weighted scales) indicate that the basic assumptions of zero mean for the error term and uncorrelated error terms hold true.

Possible collinearity of D and H in equations 1 and 2 was considered. Although the relationship between D and H approaches linearity, it is not perfectly linear, nor particularly strong. Thus, the estimated regression coefficients for equations 1 and 2 are valid (unbiased).

Other than heteroscedasticity, no other violations of remaining basic assumptions for linear regression models were noted.

There were some coefficients for independent variables not statistically different from zero, all occurring in fitting equation 5 and involving the D variable. The use of probability weights produced 3 of the 4 instances where the D coefficient was statistically insignificant at the .95 probability level. Alternately, 3 of the 4 optimally weighted regressions for equation 5 had D coefficients that were statistically significant.

Negative intercept terms occurred in most regressions for equations 5 and 7. The intercept term for equation 1 (probability weighted) fitted to post oak data was also negative.

Negative signs for D occurred in all regressions for equations 1, 2, and 5. In each of these equations D is paired with D^2; coefficients for the latter variable were positive in all cases.

Comparison of Predicted Volumes for the Test Sample

The test sample was used to compare predicted volumes with actual volumes calculated from Smalian's formula. Three different criteria were used-the occurrence of unreasonable values (i.e., negative or zero tree volumes), the prediction of total inventory volumes within a 1 percent standard, and the prediction of one-inch diameter class volumes within a 10 percent standard. Inventory volumes were estimated by first calculating individual tree volumes, using regression coefficients, then obtaining the product of tree volume and the number of trees per acre for the 37.5 basal area factor prism sample. Finally, values

Table *2.-Optimum d.b.h. weighting values and confidence intervals* for linear regressions, ***by species and regression model***

Linear regression model	Weight variable	Optimum values and confidence intervals[1] for λ			
		Shortleaf pine	Loblolly pine	Sweetgum	Post oak
1	$(1/D^\lambda H)^2$	2.2(2.05-2.325)	2.325(2.225-2.425)	2.325(2.175-2.475)	2.575(2.375-2.775)
2	$(1/D^\lambda H)^2$	1.925(1.775-2.05)	1.95(1.85-2.075)	2.175(2-2.35)	2.5(2.325-2.675)
3	$(1/D^\lambda H)^2$	2.075(1.975-2.175)	1.975(1.9-2.075)	2.35(2.2-2.5)	2.375(2.175-2.575)
4	$(1/D^\lambda H)^2$	1.95(1.85-2.075)	1.75(1.65-1.85)	2.025(1.875-2.175)	2.3(2.125-2.475)
5	$(1/D^\lambda)^2$	2.375(2.225-2.5)	2.225(2.125-2.35)	2.525(2.35-2.7)	2.45(2.25-2.65)
6	$(1/D^\lambda)^2$	2.525(2.375-2.675)	2.55(2.425-2.65)	2.8(2.625-2.95)	2.7(2.45-2.9)
7	$(1/D^\lambda)^2$	2.325(2.2-2.45)	2.17X2.075-2.275)	2.5(2.3-2.65)	2.45(2.25-2.65)
8	$(1/D^\lambda)^2$	1.65(1.525-1.775)	1.55(1.425-1.65)	2.375(2.125-2.6)	1.875(1.675-2.1)

[1]At .95 probability level.

for appropriate diameter classes and species were summed. All comparisons were made for the four species individually, and for all species (all trees) combined. One-inch diameter classes ranged from 5 to 21 inches with all trees 21-28 inches being placed in the latter class.

Among the probability weighted regressions, negative or zero tree volumes occurred with equations 1 and 2, usually in the 5-inch diameter class. Such values also occurred in the 6- and 7-inch-diameter classes for the pines. No negative volumes were predicted using optimally weighted regressions. For the balance of the analysis, all negative volumes produced from the probability weighted regressions were set to zero.

The ability of the models to predict inventory volume was compared using the test sample. Inventory volumes represented by individual trees were first predicted, then mean volumes calculated both for individual species and the entire test sample. These mean volumes were finally divided by the mean Smalian formula inventory volume (CF *x* trees per acre); the results are displayed in table 3. Instances where the predicted volume was within 1 percent of actual volume are noted.

Among equations including both D and H components, the probability weighted regressions generally performed better than optimally weighted regressions in predicting inventory volume. For the entire test sample (all species combined), equation 3, fitted using probability weights, predicted inventory with an error of only .1 percent; the use of optimal weights produced an error of .6 percent. Equations 1 and 2, fitted with probability weights, predicted inventory for the combined sample within 1 percent error; other regressions produced errors greater than 1 percent.

