Multivariate regression model for predicting yields of grade lumber from yellow birch sawlogs

Andrew F. Howard
Daniel A. Yaussy

Abstract

A multivariate regression model was developed to predict green board-foot yields for the common grades of factory lumber processed from yellow birch factory-grade logs. The model incorporates the standard log measurements of scaling diameter, length, proportion of scalable defects, and the assigned USDA Forest Service log grade. Differences in yields between band and circular headrigs were accounted for in the model. Lumber yields predicted using the model compared favorably with guidelines for expected yields from factory-grade logs. The coefficients presented here can be used in computer programs for the purpose of sawmill simulations, economic analyses, or log yard inventory systems.

Sawmill personnel are using computers more often for log-inventory systems, mill efficiency studies, and detailed economic analyses. Therefore, there is an increased need for readily usable models for estimating lumber grade yields from sawlogs. Until recently, lumber grade yield data were only available in the form of tables. In most of these tables, dry board-foot yields are expressed as percentages of total lumber volume. These data are not easily incorporated into computer programs and are not for green board-foot yields. Equations for predicting green yields are preferable because many mills sell only green lumber, or dry only the higher grades. Both practices preclude the use of tables for dry board-foot yields. Another advantage to estimating green yields is that drying degrade factors specific to a given mill can be used to obtain an estimate of dry yields.

Researchers have begun to apply various statistical methods to the problem of specification and estimation of these elaborate models. The objective of the study reported here was to develop a regression model for predicting yields of grade lumber from yellow birch (*Betula alleghaniensis*, Britton) sawlogs. The method employed was suggested previously (7), and represents a continuation of work completed on five other hardwood species (8,9). This study goes one step further by accounting for the effects of different headrigs (band or circular) on lumber grade yields.

Lumber grade yield data

The data used in this study were from mill studies conducted by the USDA Forest Service, Northeastern Forest Experiment Station (NEFES) (4), and in conjunction with the Sawmill Improvement Program conducted by the NEFES and the Northeastern Area of State and Private Forestry. The studies were done at sawmills in Maine, Michigan, Wisconsin, West Virginia, and Vermont. The data included logs from eight sawmills, five with band headrigs, and three with circular headrigs. One of the band sawmills was equipped with a resaw. The logs from both sources were graded to specifications for USDA Forest Service factory-grade logs (5). The data were divided randomly into two sets containing 70 percent and 30 percent of the logs. The first was used for purposes of model specification. The second was used to validate the coefficients derived for the final model specification. Summary statistics for both data sets are presented in Tables 1 through 4.

The lumber produced from the sample logs was graded immediately after sawing by graders certified by the National Hardwood Lumber Association (NHLA) using NHLA rules. The following lumber grades were used:

The authors are, respectively, Assistant Professor, Harvesting and Wood Science, Univ. of British Columbia, Faculty of Forestry, 2357 Main Mall, Vancouver, BC, Canada V6T 1W5; and Research Forester, USDA Forest Serv., Northeastern Forest Expt. Sta., 359 Main Rd., Delaware, OH 43015. THis paper was received for publication in October 1985.

Forest Prod. J. 36(11/12):56-60.

Lumber Grade	Abbreviation
First and Seconds	FAS
FAS One Face	FAS1F
Selects	SEL
Number One Common	N1C
Number Two Common	N2C
Number Three A Common	N3A
Number Three B Common	N3B

Model development and validation

Theories regarding the appropriate form of models for predicting lumber yields from logs have changed very little since investigators first addressed the problem in the early 1900s. In the development of one of the first formula log rules, the International Rule, an allowance for sawkerf and shrinkage during drying was made (2). It was shown that this allowance was proportional to the cross-sectional area of the log, or the square of the diameter. Allowances for the round form of the log, minor crook, sweep, slabbing, and edgings were shown to be proportional to the surface area of the bark, or the diameter times the length. The International Rule formed the basis for later studies in which formula rules were compared to empirical log rules, and in fact is the underlying form of most yield equations in use today.

Not until almost 40 years later did researchers begin to examine formula and empirical log rules, and lumber yield equations with an eye toward statistical accuracy and efficiency. In a study of sawmill production (6), researchers showed that a logical and workable model for estimating lumber recovery (expressed in a variety of ways) was:

$$Y = b_1 L + b_2 DL + b_3 D^2 L \qquad [1]$$

where:

Y = variable of interest
D = log scaling diameter (in.)
L = log length (ft.)
b_n = theoretical coefficients

They used the method of ordinary least squares on sawmill yield data to estimate the coefficients, thus producing an empirical log rule. After fitting the model, they observed that the variance of the residuals increased with increasing log size (heteroscedasticity). This violates one of the assumptions made when applying ordinary least squares methods. They concluded that the model had to be weighted to reduce the effects of this condition. By fitting a straight line to the standard deviations of 820 logs grouped in 30 size classes, it was shown that the standard deviation of the volume of logs was proportional to the product of the square of the diameter and the length (D^2L). Equation [1] was transformed by dividing both sides by D^2L before deriving the final coefficients.

