1983

Evaluating Nonindustrial Private Landowners for Forestry Assistance Programs: A Logistic Regression Approach

United States Department of Agriculture, Forest Service

United States
Department of
Agriculture

Forest Service

**Intermountain
Forest and Range
Experiment Station
Ogden, UT 84401**

Research Paper
INT-320

October 1983

Evaluating Nonindustrial Private Landowners for Forestry Assistance Programs: A Logistic Regression Approach

Ervin G. Schuster

THE AUTHOR

ERVIN G. SCHUSTER is research forester and project leader with the Forest Economics Research Work Unit at the Forestry Sciences Laboratory, Missoula, Mont. Since receiving his Ph.D. in forest economics from Iowa State University in 1971, he has worked at Weber State College, the University of Idaho, the University of Montana, and with the Intermountain Station. He was responsible for all phases of this study.

RESEARCH SUMMARY

As budgets in forest management agencies become more restrictive, cost-effective programs become more important. This paper describes a quantitative tool for setting priorities for the forestry assistance program administered by the Montana Division of Forestry. Logistic regression was used to better identify the type of forest owners to which assistance should be directed. (In logistic regression, the dependent variable is a probability that a certain event or activity will occur.) Data supporting model development were obtained from a questionnaire survey of forest landowners in the western portion of Montana. Four models were developed that pertain to past use of technical assistance, intention to harvest timber, and timber benefits as motivation for forest ownership. The most consistently useful independent variables were geographic region and past timber harvest activity. The author discusses procedures for interpreting results and for rating land ownerships for assistance. One model is discussed in detail, but the discussion is applicable to the other three models. Supporting data are presented for all models.

Evaluating Nonindustrial Private Landowners for Forestry Assistance Programs: A Logistic Regression Approach

Ervin G. Schuster

INTRODUCTION

Public programs providing technical forestry assistance to owners of nonindustrial forest land have become part of the forest economy in the United States. The Private Forestry Assistance (PFA) program administered by State Foresters (formerly known as the Cooperative Forest Management [CFM] program) along with extension forestry within the USDA Cooperative Extension Service, and to a lesser extent the State and Private Forestry division of the USDA Forest Service, provide the bulk of assistance. Assistance is ostensibly aimed at enabling the landowner to make informed decisions to accomplish personal objectives. Although the programs have multiple-use goals, the landowners' objectives usually favor timber growing, harvesting, and marketing. These programs, therefore, affect timber supply.

Recently, renewed interest in small, privately owned timber holdings coupled with static or declining assistance program budgets have compelled a closer look at the processes by which technical assistance is delivered to forest landowners. Increasingly, assistance must be delivered in a more cost-effective manner. Undersecretary of Agriculture John B. Crowell, Jr., recently spoke of the need to "improve the effectiveness of public programs aimed at encouraging more productive management of nonindustrial, private lands" (speech to the Forest Industries Committee on Timber Valuation and Taxation, Scottsdale, Ariz., November 4, 1982). Traditional programs will not meet that challenge.

Assistance programs would be improved if foresters could identify the landowners who will be most responsive to assistance. Better targeting of efforts and the rating of applicants would help. Given an appropriate data base, a logistic regression model is well suited to this need. This paper reports development of such models for western Montana and use by the Montana Division of Forestry, Department of State Lands. Although a few similar efforts can be found in Eastern States (see for examples Jones and Thompson 1981; Trokey 1981), none are known for the Intermountain West. The technique described in this paper, therefore, has the potential for widespread application.

METHODS

During the late 1970's, the Montana Division of Forestry and the USDA Forest Service undertook a cooperative study of the attitudes and activities of private landowners in Montana. A questionnaire was mailed to a stratified random sample drawn from the listing of forest landowners maintained by the Division of Forestry for use in its fire protection program. Responses from owners of less than 40 acres of forest land and from owners in eastern Montana were eliminated from the data base due to sampling problems. The final 41 percent response rate was explicitly analyzed for response bias; no statistically significant bias was found. Results were published in 1978 (Schuster 1978). The 499 completely usable responses from that study constitute the data base of this present study.