Among equations using only D as an explanatory component, three—5, 6, and 7-predicted total inventory volume within 1 percent of actual values, using both weighting schemes. Equation 5 (optimal weights) and equation 6 (probability weights) predicted total inventory with no error. Equations 5 and 7 (probability weights) were within .5 percent error.

Analysis of mean predicted inventory volume by diameter class indicates distinct differences among equations. Predicted volumes for individual trees of the test sample, calculated using regression coefficients, were multiplied by the appropriate factor for trees per acre to develop mean inventory volumes for the 17 one-inch diameter classes. These mean volumes were then compared to the respective Smalian formula inventory volumes using a Chi-square test as described by Freese (1960). This test was formulated to estimate ability to simultaneously predict all 17 diameter class inventory volumes within 10 percent of actual volume. The Chi-square statistics, presented in table 4, were calculated as follows:

$$\chi^2_{17df,\ .05} = (1.96^2/.10^2) \sum_{i=1}^{17} [(x_i/\mu_i) - 1]^2$$

where $\chi^2_{17df,\ .05}$ = Chi-square distribution with 17 degrees of freedom, a = .05;
x_i = predicted inventory volume for diameter class i;
μ_i = actual inventory volume for diameter class i.

The critical χ^2 value for these calculations is 27.6; statistics less than this value indicate the model will predict all 17 diameter class volumes with less than 10 percent error, unless a l-in-20 chance of a random event occurs.

Only equation 3, using either probability weights or optimal weights, was able to predict inventory volumes by diameter class with no greater than 10 percent error for all test groups (the 4 individual species and total sample). Equations 4 and 5 also did well relative to other equations, meeting this test for the total sample and at least 2 individual species; equation 4, optimally weighted, met the test for 3 individual species as well as the total sample. Equations 2, 6, and 8 did not predict volumes for any of the test groups within this standard using either probability weights or optimal weights. Equation 1 produced only one instance under each weighting scheme where the 10 percent standard for diameter class volume prediction was met.

Table 3.—*Predicted mean inventory volumes as a proportion of actual mean inventory volumes, by species and type of regression (test sample)*

Equation	Probability weighted regressions					Optimally weighted regressions				
	Entire sample (all trees)	Shortleaf pine	Loblolly pine	Sweetgum	Post oak	Entire sample (all trees)	Shortleaf pine	Loblolly pine	Sweetgum	Post oak
1	.993*	.994*	.979	1.019	1.003*	.938	.938	.919	.982	.942
2	.991*	1.001*	.971	1.011	1.003*	.917	.930	.889	.946	.938
3	1.001*	1.004*	.993*	1.022	.993*	1.006*	1.002*	1.000*	1.030	1.012
4	.973	.994*	.943	.997*	.973	1.017	1.009*	1.006*	1.058	1.025
5	.997*	1.004*	.975	1.023	1.028	1.000*	1.003*	.981	1.025	1.028
6	1.000*	.987	1.003*	1.024	.995*	1.007*	.989	1.004*	1.039	1.040
7	.998*	1.005*	.974	1.023	1.029	.994*	1.002*	.973	1.010*	1.025
8	1.131	1.108	1.127	1.210	1.103	1.033	1.045	1.060	.947	1.017

*Predicted mean within 1 percent of actual means.

EVALUATION OF MODELS AND WEIGHTS

There are tradeoffs between the probability weights and optimal weights used for linear regression in this study which dealt with a prism sample for the purpose of inventory estimation. Probability weights are used for regression in order to realize the advantages provided by emphasizing the selection of large, high-volume trees. Optimal weights for correcting heteroscedasticity will provide estimates of regression coefficients with smaller variances, although this did not appear to be an important advantage in this study because variances were typically very small. The optimal weights in this study, however, did reduce the influence of large trees, when compared to probability weights. Coefficients estimated using both weighting schemes can be expected to represent true population values. Furthermore, the probability weights reduce, but do not eliminate, the effects of heteroscedasticity.

Correct specification of linear regression models requires linear relationships between independent and predicted variables. Although equation 5 may appear to violate this specification requirement (D is not linearly related to CF), the quadratic form of the equation apparently provides a powerful linear transformation for volume prediction where D is the only explanatory component available. One reason for postulating equation 6 was the linear relationship implied between the units of the D^3 and CF variables. Results from the test sample, however, indicate that while equation 6 predicts total inventory volume accurately, it does not perform well for predicting volumes by diameter classes.

