United States
Department of
Agriculture

Forest Service

**Northeastern Forest
Experiment Station**

**Research
Paper NE-536**

1983

Multivariate Regression Model for Predicting Lumber Grade Volumes of Northern Red Oak Sawlogs

Daniel A. Yaussy
Robert L. Brisbin

Abstract

A multivariate regression model was developed to predict green board-foot yields for the seven common factory lumber grades processed from northern red oak *(Quercus rubra* L.) factory grade logs. The model uses the standard log measurements of grade, scaling diameter, length, and percent defect. It was validated with an independent data set. The model can be modified to predict various combinations of lumber grades.

The Authors

Daniel A. Yaussy, research forester, received his B.S. degree in Natural Resources (Forestry) from the Ohio State University in 1976 and M.S. degree in Forest Biometrics from Virginia Polytechnic Institute and State University in 1978. He joined the Forest Service in 1979 and is presently a member of the non-industrial private forestry project at Delaware, Ohio.

Robert L. Brisbin, research forest products technologist, received his B.S. degree in forestry from Iowa State University in 1961 and M.S. degree in wood technology from the Pennsylvania State University in 1969. He has been with the Northeastern Forest Experiment Station since 1961, and is now on the staff of the non-industrial private forestry project at Delaware, Ohio.

Manuscript received for
publication 27 June 1983

Introduction

Tables for estimating northern red oak dry lumber grade yields from factory grade logs have been developed in the past (Hanks et al. 1980; Schroeder and Hanks 1967). These tables were developed to provide estimates of lumber grade yields and values for graded hardwood logs. The yields by lumber grade were expressed as a percentage of total dry lumber tally volume. To apply the percentages, an independent estimate of lumber tally volume was required, using an appropriate log rule and estimated mill overruns. The tables provide the means to calculate good estimates of log value, but are difficult to use in mill efficiency simulation and economic models. Most mill efficiency models also require estimates of green yields rather than dry yields so that in the analysis the drying process can be separated from the sawing process.

We have developed a multivariate regression equation for predicting green board-foot lumber grade yields that is easy to use with computer analysis techniques. This report does not supersede the previous reports. It provides a tool for analyzing log quality implications more efficiently in the manufacture of hardwood factory lumber.

Data

The logs used in this study were sawn at eight band sawmills located in four states — Virginia, West Virginia, North Carolina, and Tennessee. This data set is part of the data used in the study by Hanks et al. (1980).

Although many independent variables were originally recorded for each log, only four proved important for yield predictions:

1) Factory log grade, as defined by Rast et al. (1973);
2) Scaling diameter (D) of the small end of the log to the nearest inch
3) Log length (L) in feet
4) Percent defect (P) expressed in decimal form (Rast et al. 1973).

The lumber resulting from milling the logs was graded in the green condition by graders certified by the National Hardwood Lumber Association. The lumber grades used are listed below with their abbreviations:

Lumber Grade	*Abbreviation*
Firsts and Seconds	FAS
FAS one Face	F1F
Selects	SEL
One Common	1C
Two Common	2C
Three A Common	3A
Three B Common	3B

A data set of 1053 logs was used in developing the model. Table 1 presents the average log measurements by log grade.

Table 1.—Averages and ranges of the log measurements of the model development data set

Variable	Average	Minimum	Maximum	Standard deviation
		Log Grade 1 (325 logs)		
Scaling diameter (inches)	19.7	13.0	31.0	3.7
Length (feet)	14.1	10.0	17.0	2.2
Scalable defect (proportion)	0.04	0.00	0.36	0.06
		Log Grade 2 (435 logs)		
Scaling diameter	15.6	10.0	29.0	3.6
Length	12.6	8.0	16.9	2.5
Defect	0.07	0.00	0.42	0.09
		Log Grade 3 (293 logs)		
Scaling diameter	12.8	8.0	27.0	3.6
Length	11.9	8.0	16.8	2.6
Defect	0.08	0.00	0.59	0.10

Model Development

Multivariate regression analysis was appropriate for the data set since there were multiple dependent variables, seven lumber grade yields for each log. The model is of the general form

$$y = xB$$

where

y = a row vector of the seven lumber grade yields
x = a row vector of the independent variables
B = a matrix of coefficients determined by regression analysis.

Since log grade is a classification variable with three levels, dummy variables, G_1 and G_2 were used to account for lumber grade yield variations due to differences in log grades (Draper and Smith 1981). The different combinations of G_1 and G_2 for the three log grades are:

Log Grade	G_1	G_2
One	1	0
Two	0	1
Three	0	0.

A weighted regression model proposed by Bruce (1970) was expanded to include dummy variables and multiple yields. It had the form:

$$Y_i/(D^2L) = b_{i0} + b_{i1}/D + b_{i2}/D^2 + b_{i3} \bullet P + b_{i4} \bullet G_1 + b_{i5} \bullet G_1/D + b_{i6} \bullet G_1/D^2 + b_{i7} \bullet G_1 \bullet P + b_{i8} \bullet G_2 + b_{i9} \bullet G_2/D + b_{i10} \bullet G_2/D^2 + b_{i11} \bullet G_2 \bullet P$$

where

Y_i = green board foot volume for the *ith* lumber grade, $i = 1, \ldots, 7$; Y_i = FAS, . . ., 3B
D^2L = weighting variable
b_{ij} = coefficients determined by regression for the *ith* lumber grade and the *jth* independent variable.