The Montana Division of Forestry requested that the Intermountain Forest and Range Experiment Station reanalyze data from the earlier study. The new objective was to develop information and relationships that would enable service foresters to better identify landowners that not only wanted and needed technical assistance, but who would also be likely to use or apply the assistance provided. Unfortunately, the latter question was not addressed in the original questionnaire.

Specific questions in the following categories were selected from that survey as the best indicators of landowner desire for and acceptance of technical assistance:

- Landowner's previous use of forestry assistance, either public assistance or private consultant.
- Landowner's stated intention to harvest timber products in the future.
- Landowner's stated reasons for owning forest land related to production of timber products.

The first category was selected because it obviously and explicitly deals with using technical assistance. The latter two categories were included because of the strong timber and wood products orientation of participants in assistance programs. Although the specific questions were linked to the assistance program, each stands alone and may be used to assess other issues. Responses to selected questions from these areas were used to represent the dependent (Y) variables, the variables to be

predicted in this research. Note that these variables were not modeled to predict behavior of landowners who responded to the original survey. Rather the purpose is to model responses from previous participants as an indication of behavior of other landowners.

The questionnaire also contained information about landowners and their forest holdings that would be useful in predicting the key indicators of landowner response to assistance:

- Ownership-size class.
- Timber-size class.
- Previous timber harvest activity.
- Landowner age.
- Landowner education.
- Landowner income.
- Landowner occupation.
- Geographical location of forest land.

This list represents potential independent (X) variables.

Two analytical techniques are particularly well suited to the type of prediction needed in this research—the discriminant function and the logistic function. The difference can be illustrated with the question: Will a specific landowner use technical forestry assistance? Given measurements on the independent (X) variables reflecting landowner characteristics, the discriminant function will predict an outcome (the Y) as being either yes or no. Given the same set of measurements, the logistic function will predict the numerical probability. For example, given a set of landowner characteristics, the discriminant function might predict an outcome of "no," will not use assistance; whereas the logistic function might predict the outcome as 0.15, a 15 percent probability that assistance will be used. The logistic function, sometimes referred to as a "logit model," was judged more suitable for this study.

The logistic function resembles a typical multiple linear regression function, but also differs from it. Three aspects warrant mention. First, while the multiple regression function is a linear function, the logistic function is nonlinear. Second, in the case of multiple regression, the statistical model is of the form:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + ... + \beta_iX_i \qquad (1)$$

The regression coefficients (β's) show the linear relationship between the independent variables (X_i) and the dependent variable (Y). A logistic regression model instead estimates a probability. This is done by means of the ratio of natural logarithms:

$$P(E) = \frac{e^{Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + ... + \beta_iX_i}}{1 + e^{Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + ... + \beta_iX_i}} \qquad (2)$$

All symbols are as before, except for P(E), the probability of an event, which lies between 0.0 and 1.0; and "e", the base of natural logarithms, which is approximately 2.718. Third, interpretation of the regression coefficients is different. In the case of multiple linear regression, each coefficient (β_i) can be directly interpreted as the effect of a unit change in X_i on Y, when all other variables are held constant. For the logistic model, β_i represents the effect of a unit change in X_i on the exponent of "e". This attribute makes it somewhat more difficult to interpret coefficients. For a more complete discussion of the logistic function, see Pindyck and Rubinfeld (ch. 10, 1981).

Study data were analyzed by means of the Stepwise Logistic Regression feature of BMDP Statistical Software (Dixon 1981). Each dependent variable was transformed to take on only 0 or 1 values. All independent variables were formulated in terms of categories or classes. For example, the variable, landowner age, has three classes, one of which is 65 years and older. All dependent variables together with their class designations are shown later as part of table 2.

Initial model construction involved unrestricted entry and exit of variables until no additional variable could achieve statistical significance, based on the F statistic with $\alpha = 0.10$. Many sets of observations (cases) had "missing" values for one or more independent variables (some respondents did not answer some questions in the original questionnaire). Because the computer program automatically excluded any case with missing values, effective sample size was frequently reduced to about 300. Final model construction involved refitting all data to models containing only the statistically significant variables identified in initial model construction; the stepwise procedure was not used. This increased effective sample size from 300 to between about 350 and 500 cases.