One advantage of through-the-origin regression models (equations 2, 4, and 8) was thought to be the avoidance of negative volume prediction. Since equations 3 and ‘I-comparable to equations 4 and 8, but with an intercept term-produced no negative volumes, the use of the through-the-origin models for this purpose was unnecessary. Equation l-comparable to equation 2, but with an intercept term-did produce negative volumes, but only when fitted with probability weights. Equation 2 also predicted negative volumes when fitted with probability weights; these values resulted because of the size of the D coefficient, which was negative. Thus, through-the-origin models would appear either ineffective or unnecessary for this purpose.

Table 5 displays summary test information by indicating (1) the number of test groups where total inventory volumes were predicted within 1 percent of actual volume; and (2) the number of test groups where one-inch diameter class volumes were predicted (simultaneously) within 10 percent of actual volume, as indicated by the Chi-square tests. Test groups are the four individual species, plus the entire combined test sample, i.e., the columns in tables 3 and 4.

In general, it appears that optimally weighted regressions provide no clear advantage over the probability weighted regressions. Although the use of optimally weighted regressions slightly improved predictions for two equations (4 and 7), other equations provided better estimates when fitted with probability weights (1, 2, 3, and 6). Table 5 indicates no difference for the two weighting schemes for equations 5 and 8.

Based on the evidence summarized in table 5, equation 3, fitted with probability weights, is clearly the best performer. Equation 3 also performed well when fitted with optimal weights.

Among models using only D as a component, equation 5, fitted with probability weights, and equation 7 fitted with optimal weights, are the best performers.

Coefficients for equations 3, 5 and 7, derived using both weighting schemes, are presented in table 6.

In summary, results from this study indicate that linear regressions can be an effective tool for inven-

Table 4.—*Chi-square statistics for predicting inventory volume within 10 percent of actual volume for 17 one-inch diameter classes simultaneously at the 95 percent confidence level (a = .05)[1]; by species and type of regression (test sample)*

Equation	Probability weighted regressions: Entire sample (all trees)	Shortleaf pine	Loblolly pine	Sweetgum	Post oak	Optimally weighted regressions: Entire sample (all trees)	Shortleaf pine	Loblolly pine	Sweetgum	Post oak
1	51.5	27.1*	62.6	30.0	43.1	35.2	34.9	59.7	25.3*	31.7
2	199.7	154.5	139.7	102.9	44.4	131.7	111.7	158.8	167.1	45.4
3	11.9*	5.0*	11.7*	19.2*	22.1*	1.8*	5.8*	2.9*	16.4*	17.6*
4	24.2*	18.2*	42.5	33.7	16.0”	18.1*	18.1*	26.7*	52.5	18.1*
5	15.2*	35.4	21.3*	44.9	17.3*	6.2*	32.0	14.2*	59.7	18.3*
6	809.6	463.6	568.1	286.3	300.4	185.6	296.8	191.5	450.4	392.7
7	39.2	52.0	39.5	56.0	18.6*	27.1*	43.7	37.3	48.7	17.4*
8	1435.1	1183.8	1614.0	962.4	369.5	1012.4	1122.7	1580.8	775.9	320.9

[1] $\chi^2_{17df,\ .05} = (1.96^2/.10^2) \sum_{i=1}^{17} [(x_i/\mu_i) - 1]^2$, where x_i = predicted inventory volume; μ_i = actual inventory volume. Note: asterisks (*) indicate statistics less than the critical value of 27.6; these predictions are within the 10 percent limit at the indicated confidence level.

Table 5.—*Number of test groups[1] meeting 1 percent total inventory and 10 percent diameter class volume standards, by type of regression*

Equation	Probability weighted regressions		Optimally weighted regressions	
	Number within 1% total inventory volume	Number within 10% volume standard for diameter class	Number within 1% total inventory volume	Number within 10% volume standard for diameter class
1	3	1	0	1
2	3	0	0	0
3	4	5	3	5
4	2	3	2	4
5	2	3	2	3
6	3	0	2	0
7	2	1	3	2
8	0	0	0	0

[1]Five test groups are included: 4 individual species and all trees combined.

