United States
Department
of Agriculture

Forest Service

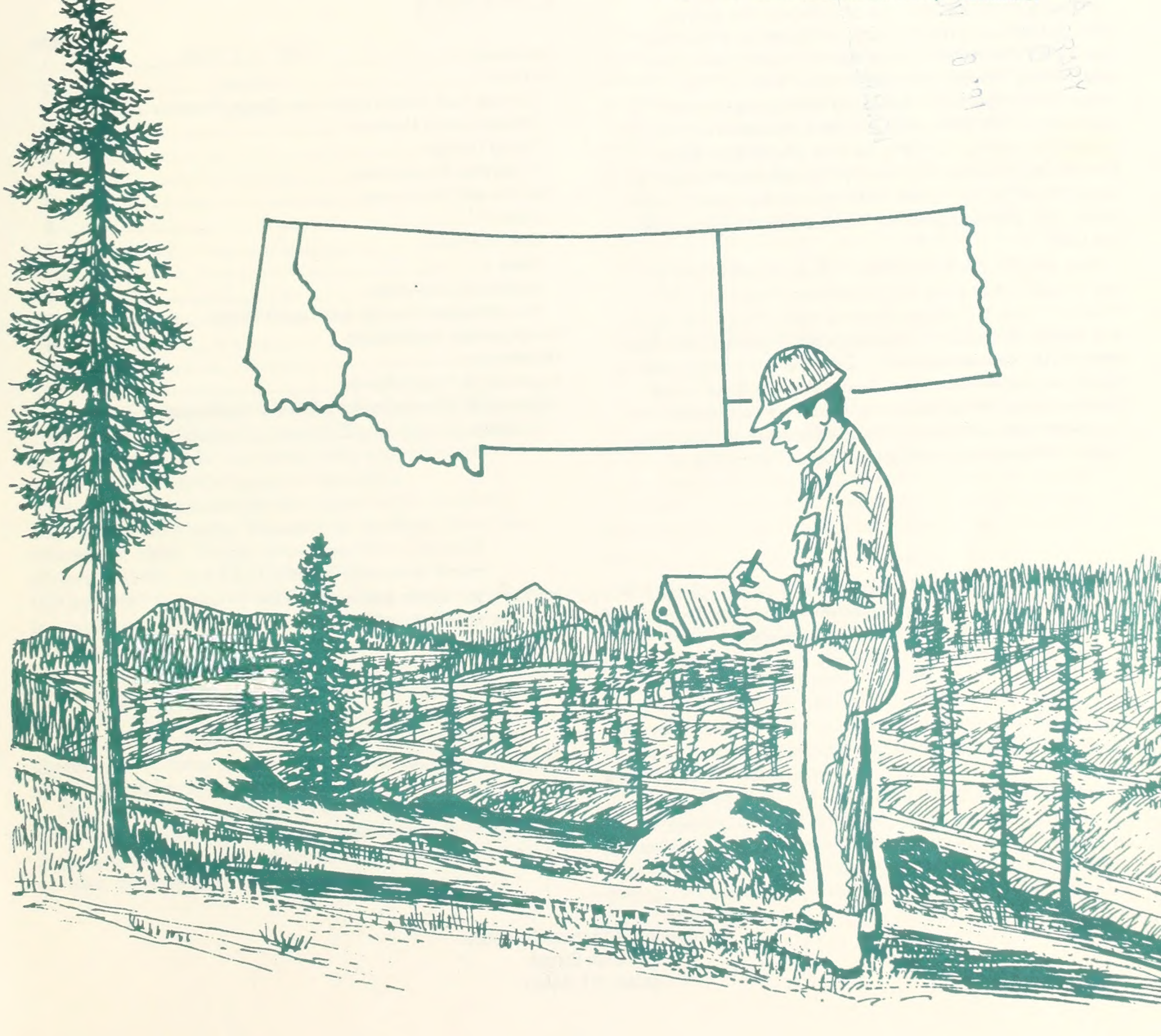
Intermountain
Research Station

Research Paper
INT-418

December 1989

Predicting Salability of Timber Sale Offerings in the Forest Service Northern Region

Michael J. Niccolucci



THE AUTHOR

MICHAEL J. NICCOLUCCI is an economist with the Economics Research Work Unit, Intermountain Research Station, Forestry Sciences Laboratory, Missoula, MT. He received academic training in economics at the University of Montana, where he earned his M.A. His research includes timber sale design and appraisal.

RESEARCH SUMMARY

Timber sale planning is a complex, expensive process. Developing a sale to the point where it is ready for auction requires the efforts of many natural resource specialists, many hours, and many dollars. The common expectation is for the sale to sell at initial auction. But many timber sales receive no bids, meaning they do not sell at their initial offering. Given the large investment involved, the occurrence of unsold sales is not desirable. Unsold sales also raise the question of organizational competence.

Knowing the likely outcome of a timber sale offering, particularly in early design stage, is important to the manager. This information can be used to modify the timber sale thereby increasing its likelihood of selling. The research reported here developed and compared two approaches to statistical classification, intended to predict salability at various points in the sale planning process. Classification results were statistically compared based on geographical zone models, models at various points in the timber sale planning process, and the classification methods used.