This model resulted in logical predictions of green board-foot volumes for each lumber grade. Two of the dummy variables ($G1/D^2$ and $G2/D$) were not significant at the 0.05 level of significance. Therefore, the equation was reduced to the following form for testing (coefficients will be presented later):

$$Y_i/(D^2L) = b_{i0} + b_{i1}/D + b_{i2}/D^2 + b_{i3} \bullet P + b_{i4} \bullet G_1 + b_{i5} \bullet G_1/D + b_{i6} \bullet G_1 \bullet P + b_{i7} \bullet G_2 + b_{i8} \bullet G_2/D_2 + b_{i9} \bullet G_2 \bullet P. \quad (1)$$

Validation

The set of data from a ninth mill was excluded from the model development phase of the study and used as a validation data set. This set consisted of 98 logs. Averages of log measurements are listed in Table 2. The method used for testing the equation involves the calculation of coefficients for the proposed model based on the validation set (Seegrist 1975). These validation coefficients are then compared to the coefficients from the developed model. If there is no statistically significant difference between the two sets of coefficients based on the Wilks-Lambda statistic, the model has passed the validation test.

The Wilks-Lambda statistic resulting from the test was 0.89. Since this is larger than the critical value, 0.84, at the 0.05 level of significance, the hypothesis that there was no difference in the coefficients between the two data sets was not rejected; therefore, the model was validated.

Table 2.—Averages and ranges of the log measurements of the validation data set

Variable	Average	Minimum	Maximum	Standard deviation
		Log Grade 1 (26 logs)		
Scaling diameter (inches)	18.6	13.0	25.0	2.9
Length (feet)	13.4	10.0	15.0	1.4
Defect	0.07	0.00	0.24	0.7
		Log Grade 2 (40 logs)		
Scaling diameter	15.1	11.0	24.0	2.8
Length	13.1	8.1	17.4	2.2
Defect	0.05	0.00	0.20	0.06
		Log Grade 3 (32 logs)		
Scaling diameter	11.5	8.0	16.0	2.2
Length	11.6	8.0	16.7	2.5
Defect	0.06	0.00	0.27	0.06

Results

The regression was rerun using the data from all the mills. All of the coefficients were significantly different from zero at the 0.05 level, with an overall Wilks-Lambda statistic of 0.005, indicating a very good fit of the model. For use, the weighted model (1) is rewritten by multiplying both sides of the reduced equation by D^2L. The final equation is then of the form:

$$Y_i = b_{i0} \cdot D^2L + b_{i1} \cdot D \cdot L + b_{i2} \cdot L + b_{i3} \cdot D^2 \cdot L \cdot P + b_{i4} \cdot D^2 \cdot L \cdot G_1 + b_{i5} \cdot D \cdot L \cdot G_1 + b_{i6} \cdot D^2 \cdot L \cdot P \cdot G_1 + b_{i7} \cdot D^2 \cdot L \cdot G_2 + b_{i8} \cdot L \cdot G_2 + b_{i9} \cdot D^2 \cdot L \cdot P \cdot G_2 \qquad (2)$$

The final coefficients from this pooled data set are listed in Table 3. A comparison of the actual and predicted average yields for the pooled data set is presented in Table 4.

To further simplify the model, the dummy variables for log grades can be eliminated, resulting in three separate equations for each lumber grade:

log grade 1, lumber grade i

$$Y_i = (b_{i0} + b_{i4})D^2 \cdot L + (b_{i1} + b_{i5})D \cdot L + b_{i2} \cdot L + (b_{i3} + b_{i6})D^2 \cdot L \cdot P$$

log grade 2, lumber grade i

$$Y_i = (b_{i0} + b_{i7}) \cdot D^2 \cdot L + b_{i1} \cdot D \cdot L + (b_{i2} + b_{i8}) \cdot L + (b_{i3} + b_{i9})D^2 \cdot L \cdot P$$

log grade 3, lumber grade i

$$Y_i = b_{i0} \cdot D^2 \cdot L + b_{i1} \cdot D \cdot L + b_{i2} \cdot L + b_{i3} \cdot D^2 \cdot L \cdot P.$$

Table 3.—Matrix of coefficients (b_{ij}) for the final model

Lumber grades	(i)	Independent variables (j)									
		D^2L	DL	L	D^2LP	D^2LG_1	DLG_1	D^2LPG_1	D^2LG_2	LG_2	D^2LPG_2
		0	1	2	3	4	5	6	7	8	9
FAS	1	0.0049	–0.0815	0.3414	–0.0006	0.0284	–0.2407	–0.0279	0.0058	–0.5122	–0.0071
F1F	2	0.0066	–0.1049	0.4188	–0.0010	0.0023	0.0403	0.0025	0.0028	–0.0873	–0.0070
SEL	3	0.0007	–0.0132	0.0894	0.0009	0.0012	–0.0086	–0.0015	–0.0002	0.1684	–0.0022
1C	4	0.0267	–0.3261	1.1441	–0.0066	–0.0107	0.1758	0.0190	0.0054	–0.1512	–0.0082
2C	5	–0.0028	0.4062	–2.4485	–0.0050	–0.0060	0.0213	0.0041	–0.0040	0.4528	0.0043
3A	6	–0.0011	0.1663	–0.1109	–0.0047	0.0004	–0.1081	0.0109	–0.0027	–0.2437	0.0083
3B	7	0.0053	–0.0226	0.2609	–0.0012	–0.0041	–0.0008	0.0025	–0.0034	0.0864	0.0056