Table 6.—*Coefficients for gross cubic volume equations for probability and optimally weighted regression models 3, 5,* and 7

Species	Regression model, variables[1]	Coefficient estimates		
		β_0	β_1	β_2
Shortleaf pine	$(3p)\beta_0,\beta_1(D^2H)$	0.2578117	0.002676204	....
	$(3o)$	0.1971193	0.002684152	....
	$(5p)\beta_0,\beta_1(D),\beta_2(D^2)$	-2.52757	-0.2882	0.1898756
	$(5o)$	-1.05001	-0.619361	0.2063557
	$(7p)\beta_0,\beta_1(D^2)$	-3.96955	0.1771888	....
	$(7o)$	-3.63574	0.1734781	....
Loblolly pine	$(3p)\beta_0,\beta_1(D^2H)$	0.8668721	0.002384048	....
	$(3o)$	0.3031817	0.002529929	....
	$(5p)\beta_0,\beta_1(D),\beta_2(D^2)$	-3.2068	-0.0931619	0.1645083
	$(5o)$	0.479005	0.681579	0.1917292
	$(7p)\beta_0,\beta_1(D^2)$	-3.68215	0.1606684	....
	$(7o)$	-3.3994	0.1574598	....
Sweetgum	$(3p)\beta_0,\beta_1(D^2H)$	0.2976087	0.002579246	....
	$(3o)$	0.2848158	0.002605488	....
	$(5p)\beta_0,\beta_1(D),\beta_2(D^2)$	0.63616	-0.908149	0.1940991
	$(5o)$	2.268433	-1.30103	0.215419
	$(7p)\beta_0,\beta_1(D^2)$	-3.54821	0.1508756	....
	$(7o)$	-2.69169	0.1375616	....
Post oak	$(3p)\beta_0,\beta_1(D^2H)$	0.3127841	0.002961708	....
	$(3o)$	0.1054914	0.003102175	....
	$(5p)\beta_0,\beta_1(D),\beta_2(D^2)$	-1.28018	-0.0736108	0.1071492
	$(5o)$	-0.657807	-0.217158	0.1143979
	$(7p)\beta_0,\beta_1(D^2)$	-1.6376	0.1039153	....
	$(7o)$	-1.52254	0.102326	....

[1]Probability weighted regression estimates are designated (p); optimally weighted regression estimates are desigued (o).

tory estimation including component volumes by species and d.b.h. class. While some linear models perform well using either probability weights or optimal weights for correcting heteroscedasticity, other linear models predict volumes with substantial errors, particularly when fitted with optimal weights. Evidence from this study indicates that data from prism samples, where the probability of selection is a function of tree diameter, can be effectively fitted to certain linear models using probability weights. Optimal weights do not appear to provide any distinct advantages for timber inventory estimation. The best equation in this test-using diameter and merchantable height components combined in a single variable, and fitted with probability weights-predicted total inventory within a fraction of a percent. This equation also predicted inventory volumes for individual species by l-inch diameter classes within 10 percent of actual volumes.

LITERATURE CITED

Box, G.E.P. and Cox, D.R. 1964. An analysis of transformation. J.R. Stat. Soc. B, 26: 211-243.

Cunia, T. 1964. Weighted least squares method and construction of volume tables. For. Sci., 10:180-191.

DuMouchel, W.H. and Duncan, G.J. 1983. Using sample survey weights in multiple regression analysis of stratified samples. J. of the Am. Stat. Assoc., 78:535–543.

Freese, F. 1960. Testing accuracy. For. Sci. 6:139-145.

Grosenbaugh, L.R. 1964. STX-FORTRAN 4 program-for estimates of tree populations from 3P sample-tree-measurements. Pacific SW Forest and Range Expt. Sta. U.S. For. Serv. Res. Pap. PSW-13. 49 p.

Holt, D.; Smith, T.M.F.; and Winter, P.D. 1980. Regression analysis of data from complex surveys. J. R. Stat. Soc., 143:474–487.

Kmenta, J. 1971. Elements of econometrics. Macmillan Publ. Co., New York. 655 p.

Meng, C.H. and Tsai, W.Y. 1986. Selection of weights for a weighted regression of tree volume. Can. J. For. Res., 16:671–673.

Spurr, S.H. 1952. Forest Inventory. Ronald Press, New York. 476 p.

USDA Forest Service. 1985. Forest survey inventory work plan. Southern Forest Experiment Station, Forest Inventory and Analysis Res. Work Unit, Starkville, MS. Unpubl. 57 p.

Kelly, John F., and Beltz, Boy C. 198'7. A comparison of tree volume estimation models for forest inventory. Res. Pap. SO-233. New Orleans, LA; U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 9 p.

Tests are analyzed for eight different linear regression models to determine abilities to predict inventory for four tree species in Northeast Texas.

☆U.S GOVERNMENT PRINTING OFFICE:1 987 -761 -0 86/ 60 0 32