Data used in this study came from a sample of 389 sold and unsold timber sales in the Northern Region of the Forest Service, U.S. Department of Agriculture. The region was further divided into two geographical zones—east and west of the Continental Divide. Discriminant analysis and logistic regression were used to develop statistical equations to classify timber sales into groups of sold and unsold. Equations were designed to be used at three points in the "Gates" timber sale planning process, all before the actual bidding. The quantity and quality of information increases as the sale proceeds from one gate to another.

The accuracy of the equations increased as the timber sale progressed through the Gates process. Equations at the first gate, approximately 7 to 10 years before the auction date, correctly classified about 65 percent of the sales, based on only general sale characteristics. The equations for the next gate, 1 to 3 years before the auction, correctly classified 74 percent of the sales. Equations at the final gate before the auction correctly classified about 84 percent of the sales.

Statistical analyses were conducted to test for differences in classification success between geographical zones within the Northern Region, between statistical modeling techniques, and between phases in the timber sale planning process. Statistically significant results were found between geographical zones and sale phases, but no statistical difference in classification success could be found between statistical modeling techniques.

CONTENTS

Intermountain Research Station
324 25th Street
Ogden, UT 84401

Predicting Salability of Timber Sale Offerings in the Forest Service Northern Region

Michael J. Niccolucci

INTRODUCTION

Timber sales play an important role in achieving forest management goals and objectives—providing wood raw materials, nontimber outputs, and government revenue. The sale is planned by foresters, engineers, and other specialists. It takes approximately 2 to 10 years to develop a timber sale, depending upon sale size and complexity. Once the timber sale plan is finalized, it is offered to the public for purchase, fully expecting the sale to sell at its initial offering.

But sometimes a timber sale offering receives no bids. The sale remains unsold. During 1980 through 1985, approximately 20 percent of volume offered for sale received no bid. An offering that fails to sell, given all of the hours of planning needed, can be viewed as a waste of dollars and human resources. Even if revised, reoffered, and later sold, these unsold offerings may be obviously seen as evidence of poor planning, professional incompetence, or timber supply exceeding demand. Revisions require added expense and still do not guarantee a more appealing offering.

Forest Service managers need reliable, defensible tools to better predict salability. Also, given the years needed to plan a sale, the sooner salability is known, the better. Timber sale specialists can then either modify the sale design or simply offer the timber sale without further modification in light of expected salability.

Currently, Forest Service managers in the Northern Region have been using Transaction Evidence Equations (Merzenich 1985), Timber Sale Feasibility Analysis (Peterson 1980), and DLOGPRICE Economic Model (Artley 1986) to quantify salability. Transaction evidence is a multiple regression approach used to predict stumpage value. Because these equations are typically based on sold sales only, they cannot be used to make reliable statistical predictions regarding unsold sales. The sale feasibility and DLOGPRICE models both rely on the calculation of a value-cost ratio. Because the statistical significance of these ratios cannot be tested, these procedures cannot be rigorously defended.

The research being reported here was designed to develop and compare two statistical approaches to predicting timber salability at various points in the sale planning process. Several questions were addressed:

1. Is one approach consistently better than the other?
2. Does our ability to predict salability improve as the sale gets closer to the auction date?
3. Does geographical zone variation affect salability prediction?

METHODS

Methods used in this study fundamentally reflect research choices about statistical classification methods and how the timber sale planning process is envisaged. The planning process provided this study a timeframe within which to analyze timber sales. Fortunately, the Forest Service currently uses a planning process that contains the desired timeframe. Because a timber sale selling or not selling is a dichotomous event, the desired statistical technique should be capable of classifying events into those classes. Logistic regression and discriminant analysis were the classification methods selected.

Timber Sale Preparation—The Gates Process

The Forest Service currently uses a planning and decision making process called "Gates" to design timber sales (USDA FS 1985, 2431.2). This process encompasses a series of activities that begins with the identification of a sale area and ends with a sale award. Sale planning activities must pass through six reporting points, called "gates."

Gate 1: Sale preparation begins. This entails identifying the purpose and the need for the sale; identifying public issues; identifying the resource opportunities in the sale area, and so on. Preliminary sale volume and sale area estimates are also produced.

Gate 2: Alternative sale area designs are developed. Environmental effects are analyzed and a preliminary economic analysis is completed. Gate 2 results in selection of a preferred alternative.

Gate 3: The activities leading to sale plan implementation are performed. Preparation of the contract, data gathering, and the necessary outline to support the appraisal are examples of these activities. The sale passes through gate 3 when the fieldwork and the timber sale report are completed.

Gate 4: Necessary engineering, logging, and environmental cost information is gathered. Timber value is set, the appraisal is prepared, and the total sale package is reviewed.

Gate 5: Bids are accepted and the successful bidder is determined. The output is the bid report.

Gate 6: The winning bidder is evaluated with respect to financial qualifications, Equal Employment Opportunity clearance, and so on. The timber sale passes through gate 6 when all requirements are met; the award of the contract is the output.

The "gates" are important to this study because they not only progress temporally toward the actual implementation of the timber sale, but they also depict increasing quantity and quality of information that can be utilized by statistical classification models. Because gates 5 and 6 occur after the sale is sold (or not sold) they are of no use in predicting salability and will not be considered further.