Table 4.—Averages of actual and predicted green board-foot volumes for the pooled data set

Lumber grade	Log Grade 1		Log Grade 2		Log Grade 3	
	Actual	Predicted	Actual	Predicted	Actual	Predicted
FAS	97.2	94.5	15.3	14.7	1.9	1.9
F1F	38.0	38.5	12.3	12.2	2.7	2.8
SEL	6.5	6.2	2.2	2.2	0.7	0.7
1C	66.2	67.2	48.8	48.4	19.2	19.0
2C	34.1	34.3	32.8	32.6	26.6	25.9
3A	12.1	12.2	16.6	16.7	20.3	20.8
3B	4.1	4.1	6.9	7.0	10.1	10.6
Total	258.2	257.0	134.9	133.8	81.5	81.7

Table 5 is an example of the tabular volumes estimated from the multivariate model for grade 1 12-foot logs with 5 percent scalable defect. This table cannot be directly compared with the tables developed by Hanks et al. (1980) until the volumes are converted to percentages and green to dry degrade factors are applied.

By substituting the average log length and defect of the validation set (Table 2) into the final model (2), we can derive an estimate of total green board foot lumber tally by scaling diameter. The total lumber tally estimates from the model are plotted with the actual data from the validation set in Figures 1, 2, and 3. We also plotted the appropriate volumes from the net International 1/4-inch and Doyle log rules. The relationships between estimated lumber tally volume and the log rules are reasonable for an efficient band sawmill producing standard factory lumber.

One common way to judge the effectiveness of different quality standards is to compare their estimates of volume in an index lumber grade such as No. 1 Common and Better. In Figures 4, 5, and 6 the predicted volume of 1C and Better from the multivariate model is plotted over the actual validation data. There is a distinct and uniform difference among the three log grades.

Table 5.—Predicted green board-foot lumber grade yields (log grade 1, log length 12, defect 0.05)

Scaling	Lumber grade						
diameter	FAS	F1F	SEL	1C	2C	3A	3B
13	18	13	1	24	19	7	2
14	25	15	2	28	22	8	2
15	32	18	2	32	24	8	2
16	40	20	3	36	26	9	3
17	49	23	3	41	27	9	3
18	58	26	4	46	29	10	3
19	69	29	4	51	30	10	3
20	80	33	5	57	31	11	4
21	92	36	5	64	32	11	4
22	104	40	6	71	32	12	4
23	117	44	7	78	32	12	5
24	132	48	8	85	33	13	5
25	146	53	9	93	33	13	6
26	162	58	9	102	32	14	6
27	179	63	10	110	32	14	7
28	196	68	11	120	31	15	7
29	214	73	12	129	30	15	8
30	232	79	13	139	29	15	8

Applications

As shown in Figures 1, 2, and 3, the model provides good estimates of total green lumber tally volume. This eliminates the need to calculate log scale volume and overrun in some computer analysis applications. Sawmill managers with computer facilities may also use the model to predict expected total lumber tally volume and volume by lumber grade from a group of graded logs.

A multivariate model is flexible enough that appropriate lumber grade coefficients can be combined if some of the seven grades are not separated at a particular mill. For example, if No. 3 common (3C) lumber is not broken down into 3A and 3B lumber grades, the coefficients for 3A and 3B (rows 6 and 7 in Table 3) may be added together. This results in a 6 × 10 matrix of coefficients that predicts FAS, F1F, SEL, 1C, 2C, and 3C. The equation for predicting the volume of 3C lumber is then:

$$
\begin{aligned}
Y_6 + Y_7 = {} & 0.0042(D^2 \cdot L) + 0.1437(D \cdot L) \\
& + 0.1500(L) - 0.0059(D^2 \cdot L \cdot P) \\
& - 0.0037(D^2 \cdot L \cdot G_1) - 0.1089(D \cdot L \cdot G_1) \\
& + 0.0134(D^2 \cdot L \cdot P \cdot G_1) - 0.0061(D^2 \cdot L \cdot G_2) \\
& - 0.1573(L \cdot G_2) + 0.0139(D^2 \cdot L \cdot P \cdot G_2).
\end{aligned}
$$

As this equation illustrates, the same result can be obtained by predicting 3A and 3B lumber yields and adding these yields together to predict 3C lumber yield. The above methods can be used to predict other grade combinations such as 1C and better or F1F and better.

(text follows on page 11)

Figure 1.—Comparison of total board volume predicted by the model, and by two popular log rules, with the actual volume of validation data set (Log Grade 1).