For the estimation to be a best linear unbiased estimate (BLUE) the sum of squared residuals of the regression should be weighted by the inverse of the variance. For the example used above:

$$W_i = 1/V_i \qquad [2]$$

where:

W_i = weight, group i
V_i = variance, group i

TABLE 1. — *Summary statistics for logs from sawmill yield studies. Model specification data set, band headrigs, 460 logs.*

Log grade and variable	Mean value	Standard deviation	Maximum value	Minimum value
Log grade 1 (88 logs)				
length (ft.)	13.6	2.2	16.5	10.0
scaling diameter (in.)	16.2	3.2	26.0	13.0
defect (%)	3.5	5.3	20.0	0.0
Log grade 2 (216 logs)				
length (ft.)	11.6	2.7	16.5	8.0
scaling diameter (in.)	13.9	3.1	29.0	10.0
defect (%)	5.9	9.0	46.0	0.0
Log grade 3 (156 logs)				
length (ft.)	11.1	2.4	16.7	8.0
scaling diameter (in.)	11.9	3.2	23.0	8.0
defect (%)	5.1	7.7	51.0	0.0

TABLE 2. — *Summary statistics for logs from sawmill yield studies. Model specification data set, circular headrigs, 351 logs.*

Log grade and variable	Mean value	Standard deviation	Maximum value	Minimum value
Log grade 1 (52 logs)				
length (ft.)	13.4	1.9	16.5	10.3
scaling diameter (in.)	16.6	3.0	25.0	13.0
defect (%)	5.5	7.1	36.0	0.0
Log grade 2 (116 logs)				
length (ft.)	11.5	2.4	16.5	8.0
scaling diameter (in.)	14.2	2.4	21.0	10.0
defect (%)	5.6	7.8	30.0	0.0
Log grade 3 (183 logs)				
length (ft.)	10.4	2.4	16.5	7.7
scaling diameter (in.)	10.9	2.3	19.0	8.0
defect (%)	6.6	9.3	45.0	0.0

TABLE 3. — *Summary statistics for logs from sawmill yield studies. Validation data set, band headrigs, 218 logs.*

Log grade and variable	Mean value	Standard deviation	Maximum value	Minimum value
Log grade 1 (26 logs)				
length (ft.)	14.5	2.5	16.6	10.0
scaling diameter (in.)	16.0	3.3	27.0	13.0
defect (%)	2.7	4.3	14.0	0.0
Log grade 2 (100 logs)				
length (ft.)	12.2	2.8	16.7	8.0
scaling diameter (in.)	13.2	2.0	19.0	10.0
defect (%)	5.3	10.2	54.0	0.0
Log grade 3 (92 logs)				
length (ft.)	11.8	2.6	16.4	8.0
scaling diameter (in.)	11.3	2.8	24.0	8.0
defect (%)	5.8	10.3	56.0	0.0

TABLE 4. — *Summary statistics for logs from sawmill yield studies. Validation data set, circular headrigs, 130 logs.*

Log grade and variable	Mean value	Standard deviation	Maximum value	Minimum value
Log grade 1 (18 logs)				
length (ft.)	13.2	2.3	16.5	10.3
scaling diameter (in.)	16.2	2.6	21.0	13.0
defect (%)	5.4	8.8	35.0	0.0
Log grade 2 (51 logs)				
length (ft.)	11.7	2.4	16.5	8.3
scaling diameter (in.)	14.7	2.5	20.0	11.0
defect (%)	7.4	10.1	47.0	0.0
Log grade 3 (61 logs)				
length (ft.)	10.7	2.2	14.4	8.0
scaling diameter (in.)	11.3	2.8	26.0	8.0
defect (%)	7.7	10.9	45.0	0.0

and

$$SSR = W_i (y_i - b_0 - b_1 x_i)^2 \quad [3]$$

where:

SSR = the sum of squared residuals from regression
y_i = the observed values, group i
b_n = the theoretical coefficients
x_i = the explanatory variable, group i

or

$$SSR = (\sqrt{W_i}\, y_i - \sqrt{W_i}\, b_0 - \sqrt{W_i}\, b_1 x_i)^2 \quad [4]$$

The result obtained in Equation [4] indicates that the model should be transformed by multiplying by the inverse of the square root of the variance, or the inverse of the standard deviation. Because it was previously demonstrated that D^2L is proportional to the standard deviation (6), in practice the weighting is accomplished by dividing both sides of the equation by D^2L. When expected values are computed with equations derived using the transformed model, it is necessary to multiply the results by D^2L if values in the original units are desired. A detailed presentation of this procedure is available (3). Further discussion of the relationship between formula rules, empirical log rules, and product yield equations is also available (1).