Traditional statistical measures of model goodness, such as R^2, are not very useful to assess logistic regression models. Rather, their overall ability to correctly predict the event being studied, for example as reflected by Chi-square, is a more useful measure. This aspect will be discussed along with other study results.

RESULTS

This study estimated four logistic regression models whose dependent variables had been identified as being important to administration and implementation of the Private Forestry Assistance program in Montana. Estimates for dependent variables should be interpreted as the probability of a landowner behavior event occurring (P[E]). The four landowner events studied pertain to:

E1. **Using** the services of a Private Forestry Assistance (PFA) forester.

E2. **Using any** technical assistance services, either from a PFA forester or a private forestry consultant.

E3. **Harvesting** timber from forest land at any future time.

E4. **Currently** owning forest land either for **timber** production (income from the growth and sale of timber or other forest products) or for farm or domestic use (source of forest products for own use—firewood, fenceposts, etc.).

Logistic Models

Although the specific details of the four logistic regression models are different, the form of the results and their interpretation process are identical. Additionally, some models are sufficiently complex so that narrative presentation is too cumbersome. For these reasons, results for only the first (E1) model, using a PFA forester, will be presented. But the discussion also applies to the other models. Data needed to interpret those models will be displayed in tables and figures.

The likelihood of a forest landowner using the services of a PFA forester was found to significantly vary as a function of size of ownership and region of location. The region variable has three class categories: northwest, southwest, and central, as displayed in figure 1. The ownership size variable also has three classes: 40-159 acres (16.2-64.3 ha), 160-639 acres (64.8-258.6 ha), and 640 or more acres (259.0 or more ha). Other factors (tree size, owner age, income, etc.) probably influence use of PFA, but did not increase predictability by a statistically significant amount over ownership size and region.

Overall, only about 18.8 percent of western Montana forest landowners have used the services of a PFA forester, but substantial differences exist between regions and size-classes. Table 1 shows the effect of these differences and the probability of using the PFA program. There is a pronounced regional effect wherein, regardless of size-class, landowners in the southwest region have a higher probability of use than in the northwest and both greatly exceed the central region. Similarly, owners in the middle size-class, independent of region, have the highest probability of use; the smallest size-class has the lowest. Consequently, middle size-class owners from the southwest region have the highest probability of use, while central region owners in the smallest size-class have the lowest use probability.

Table 1.—Probability of Montana forest landowner using services of a PFA forester, by region and size-class

Size-class	Region		
	Northwest	Southwest	Central
Acres			
40-159	0.127	0.150	0.037
160-639	.271	.311	.089
640 +	.264	.303	.086

The probabilities of using the PFA program (E1) are easily displayed. Only two independent variables were statistically significant, each with three classes or categories. Results could be displayed in a 3 X 3 table. But those are the **only** easily displayed results. Table 2 shows all logistic regression models and information pertaining to the statistically significant variables. Rather than presenting a series of complex tables to display results, the process of computing the probabilities shown in table 1 from the data for the equivalent model (E1) in table 2 will be fully explained. Probabilities analogous to those in table 1 could easily be computed for any model, as desired by the reader.

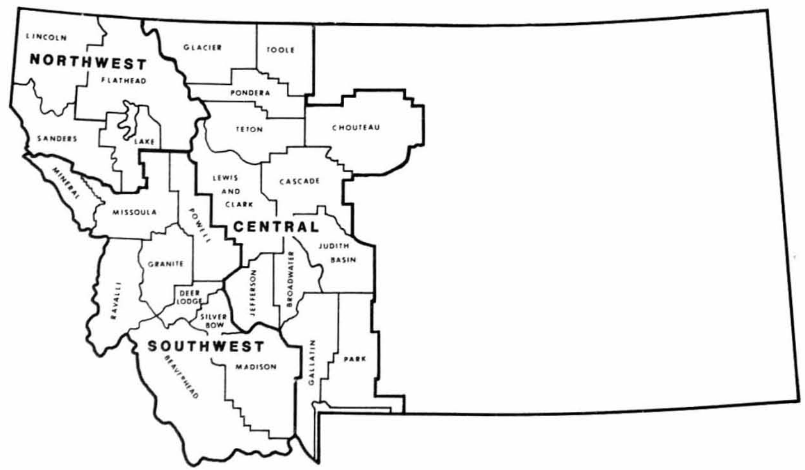

Figure 1.—Three geographical regions within Montana.