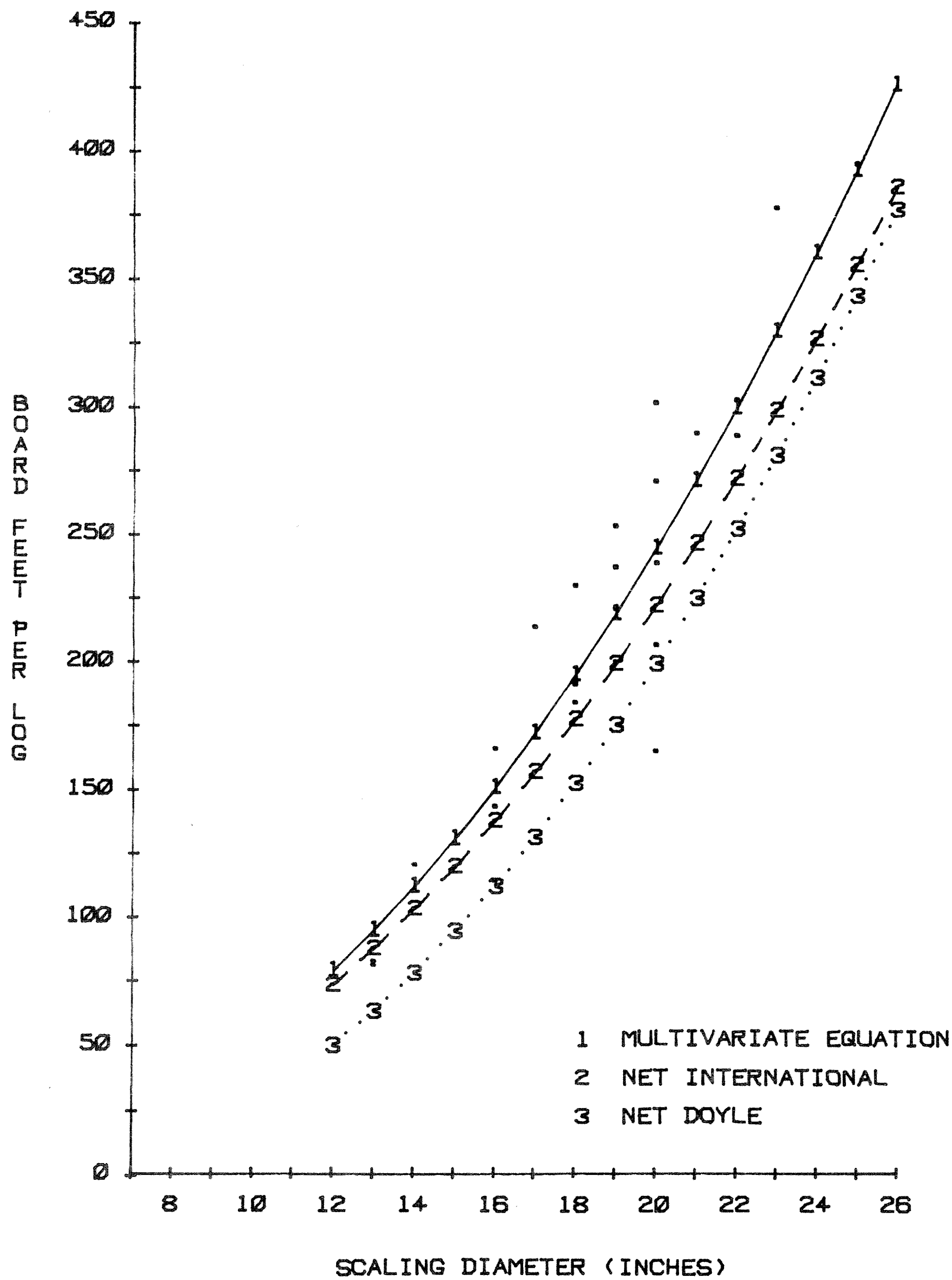


Figure 2.—Comparison of total board volume predicted by the model, and by two popular log rules, with the actual volume of validation data set (Log Grade 2).