At gate 1 an area is brought into the planning process through development of a position statement—a document that is a prerequisite to listing a proposed timber sale on the timber sale action plan (USDA FS 1985, 2414.27). At this point, 10 years from the auction date, very little site-specific information is known. Examples of information known at this gate are slope, elevation, and acreage within the proposed sale area. Over the long time span of sale development many external influences may alter salability.

Gates 2 and 3 are closely related and will be treated as a single, composite gate. They deal with developing a sale area design and preparing for sale plan implementation. At these gates specific sale characteristics are developed. Sale characteristics include number and size of the cutting units, the volume-per-acre harvested, the miles of road construction, the silvicultural systems needed, the logging method required, and so forth. Gates 2 and 3 occur about 1 to 3 years before the auction date.

Gate 4, the final gate for predicting salability, completes the package by generating the appraisal. At this gate, the planner's sale design decisions are converted into appraisal information—dollars per thousand board feet. Information generated at this gate includes stump-to-mill costs and the advertised selling rate. Gate 4 occurs about 3 months before the auction.

Classification Methods

The major factor affecting selection of statistical classification methods is the dichotomous nature of the dependent variable. In this problem, the dependent variable takes on two values (0 = unsold, 1 = sold) and identifies group membership. For this class of problem, potentially useful methods are limited to regression analysis, discriminant analysis, and logistic regression. The method of regression analysis was discarded because of the potential violations of certain key assumptions, principally the variance of the error term is not constant for all observations, and the predicted values are not guaranteed to lie in the (0, 1) interval (see Pindyck and Rubinfeld 1981). The methods of logistic regression and discriminant analysis seem well suited to the problem of predicting salability.

LOGISTIC REGRESSION

Logistic regression relates a qualitative dependent variable, such as "sold" or "unsold" timber sales, to independent predictor variables through a cumulative logistic probability function (see Maddala 1983; Pindyck and Rubinfeld 1981). Parameter estimation is based on maximum likelihood estimation. These estimates have several desirable properties, such as all parameters being consistent and efficient asymptotically (Pindyck and Rubinfeld 1981). All parameter estimators are known to be normal, therefore the t-test can be applied to test for significance. Also, research has shown that if certain discriminant function assumptions are violated, the logistic regression provides better prediction results (Press and Wilson 1978).

The logistic regression predicts a probability of an event occurring. The general model is specified as:

$$\text{Probability}_i = \frac{e^Y}{1 + e^Y} \tag{1}$$

Probability_i is the probability of an event occurring (sale selling), e is the base of natural logarithms (approximately 2.718), and Y is estimated:

$$Y = B_0 + B_1X_1 + B_2X_2 + \ldots + B_iX_i + E_i. \tag{2}$$

Equation 2 above is presented in this research. To predict a probability that an event will occur, you must first calculate a Y and then substitute that value into equation (1).

Given the predicted probability that an offering will sell, a decision rule needs to be adopted to perform classification. The common decision rule is based on a probability of one-half. If the probability is greater than or equal to 0.50, the sale is predicted to be a sold sale. If the probability is less than 0.50, the sale is predicted to be an unsold sale. This specific decision rule will generate a specific classification result.

Most of the following results are based on the 50 percent decision rule discussed above. The effect of changing this rule is also explored.

DISCRIMINANT ANALYSIS

Discriminant analysis is the traditional classification technique. The basic strategy in discriminant analysis is to form an equation (discriminant function) that uses independent variables to classify observations into designated groups (sold and unsold timber sales). The discriminant function (equation) has the following form:

$$L = B_1X_1 + B_2X_2 + \ldots + B_iX_i \tag{3}$$

where L is the dependent variable (sold and unsold sales) and X_1 through X_i are independent variables. The common expectation is the production of a discriminant function (equation) that optimally separates the designated groups.

The linear discriminant function and its classification rule depend on several statistical assumptions (Maddala 1983). If the assumptions are violated, corrective procedures are recommended (Johnson and Wichern 1982; Meddala 1983; Press and Wilson 1978). The linear discriminant functions were tested for compliance with the statistical assumptions and corrective procedures applied when necessary.

Classification can be achieved using either Fisher's linear discriminant function (equation presented above) or classification functions (see Morrison 1976 for a discussion on classification functions—classification functions presented in tables 10 through 15, appendix B). Classification depends on the calculated discriminant score and a critical value. If the calculated discriminant score is greater than zero (the critical value) it is assigned to the sold sales category, and if less than zero to the unsold sales category.

Study Design

The Gates timber sale planning process provided the framework for this analysis. As the sale progresses from gate 1 to gate 4, site, sale, appraisal, and economic information is generated. This information was used to develop the equations. The gate 1 equation is based on the general site characteristics known at that time in the planning process. The gate 2-3 equation is based on information from gate 1 (site characteristics) plus information from gates 2 and 3 (sale characteristics). The third equation, based at gate 4, uses all prior information plus the appraisal information generated at this gate. If one chooses to use measures of economic expectations as independent variables, gate 4 would be the appropriate gate. Estimates of economic expectations were not used.