Table 2.—Partial exponents for logistic regression models

Variable or factor	Classes or categories	E1	E2	E3	E4
		Using PFA	Using any	Harvest any	Current timber
Constant	—	-1.694	-2.487	1.801	-1.667
Ownership size	40-159 acres	-.614	-.564	-.585	—
	160-639 acres	.326	.259	.338	—
	640 + acres	.288	.305	.247	—
Timber size	≤5 inches	—[1]	—	—	-.002
	5-9 inches	—	—	—	-.316
	≥9 inches	—	—	—	.318
Prior harvest	Yes	—	.552	.859	1.197
	No	—	-.552	-.859	-1.197
Age	44 years and under	—	—	.484	—
	45-64 years	—	—	.446	—
	65 years and older	—	—	-.930	—
Education	1-8 years	—	—	—	-.400
	9-12 years	—	—	—	1.003
	Post-high school	—	—	—	-.646
	Bachelors degree	—	—	—	.701
	Postgraduate	—	—	—	-.659
Occupation	Professional	—	1.790	—	.580
	Administrative	—	.473	—	.232
	Sales	—	.781	—	-1.759
	Crafts	—	1.185	—	-1.021
	Operator	—	.240	—	-.573
	Laborer	—	-8.092	—	.069
	Farmer/Rancher	—	1.404	—	.636
	Retired	—	.936	—	-.261
	Other	—	1.285	—	2.100
Region	Northwest	.380	.465	—	.965
	Southwest	.573	.445	—	-.030
	Central	-.953	-.910	—	-.935

[1]Dashes (—) indicate variable as not statistically significant.

Data contained in table 2 are a condensed form of the logistic regression models, the coefficients and associated design matrixes. They constitute the contribution of each category to the logistic model. Consider the probability of 0.311 shown in table 1 for southwest region owners in the 160-639-acre (64.8-258.6-ha) size-class. Refer now to the data in table 2 for the appropriate variable category pertaining to the E1—Using PFA model:

Constant.. (-1.694)
Size—160-639 acres .... (0.326)
(64.8-25.8.6 ha)
Region—southwest .... (0.573)

The numbers (-1.694, 0.326, and 0.573) are used to quantify $\hat{Y}$ in equation 2:

$$P(E1) = \frac{e^{\hat{Y}}}{1 + e^{\hat{Y}}}$$

Simply determine $\hat{Y}$ as the sum:

$$\hat{Y} = -1.694 + 0.326 + 0.573$$
$$= -0.795$$

Hence:

$$P(E1) = \frac{e^{-0.795}}{1 + e^{-0.795}}$$
$$= 0.311 = 31.1 \text{ percent}$$

Similarly, to estimate the probability of using PFA by central region owners in the smallest size-class:

$$P(E1) = \frac{e^{\hat{Y}}}{1 + e^{\hat{Y}}} \quad \hat{Y} = -1.694 + (-0.614) + (-0.953) = -3.26$$
$$\hat{Y} = -3.26$$
$$= 0.037 = 3.7 \text{ percent}$$

Probabilities for other events, E2-E4, are determined by using the procedure just described. It is important that each significant variable have a coefficient in the summation.

For some purposes, it may not be necessary to calculate probabilities with great precision. An approximation will be sufficient. Table 3 provides a listing of $\hat{Y}$ values together with the associated probability of event values. Consider the case of 160-639-acre (64.8-258.6-ha) owners in the southwest region where $\hat{Y}$ = -0.795. Inspection of table 3 shows -0.8 to be the closest $\hat{Y}$ value; its associate P(E) is 0.310 which corresponds to 0.311 shown above. Since all probability values can be calculated exactly, use of table 3 is optional.