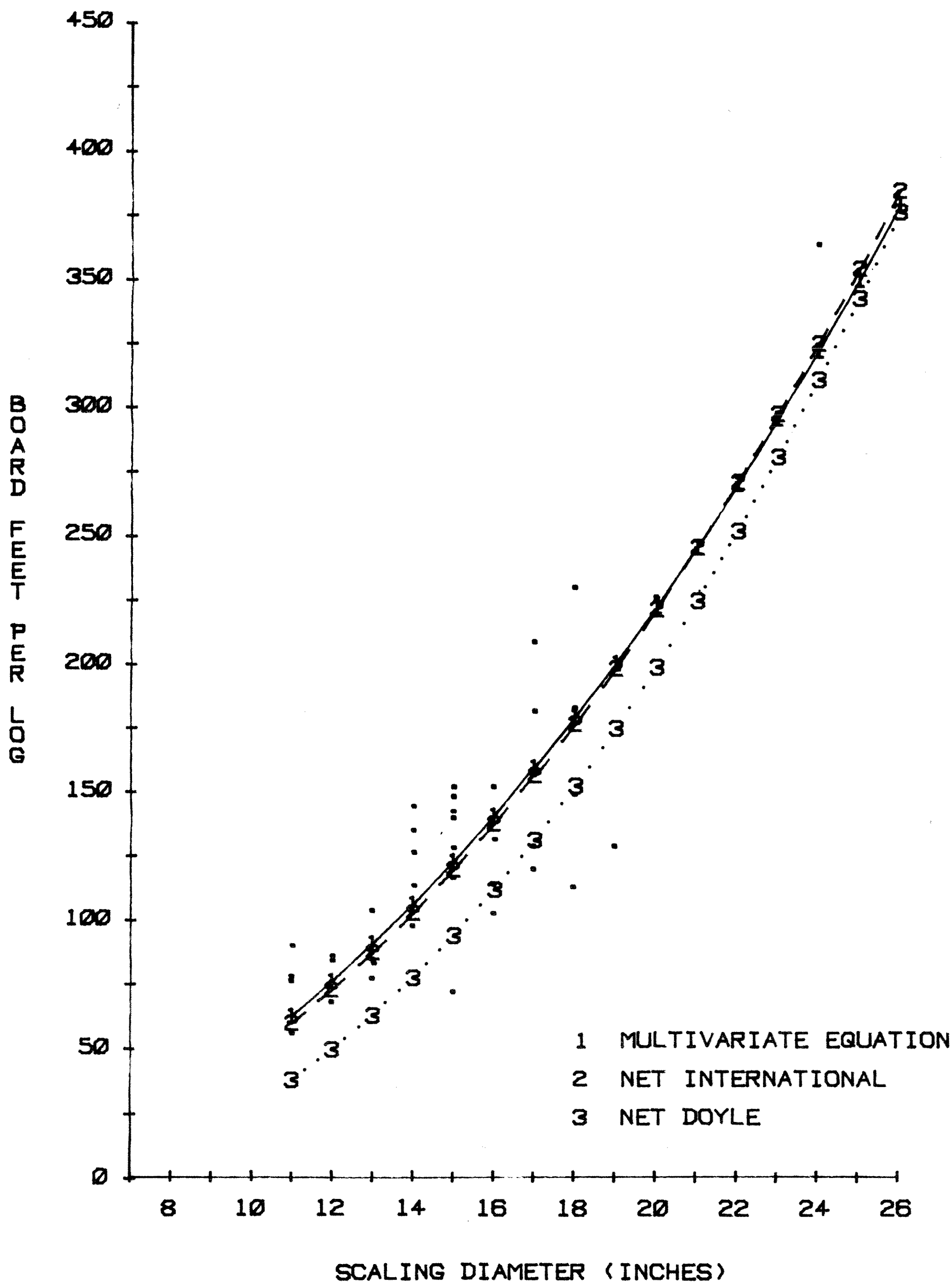


Figure 3.—Comparison of total board volume predicted by the model, and by two popular log rules, with the actual volume of validation data set (Log Grade 3).