The model specification in Equation [1] is sufficient when sawing time for logs, gross lumber volume, or some other measure of gross yield is the desired result. However, to determine the yields of individual lumber grades, and to include the effect of log grade and scalable defect, the model and analysis must be expanded. Furthermore, to account for differences in yields between band and circular headrigs due to differing sawkerfs, the model must be expanded. A multivariate formulation of the model which incorporates the effect of correlation between yields of individual lumber grades has been suggested (7). The basic form of this relationship is:

$$y = Bx \quad [5]$$

where:

y = a column vector of lumber grade yields
B = a matrix of coefficients determined by regression analysis
x = a column vector of independent variables

In recent work (8), dummy variables for log grade, and a variable for the proportion of the log with scalable defect, have been added to the column vector of independent variables. The model specification used here is similar, however, it was expanded by including dummy variables for headrig type. It is given in Equation [6].

$$\begin{aligned} Y_i = {} & b_{i0}\, DV_3 + b_{i1}\, DV_3 1/D + b_{i2}\, DV_3 1/D^2 + b_{i3}\, DV_3DEF + b_{i4}\, DV_3DV_1 + b_{i5}\, DV_3DV_1 1/D + b_{i6}\, DV_3DV_1 1/D^2 + b_{i7}\, DV_3DV_1DEF \\ & + b_{i8}\, DV_3DV_2 + b_{i9}\, DV_3DV_2 1/D + b_{i10}\, DV_3DV_2 1/D^2 + b_{i11}\, DV_3DV_2DEF + b_{i12}\, DV_4 + b_{i13}\, DV_4 1/D + b_{i14}\, DV_4 1/D^2 + b_{i15}\, DV_4DEF \\ & + b_{i16}\, DV_4DV_1 + b_{i17}\, DV_4DV_1 1/D + b_{i18}\, DV_4DV_1 1/D^2 + b_{i19}\, DV_4DV_1DEF + b_{i20}\, DV_4DV_2 + b_{i21}\, DV_4DV_2 1/D + b_{i22}\, DV_4DV_2 1/D^2 + b_{i23}\, DV_4DV_2DEF \end{aligned} \quad [6]$$

where:

Y_i = green lumber tally, grade i divided by D^2L
D = log scaling diameter (in.)
DV_1 = "1" for log grade 1, "0" otherwise
DV_2 = "1" for log grade 2, "0" otherwise
DV_3 = "1" for band headrig, "0" otherwise
DV_4 = "1" for circular headrig, "0" otherwise
DEF = one minus the ratio of net log scale to gross log scale divided by D^2L
b_{ij} = theoretical coefficients

The coefficients obtained by using ordinary least-squares regression techniques with a multivariate model are the same as those estimated by running each lumber grade separately with the same specification. However, the test for significance of the independent variables considers the effect these variables have on all lumber grades simultaneously. The r^2 values of the individual univariate regressions hold little significance in the multivariate model. The method for determining the fit of the model is to subtract the Wilks-lambda statistic (lambda) for the overall model from one. This statistic is equivalent to r^2 in the univariate case.

Equation [6] was fit to the green board-foot volume data from the sawmill yield studies, and each independent variable was tested for significance at the 5 percent level by the Wilks-lambda criterion. Variables for which parameter estimates were not significantly different from zero were dropped from the model. The resulting model was:

$$\begin{aligned} Y_i = {} & b_{i0}\, DV_3 + b_{i1}\, DV_3 1/D + b_{i3}\, DV_3DEF + b_{i5}\, DV_3DV_1\, 1/D + b_{i6}\, DV_3DV_1 1/D^2 + b_{i7}\, DV_3DV_1DEF + b_{i8}\, DV_3DV_2 + b_{i9}\, DV_3DV_2 1/D \\ & + b_{i10}\, DV_3DV_2 1/D^2 + b_{i11}\, DV_3DV_2DEF + b_{i12}\, DV_4 + b_{i13}\, DV_4 1/D + b_{i14}\, DV_4 1/D^2 + b_{i15}\, DV_4DEF + b_{i16}\, DV_4DV_1 \\ & + b_{i17}\, DV_4DV_1 1/D + b_{i18}\, DV_4DV_1 1/D^2 + b_{i19}\, DV_4DV_1DEF + b_{i20}\, DV_4DV_2 \end{aligned} \quad [7]$$