Table 3.—Probability of events for corresponding values of $\hat{Y}$

$\hat{Y}$	P(E)
-10.0	0.000
- 8.0	.000
- 6.0	.003
- 5.5	.004
- 5.0	.007
- 4.5	.011
- 4.0	.018
- 3.5	.029
- 3.0	.047
- 2.5	.076
- 2.0	.119
- 1.8	.142
- 1.6	.168
- 1.4	.198
- 1.2	.232
- 1.0	.269
- 0.8	.310
- 0.6	.354
- 0.4	.401
- 0.2	.450
0.0	.500
0.2	.550
0.4	.599
0.6	.646
0.8	.690
1.0	.731
1.2	.769
1.4	.802
1.6	.832
1.8	.858
2.0	.881
2.5	.924
3.0	.953
3.5	.971
4.0	.982

Data contained in table 2 can also be used less analytically. The numbers themselves indicate relative importance in determining the probability of an event occurring. The bigger the number, the larger the effect on probability. Consider the E1 model. The smallest numbers in table 2 are associated with the 40-159-acre (16.2-64.3-ha) size-class and the central region. Both have relatively large negative values (-0.614 and -0.953 respectively). Table 1 shows these categories have lower probabilities and when combined constitute the lowest probability. Conversely, the highest probabilities in table 1 are for the middle size-class and the southwest region, variable categories with the largest values in table 2. Table 2 values should be compared within a column, not between columns. Table 2 values contribute to the size of the exponent ($\hat{Y}$); therefore effect on probability is not proportional to size.

The quality of a logistic regression model is determined by its ability to predict outcomes correctly. Moreover, the goodness of predicted probabilities can be verified only in the context of a large number of prediction opportunities. Although the probabilities can be applied to a specific forest landowner, evaluation of the probabilities estimated by logistic regression is best done by reference to the combined outcomes over many individuals. Table 4 compares predicted and actual percentages of landowners using the PFA program, both derived from the study data base.

Table 4.—Actual and predicted participation (percentage) in PFA programs by size-class and region

Participation				
Actual	Predicted	Size-class	Region	Sample size
Percent		*Acres*		
5.7	3.7	40 - 159	Central	35
6.5	8.9	160 - 639	Central	31
8.8	8.6	640 +	Central	34
12.2	12.7	40 - 159	Northwest	90
14.6	15.0	40 - 159	Southwest	82
26.8	27.1	160 - 639	Northwest	82
27.9	26.4	640 +	Northwest	43
28.6	30.3	640 +	Southwest	42
33.3	31.1	160 - 639	Southwest	45

Consider the case of northwest region landowners in the 40-159 acre (16.2-64.3 ha) size-class. The logistic regression model predicted that 12.7 percent of the 90 landowners would use the PFA program. In fact, 11 of the 90, amounting to 12.2 percent, of the landowners did. Comparisons between predicted and observed participation for the other logistic regressions, E2-E4, are very similar to those shown for E1, but are too complicated to present here. The E1 logistic regression model yielded the best predictions; the E2 model the worst predictions, based on a Chi-square analysis.

Cut-off Points

Although the overall accuracy of the logistic regression models are revealed in the context of a large number of landowners, their application in forestry assistance is to set priorities for the assistance program. A rule or cut-off point must be established by which a class of landowner (or classes) is judged a likely (good) or unlikely (bad) prospective client for the assistance program. One must establish a probability level (P[E]) above which an associated class of landowner (or classes) is judge to be "likely" clients, below which judged "unlikely." In reality the cut-off point would be used as a guideline, a "screening" device to separate the likely from the unlikely prospects. The PFA forester can then focus time and attention on the likely prospects, deemphasizing or screening out the unlikely.