I analyzed timber sales auctioned during January 1980 through December 1985 on National Forests in the Northern Region of the Forest Service. All timber sales were competitively auctioned and were of at least $2,000 minimum value. The sampling period consisted of the high markets of the early 1980's, a major recession in 1982 and 1983, and a recovery during 1984 and 1985. Given that the major thrust of this research was the development and comparison of methods that quantify salability, the sample diversity was not a hindrance. But the equations should be used only to predict salability for sales that will be competitively auctioned.

I randomly sampled 389 sold and unsold timber sales from the 13 National Forests in the Northern Region. Of the 389 sales, 349 timber sales were sampled on the "westside" National Forests (Bitterroot, Lolo, Flathead, Kootenai, Idaho Panhandle, Clearwater, and Nez Perce). Of the 349 westside timber sales, 204 were sold timber sales; the remaining 145 timber sales were unsold. I sampled 40 timber sales on the "eastside" National Forests (Custer, Gallatin, Beaverhead, Helena, Lewis & Clark, and Deerlodge). They were composed of 26 sold and 14 unsold sales.

Analytical Procedures

MODEL CONSTRUCTION

The equations were checked for violations of the assumptions and were corrected if appropriate. Given the empirical nature of equation development, Wilk's lambda (stepwise procedure) was used to generate the final discriminant functions. For logistic regression, an "all-possibles" regression subroutine based on ordinary least squares was used to develop preliminary equations. Of course, the timber sale and economic characteristics used were considered logical variables in determining sold and unsold timber sales.

Both statistical classification methods were developed using SPSSX (Nie 1983) and BMDP (Dixon 1981) statistical software on a Vax 8600 computer.

EVALUATION CRITERIA

Using either the classification functions or Fisher's linear discriminant function, I produced a classification matrix that presents the predicted classification results. The classification matrix is a tabular method used to present the percentage correctly and incorrectly classified. Another method used to measure goodness of fit is the Holdout method (Lachenbruch 1975), also known as the Jackknife method. This method predicts the group to which a sale belongs when the sale is not involved in the model-building process. This technique generates a better estimate of the error variance than the classification method described above. Both of these methods are used to produce classification results for the discriminant functions.

The logistic regression equation predicts the probability of a sale being sold given sale attributes and market information. Since it predicts a probability, we must adopt a rule that aggregates these probabilities into groups. The decision rule adopted aggregated sales with predicted probabilities greater than or equal to 0.50 into the sold sales group. If the predicted probability is less than 0.50, the sale is predicted to be an unsold timber sale. Once the decision rule was implemented, a classification matrix was constructed, indicating how well the equation predicted sold and unsold timber sales.

CUTOFF POINTS

Particularly important is the link between probability levels, classification results, and the cost of making a decision error. The decision rule of 0.50 probability implies the costs of misclassifying sold and unsold sales is equal. But it may be more costly from a managerial standpoint to classify a sale as salable, when actually it will not sell.

Analysts and decision makers are free to adopt any decision rule. A different decision rule will lead to different classification results. Decision rules imply an underlying cost of making a decision error. These costs are not consistent from user to user. A particular land manager may view misclassification of a predicted sold sale more costly than misclassification of a predicted unsold sale. The land manager's cost of misclassification will lead to a particular decision rule, and thus, different classification results.

Seven cutoff points (probability levels) were chosen to investigate the changes in classification results: 0.20, 0.25, 0.33, 0.50, 0.67, 0.75, and 0.80. The posterior probabilities were used as the discriminant model's predicted probabilities of a sale selling. The posterior probabilities provide a continuous prediction of salability and allow the cutoff points (decision rule) to vary at the levels defined previously.

As stated previously, the logistic regression predicts a probability, and therefore adapts easily to this analysis.

CATEGORICAL ANALYSIS OF VARIANCE

Categorical analysis of variance (Bishop and others 1975) was used to analyze the dependent variable, percentage correct (*%c*), based on independent variables, classification method (*cm*), gates (*g*), and geographical zones (*z*). The results of this analysis statistically quantified the benefits derived by having employed different classification methods, gates, and geographical zones.

The mathematical equation is:

$$\%c_{ijk} = M + cm_i + g_j + z_k + I_{ijk} + e_{ijk}$$

where I_{ijk}, e_{ijk}, and M are the interaction term, error term, and the overall mean, respectively.

RESULTS AND DISCUSSION

In total, 12 classification equations were developed to predict salability—two statistical classification methods for each of three gates in the sale planning process on each of two geographical zones within the Northern Region. Table 1 provides an overall summary of the results, showing that the percent correctly classified ranged from 59 to 90.

Table 1 suggests that classification success steadily improved with progression from gate 1 to gate 4, that eastside sales are more successfully classified, and that the two analytical procedures produce similar results. In fact, these impressions are correct, as will be shown in later statistical analysis.

The 22 significant variables used in the equations are defined in table 2. As indicated earlier, measurements on sale variables were made from timber sale records and economic variables from government publications.