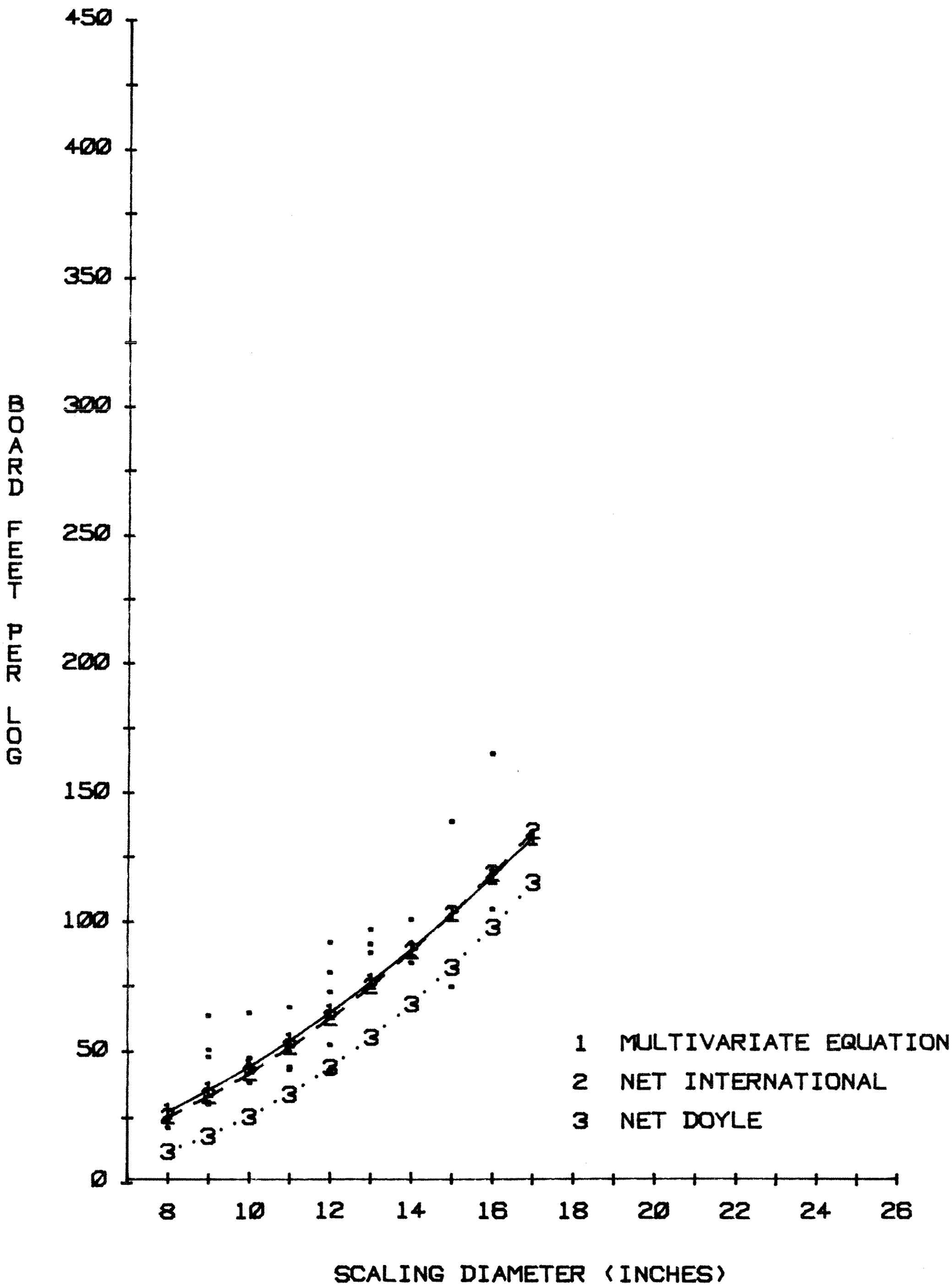


Figure 4.—Predicted volume of No. 1 Common and Better compared with actual volume of the validation data set (Log Grade 1).

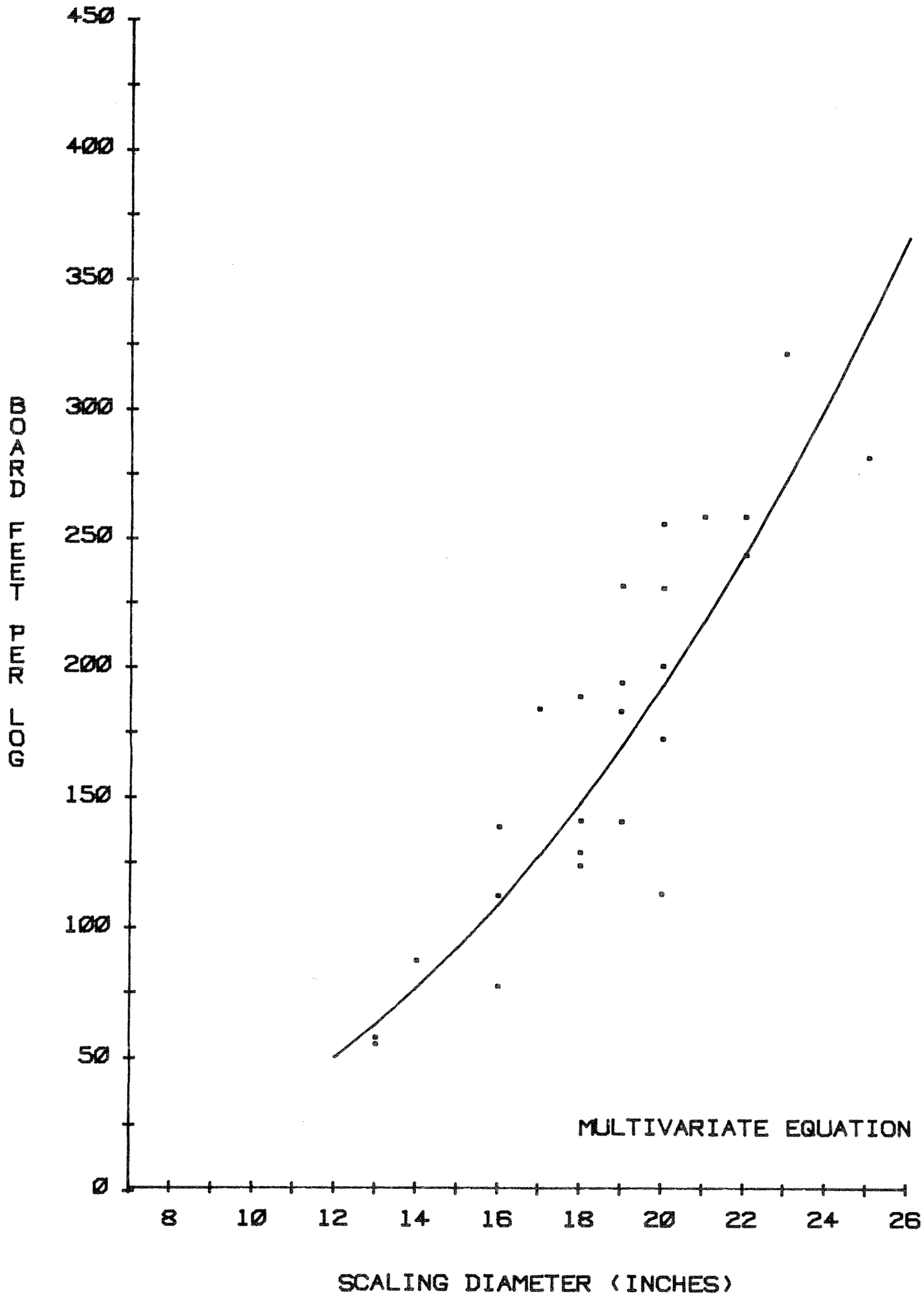


Figure 5.—Predicted volume of No. 1 Common and Better compared with actual volume of the validation data set (Log Grade 2).