Unfortunately, although individual landowners within a class have similar characteristics (region and ownership size), they do not always behave alike. Any cut-off point will result in errors. A trade-off exists between correctly identifying those landowners for which an event (using PFA services) will occur and correctly identifying those for which the event will not occur. (This problem is analogous to the statistical problem of a Type I and Type II error.) Figure 2 graphically depicts these trade-offs for each of the four logistic models being presented. Shown in the frame pertaining to the E1 model, a lower cut-off point (say P(E1) = 0.05) will correctly identify all individuals that did use PFA forester services (coded "Using PFA"). But it fails miserably at identifying those that did not (coded "Not Using"). If the cut-off point were set at P(E1) = 0.05, then any class of landowners where the associated P(E1) were greater than 5 percent would be judged as likely clients. The cut-off point would correctly identify about 95 percent of the users, but only about 10 percent of the nonusers. That

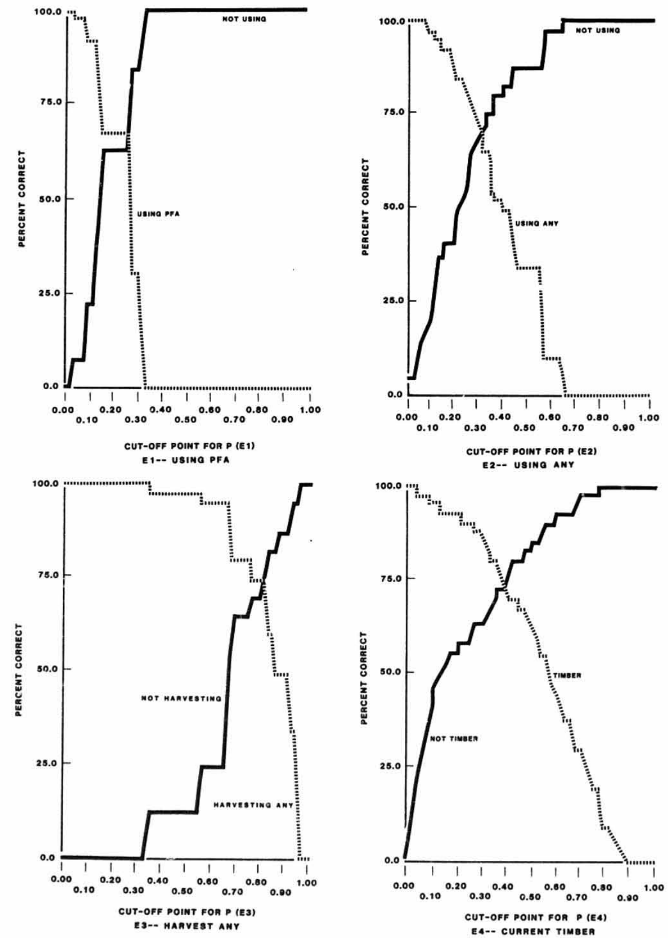

Figure 2.—Cut-off point analyses for logistic regressions

is, about 90 percent of the nonusers would be mistakenly identified as users.

The low cut-off level screens nobody and leads to the conclusion that virtually everyone will use PFA services. On the other hand, a high cut-off level leads to the conclusion that nobody will use the services of a PFA forester. Figure 2 shows that if the cut-off level were set anywhere between a probability value (P[E1]) of 15 to 25 percent, about 67 percent of the users and about 62 percent of the nonusers would be correctly identified. Below that level users would be better identified, but many nonusers would be misidentified as likely users. Conversely, above that level, likely users would increasingly be misidentified as likely nonusers.

Cut-off points must be decided for all logistic regressions (E1-E4) presented in this paper. The probability level selected should relate to the consequences (or costs) of misidentification. For example, if it is very important to correctly identify all likely PFA program users and it is not particularly costly to identify nonusers as users, a low cut-off point is appropriate. Alternatively, if the capability to provide assistance is limited such that correct identification of unlikely clients is critical, a relatively high cut-off point is appropriate. The four frames of figure 2 provide the information for determining cut-off points for all (E1-E4) logistic regression models.

DISCUSSION AND APPLICATIONS

This paper presents results from logistic regression models pertaining to landowner use of the forestry assistance program provided by the Montana Division of Forestry, Department of State Lands. Unfortunately, the available data base did not exactly address that topic. Several surrogate models were developed, each of which only partially related to desired topic. A reasonably simple model, E1—Using PFA, was used to illustrate how questionnaire-type data can be easily converted to probability estimates. The interpretive approach shown for that model should be applied to the other models, as dictated by the user needs. Additionally, users such as the Montana Division of Forestry will have to evaluate these results and develop guidelines for application. Questions must be addressed. For example: Which model or models (E1-E4) should be emphasized and what cut-off points are appropriate?