Table 1—Overall summary of classification success

Gate	Subregion	Logistic regression	Discriminant analysis
		Percent correctly classified	
1	Eastside	65.0	72.5
	Westside	63.3	59.0
2-3	Eastside	80.0	77.2
	Westside	70.2	68.8
4	Eastside	85.0	90.0
	Westside	77.4	77.9

Table 2—Independent variables used in study

Variable	Description	Units
TOTVOL	Total sale volume harvested	M bd. ft. (Scribner)
TOTSALE	Total sale area	Acres
AVGSLOPE	Average slope	Percent
ALPM	Average logs per thousand	Number
TOTROAD	Total road construction	Miles
NEW	New road construction	Miles
RECON	Old road reconstruction	Miles
ACRES	Acres harvested in sale	Acres
DENSE	Acres harvested divided by total sale area	Number
VPA	Volume per acre harvested	M bd. ft. (Scribner)
DEAD	Percent volume dead white pine or lodgepole pine	Percent
%TRAC	Percent volume tractor yarded	Percent
%CABLE	Percent volume cable yarded	Percent
TRACDIST	Average maximum tractor yarding distance	Feet
STUMPMILL	Felling and bucking + skidding and loading + haul + slash + road + advertised rate	$/M bd. ft.
ADVRATE	Minimum bid price	$/M bd. ft.
SPLT	Selling price, lumber tally	$/M bd. ft.
PMETH	Contract price escalation clause	1 = Yes 0 = No
HAULRAT	Haul distance to primary appraisal point divided by haul distance to secondary appraisal point	Number
$UNCUT_{t-3}$	Uncut volume under contract lagged 3 months	Number
$EXCH_{t-3}$	U.S./Canadian exchange rate lagged 3 months	Number
COMPMILL	Competing mills at appraisal point	1 = Yes 0 = No
LMBRPROD	12-month percentage change in Inland region lumber production	Number

Gate 1

Gate 1 provides very little information that can be used to develop the equation. Only general site information (slope and elevation) and early volume estimates are known at this time. It is therefore difficult to predict group membership at this point.

EASTSIDE EQUATIONS

The gate 1 eastside equations and classification results are found in table 3. In general, the eastside equations indicate that sale size (TOTVOL and TOTSALE) has a positive influence on salability meaning that sales with larger volumes and of larger size increase the likelihood of selling. The standardized discriminant coefficients indicate that total volume (TOTVOL) is the most important determinant of salability.

In terms of significant variables, the logistic regression and the discriminant function are very similar. Their prediction results are noticeably different, however (table 3b). The logistic regression correctly classifies 42.9 percent of the unsold sales, while the discriminant function correctly classifies 85.7 percent. The logistic regression has the advantage of predicting sold sales, but the difference is not substantial. Given its unsold sales prediction accuracy, the discriminant function correctly classifies a higher percentage of all the sales, 72.5 percent compared to 65.0 percent. Also, the holdout method indicates that the discriminant error rates (percent correctly classified) are quite stable, with the largest percentage change occurring within the unsold sales class.

WESTSIDE EQUATIONS

Table 4 presents the westside equations and classification results. Both equations contain the same statistically significant variables, total volume (TOTVOL), total sale acres (TOTSALE), and average slope (AVGSLOPE). The westside equations indicate that sale size and slope are significant determinants of sold and unsold sales. On the westside forests, however, the total sale acres have a negative effect on salability. The average slope indicates offerings found on steep slopes are more likely to be unsold. The standardized discriminant coefficients indicate that total sale acres is the most important determinant of salability.

Table 4b presents the westside gate 1 classification results. A higher percentage of unsold sales are correctly classified by the discriminant function. But the logistic regression correctly classifies a higher percentage of sold sales. Overall the logisitic regression correctly classifies 63.3 percent of the sales, in comparison to 60.2 percent, for the discriminant function. The holdout classification results indicate the discriminant results are quite stable; the percentage correctly classified is identical for both classification measurements.

Table 3—Gate 1 eastside equations and classification results

A. Equations

	Logistic regression		Discriminant analysis	
Variable	Coefficient	(Std Err)	Coefficient	(Std Coeff)
TOTVOL	0.00048	(0.00022)		
$(\mathrm{TOTVOL})^{1/2}$			0.027	(0.575)
TOTSALE			.0003	(.487)
Constant	–.276	(.498)	–1.582	

B. Classification results

		Logistic regression		Discriminant analysis	
Actual group	Total sales	Correct predict	Percent correct	Correct predict	Percent correct
Unsold	14	6	42.9	12	85.7 (78.6)[1]
Sold	26	20	76.9	17	65.4 (61.5)[1]
All sales	40	26	65.0	29	72.5 (67.5)[1]

[1]Indicates percent correctly classified using the holdout method.

Table 4—Gate 1 westside equations and classification results

A. Equations

	Logistic regression		Discriminant analysis	
Variable	Coefficient	(Std Err)	Coefficient	(Std Coeff)
TOTVOL	0.00008	(0.00004)		
Ln(TOTVOL)			0.496	(0.742)
$(\mathrm{TOTSALE})^{1/2}$	–.032	(.008)		
Ln(TOTSALE)			–.734	(1.135)
AVGSLOPE	–.032	(.010)	–.053	(.659)
Constant	1.984	(.380)	2.562	

B. Classification results

		Logistic regression		Discriminant analysis	
Actual group	Total sales	Correct predict	Percent correct	Correct predict	Percent correct
Unsold	145	54	37.2	92	63.4 (63.4)[1]
Sold	204	167	81.9	114	55.9 (55.9)[1]
All sales	349	221	63.3	206	59.0 (59.0)[1]

[1]Indicates percent correctly classified using the holdout method.