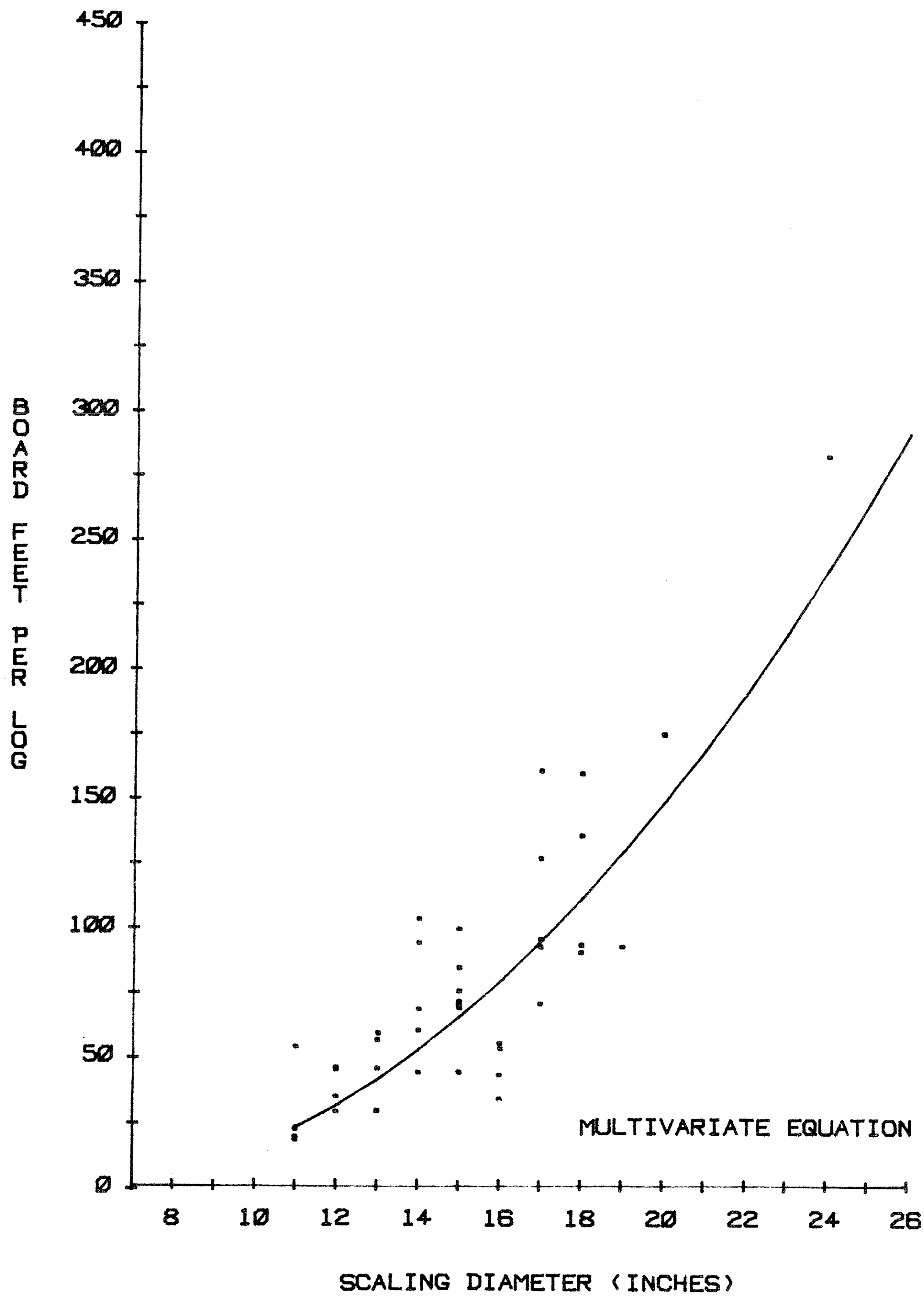
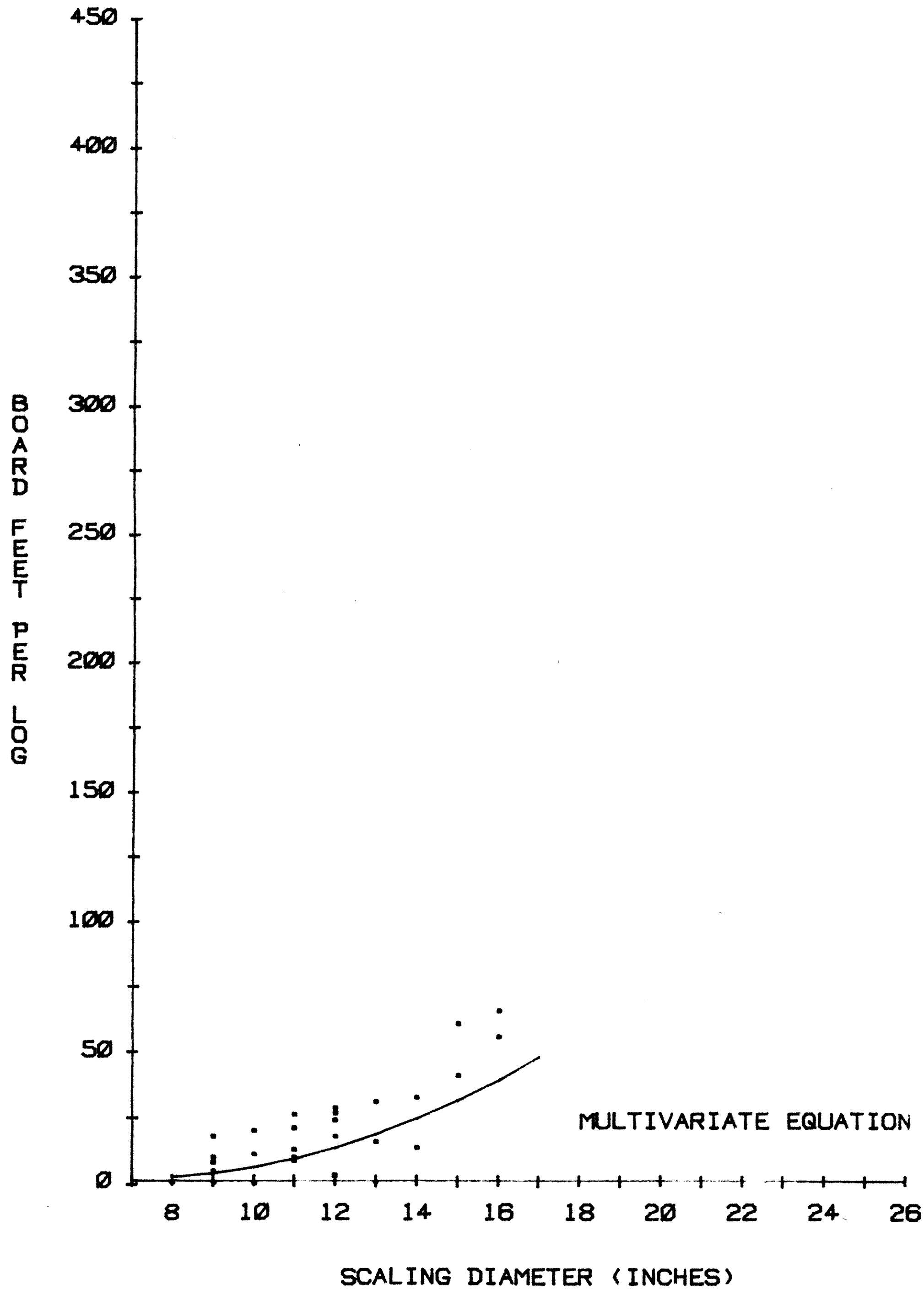