The model was validated according to methods previously described (7). A unique matrix of coefficients was derived using the validation data set. These were compared to the coefficients derived from the test data set. If there is no significant difference between the two sets of coefficients, the model has been validated. The validation test of the model resulted in a lambda of 0.90 which is greater than the 0.05 level critical value of 0.85. Because the sample size is large (n = 1,159), a slightly modified Wilks-lambda statistic has a chi-square distribution. The approximate chi-square statistic is 121.4 which, with 133 degrees of freedom, is also significant at the 0.05 level. Therefore, the hypothesis that there was no difference between the two sets of coefficients was not rejected by either measure; the model was validated.

Final estimates of the parameters were made using both data sets. All coefficients were significant at the 0.05 level using the Wilks-lambda criterion. The Wilks-lambda statistic for the overall model was 0.2014. The equivalent r^2 value was 0.7986, which indicates a relatively good fit.

TABLE 5. — *Matrix of coefficients (b_{ij}) for the final model: band headrigs.*

Independent variable	(j)	Lumber grades(i)[a] FAS 1	FAS1F 2	SEL 3	N1C 4	N2C 5	N3A 6	N3B 7
D^2L	0	0.008	0.0042	−0.0005	0.0183	0.0128	−0.0012	−0.001
DL	1	−0.0298	−0.038	0.0262	−0.01289	−0.0239	0.1141	0.1065
DEF	3	−1.5189	−2.512	−12.506	−24.058	−10.663	−1.2704	17.025
DV_1DL	5	0.6008	0.1855	−0.1224	−0.0436	−0.2345	−0.0074	0.0168
DV_1L	6	−6.9913	−2.4734	2.6625	0.576	2.2886	−0.3706	−1.0585
DV_1DEF	7	−93.578	55.562	−59.729	35.813	56.919	−3.0775	9.4166
DV_2D^2L	8	0.0345	0.0127	−0.0037	0.0187	−0.024	−0.0104	−0.0016
DV_2DL	9	−0.7733	−0.2739	0.1045	−0.595	0.5517	0.2947	0.0477
DV_2L	10	4.5162	1.5835	−0.3068	4.5694	−3.3611	−2.3573	−0.6736
DV_2DEF	11	−21.661	−5.4617	−14.663	−2.9576	22.888	9.5605	−6.7238

[a]FAS = First and Seconds; FAS1F = FAS One Face; SEL = Selects; N1C = Number One Common; N2C = Number Two Common; N3A = Number Three A Common; N3B = Number Three B Common.

TABLE 6. — *Matrix of coefficients (b_{ij}) for the final model: circular headrigs.*

Independent variable	(j)	Lumber grades(i)[a] FAS 1	FAS1F 2	SEL 3	N1C 4	N2C 5	N3A 6	N3B 7
D^2L	12	0.008	0.0055	−0.0021	0.0011	−0.0043	−0.0059	0.0113
DL	13	−0.151	−0.078	0.062	−0.0376	0.4473	0.2916	−0.2844
L	14	0.7298	0.2774	−0.3294	−0.2089	−2.867	−1.5176	2.692
DEF	15	−4.094	−3.1082	−2.1622	−15.67	−13.835	−2.4967	2.9335
DV_1D^2L	16	0.0744	−0.0486	0.017	−0.0079	0.0018	0.0043	−0.0238
DV_1DL	17	−1.9282	1.7539	−0.5518	0.1633	−0.2643	−0.2072	0.7246
DV_1L	18	13.766	−13.558	4.764	0.2849	1.8997	1.0791	−5.8327
DV_1DEF	19	−86.361	−28.122	−10.585	−0.0007	15.448	31.867	−1.5584
DV_2D^2L	20	0.0017	0.0023	0.0008	0.0042	−0.0032	−0.0023	−0.0009

[a]See footnote in Table 5.

Results and discussion

The coefficients for the final model are given in Tables 5 and 6. For use, the weighted model given in Equation [6] is rewritten by multiplying both sides of the reduced form (Equation [7]) by D^2L. Actually, the coefficients were estimated for the nontransformed model by weighting the sum of squared residuals. The weight was designated as the inverse of the square of the quantity D^2L (see Equation [2]). The coefficients for the two sawmill types must be used separately to account for the differences in yield due to sawkerf.