Gates 2 and 3

At gates 2 and 3, the road network, size and number of cutting units, yarding system, silvicultural methods, and harvested acres are defined. These are examples of variables used to develop the gate 2-3 equations. At this point, the auction date is approximately 1 to 3 years in the future.

EASTSIDE EQUATIONS

The equations and classification results are presented in table 5. The gate 2-3 eastside equations have the same significant variables and differ only in the transformations used. The equations indicate the more volume harvested (TOTVOL), the more likely the timber sale will be sold. The standardized discriminant coefficient for total volume harvested indicates it is the most important determinant of sold and unsold sales. If the average logs per thousand board feet (ALPM) is high, it is more likely that the sale will be unsold. The ALPM variable is the number of logs to be moved; the more pieces moved the less likely the sale will be sold. The final variable—miles of total road construction (TOTROAD)—represents the initial development necessary to harvest the sale. The miles of road construction affect the sale by restricting the number of potential purchasers and introduce an additional source of risk by delaying the harvest (Johnson 1979).

The classification results for the two methods are almost identical (see table 5b). The difference lies in the number of sold sales correctly predicted. The logistic regression correctly classifies an additional sale. The discriminant function classification results are very stable. The classification results are identical for both measures of prediction accuracy.

The TOTVOL variable is the only gate 1 variable remaining. TOTSALE has been displaced by gate 2-3 variables (ALPM and TOTROAD). ALPM and TOTROAD are more specific timber sale information and allow prediction accuracy to increase.

Using gates 2 and 3 information, the equations have improved the overall classification results by 23.1 percent for the logistic regression and 6.9 percent for the discriminant equation.

WESTSIDE EQUATIONS

The westside gate 2-3 equations are quite similar in terms of significant variables. Total road construction (TOTROAD) was a more desirable variable in the logistic regression than was miles of new road construction (NEW) and miles of road reconstruction (RECON).

The equations and classification results are presented in table 6. The following variables have a positive effect on a sale selling: the ratio of acres harvested to total sale acres (DENSE), volume-per-acre harvested (VPA), percentage of volume tractor yarded (%TRAC), and the average maximum tractor yarding distance (TRACDIST). The remaining significant variables, miles of road construction (TOTROAD, NEW and RECON), percentage of volume of dead lodgepole or dead whitepine (DEAD), and the average logs per thousand board feet (ALPM) have a negative

Table 5—Gate 2-3 eastside equations and classification results

A. Equations

	Logistic regression		Discriminant analysis	
Variable	Coefficient	(Std Err)	Coefficient	(Std Coeff)
Ln(TOTVOL)	3.812	(1.448)		
$(\mathrm{TOTVOL})^{1/2}$			0.073	(1.537)
Ln(ALPM)			–.962	(.487)
ALPM	–.042	(.060)		
$(\mathrm{TOTROAD})^{1/2}$	–3.440	(1.342)	–1.105	(–1.294)
Constant	–18.426	(7.699)	2.142	

B. Classification results

		Logistic regression		Discriminant analysis	
Actual group	Total sales	Correct predict	Percent correct	Correct predict	Percent correct
Unsold	14	10	71.4	10	71.4 (71.4)[1]
Sold	26	22	84.6	21	80.8 (80.8)[1]
All sales	40	32	80.0	31	77.5 (77.5)[1]

[1]Indicates percent correctly classified using the holdout method.

Table 6—Gate 2-3 westside equations and classification results

A. Equations

	Logistic regression		Discriminant analysis	
Variable	Coefficient	(Std Err)	Coefficient	(Std Coeff)
DENSE	0.864	(0.420)		
$(\mathrm{DENSE})^{1/2}$			1.344	(0.324)
VPA	.062	(.016)		
$(\mathrm{VPA})^{1/2}$			.489	(.651)
Ln(TOTROAD)	–.200	(.112)		
$(\mathrm{NEW})^{1/2}$			–.230	(–.301)
$(\mathrm{RECON})^{1/2}$			–.064	(–.082)
DEAD	–.026	(.010)		
Ln(DEAD)			–3.944	(–.373)
$(\mathrm{ALPM})^2$	–.0004	(.00030)		
$(\mathrm{ALPM})^{1/2}$			–.261	(–.269)
%TRAC	.022	(.004)	.022	(.778)
$(\mathrm{TRACDIST})^2$	.00002	(.0000)		
Constant	–1.300	(.450)	–2.327	

B. Classification results

		Logistic regression		Discriminant analysis	
Actual group	Total sales	Correct predict	Percent correct	Correct predict	Percent correct
Unsold	145	84	57.9	101	69.7 (66.9)[1]
Sold	204	161	78.9	139	68.1 (66.2)[1]
All sales	349	245	70.2	240	68.8 (66.5)[1]

[1]Indicates percent correctly classified using the holdout method.

effect on a timber sale selling. Of the variables in the discriminant function, %TRAC is the most important discriminator (standardized coefficient = 0.778).

Comparing westside and eastside results, one can conclude that the westside timber offerings are more difficult to explain. Given the complexity, a more complicated equation is needed to produce the observed classification results. Therefore, one should not expect an identical equation structure between geographical zones.