Figure 6.—Predicted volume of No. 1 Common and Better compared with actual volume of the validation data set (Log Grade 3).



The examples point out another aspect of the model's flexibility. It can be used as a single large matrix equation to predict board-foot volumes of all seven lumber grades simultaneously for any group of factory grade logs. Only log grade, diameter, length, and percent defect are needed to solve the equations. Because there are a large number of coefficients in the complete model, at least a minicomputer with substantial memory and matrix algebra functions is necessary to solve the equations efficiently. The model can also be divided into seven separate equations, one for each lumber grade, and solved in stepwise fashion on smaller microcomputers. Finally, any number of the seven equations could be combined into broader product classes by adding the appropriate coefficients. These combined equations could be solved by some programmable hand-held calculators.

Although this model was developed for northern red oak sawed in band sawmills, it may be suitable for related species or circular sawmills if the proper conversions are applied. The model is presently being tested on other species.

Literature Cited

Bruce, David. **Predicting product recovery from logs and trees.** Res. Pap. PNW-107. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station; 1970. 15 p.

Draper, N.; Smith, H. **Applied regression analysis, 2d ed.** New York: Wiley; 1981. p. 241-243.

Hanks, L.F.; Gammon, G.L.; Brisbin, R.L.; Rast, E.D. **Hardwood log grades and lumber grade yields for factory lumber logs.** Res. Pap. NE-468. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station; 1980. 92 p.

Rast, E.D.; Sonderman, D.L.; Gammon, G.L. **A guide to hardwood log grading.** Gen. Tech. Rep. NE-1. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station; 1973; 31 p.

Schroeder, J.G.; Hanks, L.F. **Lumber grade yields for factory-grade northern red oak sawlogs.** Res. Note NE-65. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station; 1967. 7 p.

Seegrist, Donald W. **A multivariate model and statistical method for validating tree grade lumber yield equations.** Res. Pap. NE-320. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station; 1975. 19 p.

Yaussy, D.A.; Brisbin, R.L. **Multivariate regression model for predicting lumber grade volumes of northern red oak sawlogs.** Res. Pap. NE-536. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station; 1983. 11 p.

A multivariate regression model was developed to predict green board-foot yields for the seven common lumber grades processed from northern red oak factory logs. The model may be modified to predict various combinations of lumber grades.

ODC 832.15—525.1-853

Keywords: Log quality; log grades; end product yields

☆U.S. GOVERNMENT PRINTING OFFICE: 1983 605 024 23

Headquarters of the Northeastern Forest Experiment Station are in Broomall, Pa. Field laboratories are maintained at:

- Amherst, Massachusetts, in cooperation with the University of Massachusetts.
- Berea, Kentucky, in cooperation with Berea College.
- Burlington, Vermont, in cooperation with the University of Vermont.
- Delaware, Ohio.
- Durham, New Hampshire, in cooperation with the University of New Hampshire.
- Hamden, Connecticut, in cooperation with Yale University.
- Morgantown, West Virginia, in cooperation with West Virginia University, Morgantown.
- Orono, Maine, in cooperation with the University of Maine, Orono.
- Parsons, West Virginia.
- Princeton, West Virginia.
- Syracuse, New York, in cooperation with the State University of New York College of Environmental Sciences and Forestry at Syracuse University, Syracuse.
- University Park, Pennsylvania, in cooperation with the Pennsylvania State University.
- Warren, Pennsylvania.