Mean values for predicted and actual board-foot yields for each lumber grade are given in Table 7. In all but three cases, the mean predicted yields were within 5 percent of the actual mean values. The only exceptions were for combinations of log and lumber grades with expected yields below 4.0 board feet. This indicates that the predictive accuracy of the model is satisfactory.

The model was used to estimate the proportion of Number One Common and better (N1C&B) grades of lumber sawn from the three grades of logs. The results of this analysis are given in Table 8. According to the log grade specifications (5), grade 1 logs should yield 65 percent or more of N1C&B. Grade 2 logs should yield between 40 and 64 percent, and grade 3 logs from 13 to 39 percent. In all cases the predicted yields were within the suggested range. This indicates that the log grading system can effectively discriminate between classes of logs when the board-foot yields of various lumber grades are used as the criteria. Furthermore, it demonstrates the importance of using log grades to determine value classes, which can be used to decide what price to pay for logs.

Conclusions

Lumber grade yields can be estimated from the standard log measurements and the assigned USDA

TABLE 7. — *Mean values for predicted and actual board-foot yields, 1,159 logs.*

Headrig and lumber grade[a]	Log grade 1 Actual	1 Predicted	2 Actual	2 Predicted	3 Actual	3 Predicted
Band						
FAS	41.0	39.2	9.2	8.9	1.9	1.3
FAS1F	15.2	15.9	6.7	6.6	1.9	1.8
SEL	11.6	11.2	5.9	6.2	1.9	2.0
N1C	38.1	38.2	21.1	20.5	12.4	12.1
N2C	23.3	23.2	20.1	20.2	17.5	17.5
N3A	14.1	13.8	11.8	11.7	12.9	13.0
N3B	9.8	9.8	11.2	11.5	13.0	13.3
Circular						
FAS	37.8	37.3	6.8	6.7	0.4	0.5
FAS1F	28.2	28.5	9.4	9.1	1.0	1.0
SEL	6.1	6.3	2.7	3.0	0.9	0.7
N1C	41.0	40.4	28.7	28.1	7.3	7.3
N2C	18.3	18.2	21.3	21.6	14.6	14.5
N3A	8.3	8.3	10.8	10.6	9.5	9.6
N3B	8.2	8.3	9.4	9.6	10.5	10.6

[a]See footnote in Table 5.

Forest Service log grade. The model presented here can be used in computer programs to estimate the expected yields of each lumber grade from sawlogs. The model can be modified to predict any combination of the seven lumber grades by simple summing the coefficients for each independent variable (for example, Number One Common and better grades). This allows users to simplify the process of incorporating the model into existing computer programs, and affords the opportunity to customize the equation according to inventory and marketing requirements. The model can be used by sawmills with circular or band headrigs.

TABLE 8. — *Percent yield of Number One Common and Better lumber for band and circular sawmills.*

Log grade	Band No. of logs	Band N1C&B[a]	Circular No. of logs	Circular N1C&B[a]
		(%)		(%)
1	114	69.1	70	76.4
2	316	49.3	167	52.4
3	248	28.2	244	21.4

[a]N1C&B = Number One Common and better grades.

Literature cited

1. Bruce, D. 1970. Predicting product recovery from logs and trees. USDA Forest Serv. Res. Pap. PNW-107. 15 pp.
2. Clark, J.F. 1906. The measurement of sawlogs. Forestry Quart. 4(2):79-83.
3. Furnival, G.M. 1961. An index for comparing equations used in constructing volume tables. Forest Sci. 7(4):337-341.
4. Hanks, L., G. Gammon, R. Brisbin, and E. Rast. 1980. Hardwood log grades and lumber grade yields for factory lumber logs. USDA Forest Serv. Res. Pap. NE-468. 92 pp.
5. Rast, E., D. Sonderman, and G. Gammon. 1973. A guide to hardwood log grading. USDA Forest Serv. Gen. Tech. Rept. NE-1. 34 pp.
6. Schumacher, F.X. and W.C. Jones. 1940. Empirical log rules and the allocation of sawing time to log size. J. of Forestry 38(11):889-896.
7. Seegrist, D.W. 1975. A multivariate model and statistical method for validating tree grade lumber yield equations. USDA Forest Serv. Res. Pap. NE-320. 19 pp.
8. Yaussy, D.A. and R.L. Brisbin. 1983. Multivariate regression model for predicting lumber grade volumes of northern red oak sawlogs. USDA Forest Serv. Res. Pap. NE-536. 11 pp.
9. ________. 1986. Green lumber grade yields from factory grade logs of three oak species. Forest Prod. J. 36(5):53-56.