The overall classification results for the two approaches are also quite similar (see table 6b). The logistic regression correctly classifies 70.2 percent of the sales, while the discriminant function correctly classifies 68.8 percent. The differences exist in the equation's ability to classify the individual categories. The discriminant function correctly classifies a higher percentage of the unsold sales (69.7 vs. 57.9 percent), the logistic regression correctly classifies more of the sold sales (78.9 vs. 68.1 percent). The holdout method indicates that prediction accuracy for the discriminant model is 66.5 percent, a slight decrease. Also, there is a slight decrease in the individual group classification results.

The gate 2-3 equations have changed in comparison to gate 1 to reflect the better information available. After adding the information from gates 2 and 3, the overall classification results increased from 221 to 245 (10.9 percent) for the logistic regression and from 206 to 240 (16.5 percent) for the discriminant equation.

Table 7—Gate 4 eastside equations and classification results

A. Equations

Variable	Logistic regression Coefficient	(Std Err)	Discriminant analysis Coefficient	(Std Coeff)
STUMPMILL	–0.078	(0.030)		
Ln(STUMPMILL)			–6.102	(–1.132)
TOTVOL	.0012	(.0004)		
$(TOTVOL)^{1/2}$			.037	(.783)
SPLT	.088	(.038)		
Ln(SPLT)			6.656	(.832)
Ln(ALPM)			–.615	(–.169)
Ln(DEAD)			.287	(.351)
PMETH			–.667	(–.274)
Constant	–2.620	(5.488)	–2.622	

B. Classification results

Actual group	Total sales	Logistic regression Correct predict	Logistic regression Percent correct	Discriminant analysis Correct predict	Discriminant analysis Percent correct
Unsold	14	10	71.4	13	92.9 (71.4)[1]
Sold	26	24	92.3	23	88.5 (73.1)[1]
All sales	40	34	85.0	36	90.0 (72.5)[1]

[1]Indicates percent correctly classified using the holdout method.

Gate 4

At the final gate before the offering, appraisal and economic information is added to site and sale characteristics. This point in the gates process can be viewed as one that converts sale characteristics information into dollars per thousand board feet. This gate is approximately 2 to 3 months before the initial offering.

EASTSIDE EQUATIONS

In general, the gate 4 eastside equations indicate that the stump-to-mill cost (STUMPMILL), the size (TOTVOL), and the price of final product derived from the logs (SPLT) are the important factors affecting salability (table 7). The discriminant function contains three additional variables, the natural logarithm of average logs per thousand, ln(ALPM), the natural logarithm of percent volume dead white pine or dead lodgepole, ln(DEAD), and the price escalation clause used in the timber sale contract (PMETH). These characteristics indicate the number and quality of the logs, and the contractual agreement of the sale. The standardized discriminant coefficients conclude that the STUMPMILL variable is the most important factor in determining sold and unsold sales, with SPLT second.

Table 7b presents the classification results for the eastside equations. The discriminant function correctly classified 90.0 percent of the timber sales, 92.9 percent of the unsold sales, and 88.5 percent of the sold sales. The holdout method indicates that the above classification results are unstable. The overall correct classification decreases to 72.5 percent, with 71.4 percent for the unsold sales and 73.1 percent for the sold sales. The logistic regression correctly classified 85.0 percent of the timber sales, with 71.4 percent correct classification for the unsold sales and 92.3 percent for the sold sales.

With the introduction of the gate 4 information, the equations have improved the overall classification results by 6.3 percent for the logistic regression and 16.1 percent for the discriminant equation.

WESTSIDE EQUATIONS

At gate 4, all information from the Gates timber sale planning process plus market information found in other sources is used to develop the equations. Given this fact, these equations are the most complicated. This has led to an equation that produces the most accurate prediction of salability.

The gate 4 equations are displayed in table 8. One should examine these equations from the standpoint of which variables have a positive effect and which have a negative effect on salability. The following variables have a positive effect on salability: the selling price (SPLT), the ratio of haul distances to primary and secondary appraisal points (HAULRAT), whether the contract follows the WWPA price index or remains fixed (PMETH), whether the mill site is competitive (COMPMILL), the 12-month percentage change in lumber production in the Intermountain zone (LMBRPROD), the volume-per-acre harvested (VPA), and the ratio of harvested acres to sale acres (DENSE). In general, the above characteristics indicate that a sale is likely to be sold: when composed

Table 8—Gate 4 westside equations and classification results

A. Equations

Variable	Logistic regression Coefficient	(Std Err)	Discriminant analysis Coefficient	(Std Coeff)
Ln(STUMPMILL)	–5.308	(0.716)	–3.307	(0.830)
$(SPLT)^2$	.00002	(.0000)		
Ln(SPLT)			.824	(.194)
Ln(HAULRAT)	.648	(.248)		
HAULRAT			.744	(.173)
DEAD	–.036	(.012)	–.023	(.270)
Ln(DENSE)			.183	(.148)
$(VPA)^{1/2}$			.113	(.150)
Ln(ACRES)			–.114	(–.137)
$(\%CABLE)^2$	–.00018	(.00004)	–.0001	(.487)
Ln($UNCUT_{t-3}$)	–9.340	(3.632)		
($UNCUT_{t-3}$)			–.0006	(.149)
COMPMILL	.864	(.288)	.636	(.318)
PMETH	.924	(.406)	.593	(.218)
LMBRPROD	.026	(.010)		
$EXCH_{t-3}$	–5.964	(3.124)	–6.856	(.419)
Constant	110.470	(28.442)	23.213	