Assume, for example, that a judgment is made to use the E1—Using PFA as the primary model and E4—Current Timber as the secondary model. Further assume such shortage of funds that it is more critical to screen-out unlikely clients than to correctly identify all likely clients. A relatively high cut-off point would be appropriate. If figure 2 were used to set the cut-off point at 0.28, about 85 percent of the unlikely clients (nonusers) would be screened, but only about 30 percent of the likely clients (users) would be identified. That a higher percentage of likely clients was not identified might be judged acceptable under circumstance of an extreme funding shortage. Table 1 shows that only southwest region landowners in the two largest size-classes meet that standard.

If further restrictions are needed, the E4—Current Timber model could be used analytically, as was the E1 model, or nonanalytically. Inspection of the E4 portion of table 2 for the largest exponents, shows that landowners that have harvested timber, that have 9 to 12 years of education, and that have an "other" occupation will have a relatively high probability of owning forest land for timber production.

The procedure described offers a system for establishing the top priority landowner group wherein assistance would be targeted. Subsequent analysis could be used to develop a more comprehensive priority listing, as needed. The technique described in this report is not highly refined, but it does illustrate the application of logistic regression to the problem. If the general approach is deemed useful, the data base could readily be improved.

PUBLICATIONS CITED

Dixon, W. J., ed. BMDP statistical software. University of California Press; 1981. 725 p.

Jones, J. Greg; Thompson, Richard P. Characteristics, attitudes, and objectives of nonindustrial private forest owners in eastern Oklahoma. Res. Rep. P-816. Stillwater, OK: Oklahoma State University, Agricultural Experiment Station; 1981. 32 p.

Pindyck, Robert S.; Rubinfeld, Daniel L. Econometric models and economic forecasts. 2d ed. New York: McGraw; 1981. 630 p.

Schuster, Ervin G. Attitudes and activities of private forest landowners in western Montana. Rep. R1-78-03. Missoula, MT: U.S. Department of Agriculture, Forest Service, Northern Region; 1978. 76 p.

Trokey, Carol B. An analysis of timber management potentials for private nonindustrial forest landowners. Columbia: University of Missouri; 1981. 122 p. M. S. thesis.

Schuster, Ervin G. Evaluating nonindustrial private landowners for forestry assistance programs: a logistic regression approach. Res. Pap. INT-320. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station; 1983. 7 p.

Four logistic regression models were developed from questionnaire data obtained from forest landowners. The models were designed to assist the Montana Division of Forestry to rate clients for the forestry assistance program. Interpreting and applying results is discussed. Basic data are presented.

KEYWORDS: forestry assistance, forest landowners, logistic regression, logit analysis

The Intermountain Station, headquartered in Ogden, Utah, is one of eight regional experiment stations charged with providing scientific knowledge to help resource managers meet human needs and protect forest and range ecosystems.

The Intermountain Station includes the States of Montana, Idaho, Utah, Nevada, and western Wyoming. About 231 million acres, or 85 percent, of the land area in the Station territory are classified as forest and rangeland. These lands include grasslands, deserts, shrublands, alpine areas, and well-stocked forests. They supply fiber for forest industries; minerals for energy and industrial development; and water for domestic and industrial consumption. They also provide recreation opportunities for millions of visitors each year.

Field programs and research work units of the Station are maintained in:

Boise, Idaho

Bozeman, Montana (in cooperation with Montana State University)

Logan, Utah (in cooperation with Utah State University)

Missoula, Montana (in cooperation with the University of Montana)

Moscow, Idaho (in cooperation with the University of idaho)

Provo, Utah (in cooperlation with Brigham Young University)

Reno, Nevada (in cooperation with the University of Nevada)

U.S. GOVERNMENT PRINTING OFFICE: 1983—776-032/1047 REGION NO. 8