B. Classification results

Actual group	Total sales	Logistic regression Correct predict	Percent correct	Discriminant analysis Correct predict	Percent correct
Unsold	145	100	69.0	115	79.3 (78.6)[1]
Sold	204	170	83.3	157	77.0 (75.0)[1]
All sales	349	270	77.4	272	77.9 (76.5)[1]

[1]Indicates percent correctly classified using the holdout method.

of higher valued species, located near several competitive milling centers, sale contract allows the winning bid price to fluctuate with the lumber market, lumber markets in an upswing, high volumes per acre harvested, does not require undue movement of harvesting equipment and labor.

As they rise in value, or are present in the timber sale the following variables negatively influence salability: stump-to-mill costs (STUMPMILL), percentage volume dead white pine or dead lodgepole (DEAD), percentage volume cable yarded (%CABLE), U.S./Canadian exchange rates ($EXCH_{t-3}$), and the uncut volume under contract ($UNCUT_{t-3}$) for the westside National Forests of Region 1. Once again these variables indicate sale quality in terms of costs and the type of volume harvested. The most important characteristic in determining salability is STUMPMILL, with %CABLE second.

Overall correct classification is approximately 78 percent (table 8b). The logistic regression correctly classified 69 percent of the unsold sales and 83.3 percent of the sold sales. The discriminant function correctly classified 79.3 percent of the unsold sales and 77.0 percent of the sold sales. The holdout method indicates that the discriminant function's classification results are stable.

Adding gate 4 and market information allowed the overall correct classification to increase approximately 12 percent. In addition to the overall improvement in correct classification, the gate 4 equations improved individual class results. Therefore we not only gain overall classification accuracy, but also accuracy within individual groups.

Statistical Evaluation

Up to this point I have discussed the classification results achieved when modeling sold and unsold timber sales at different gates, by different classification techniques, and by geographical zones. But the question still remains as to the statistical significance of the classification results achieved by the gates, classification technique, and geographical zone. The practical significance of the results are left to the reader to determine.

In general, the categorical analysis of variance (see table 9) indicates that a statistically significant difference exists in classification results when considering the geographical zone (eastside vs. westside), and when moving from gate 1 to gate 4. But there was no statistically significant difference in classification success with respect to the classification method (logistic regression vs. discriminant analysis). Also, there were no statistically significant interaction terms in the analysis of variance equation.

A statistically significant improvement in classification results was observed when moving from gate 1 through gate 4. As better information is used in the equation development process, the classification results improve significantly. The sale can be described in terms of average elevation, average slope, total volume, and sale acres at gate 1. These measures indicate the type of yarding systems that will be needed and the general size of the sale, but the general nature of the information will not allow an accurate model to be developed. The process needs specific information with regard to the percentage of volume cable yarded, the miles of road construction, the number of pieces moved, and so forth. Also, the equation is improved as market information is added to the process.

The results show that a statistically significant higher percentage of sales were correctly classified in the eastside National Forests. In general, the eastside equations had fewer variables than the westside equations.

Table 9—Categorical analysis of variance results[1]

Source	df	Chi–square	Probability
Intercept	1	2803.82	0.0000
Region[2]	1	10.16	[3].0014
Gate[4]	2	27.51	[3].0000
Method[5]	1	.08	.7724

[1]Interaction terms were not significant.
[2]Variable that represents the eastside and westside geographical zone.
[3]Indicates a statistically significant factor.
[4]Variable that represents gates 1, 2-3, and 4.
[5]Variable that represents the statistical methods.

But the simpler eastside equations correctly classified a higher percentage of the timber sales.

It is reassuring to know that as we add more specific information to the process, moving from gate 1 to gate 4, the classification results improve.

Also, having nonsignificant results with respect to the statistical procedure chosen is reassuring.

Classification Results and Cutoff Points

The classification results presented were based on a decision rule that assigns timber sales to the unsold category if the probability is less than 0.50, and to the sold category if greater than or equal to 0.50. As discussed earlier, one would expect classification results to change as the cutoff points (decision rules) are altered. Figure 1 presents the classification results based on the seven cutoff points defined earlier. When examining the eastside discriminant function (fig. 1), the most striking feature of gate 1 is the large fluctuation in the percentage correctly classified as the probability is varied. If the decision rule establishes that a predicted probability of 0.20 or greater defines a sold sale, the percentage of the sold sales that are correctly classified by the eastside gate 1 discriminant function is 100 percent. The tradeoff, however, is that under that rule, 0 percent of the unsold sales were correctly classified. At the other end of the scale, where a probability of 0.80 or greater defines a sold sale, the percentage of the sold sales correctly classified is 7.7; and the percentage of the unsold sales that are correctly classified is 100. The eastside gate 4 discriminant function at a predicted probability of 0.20 correctly classifies 64.3 percent of the unsold sales and 96.2 percent of the sold sales. At a probability of 0.80 sold, the classification results are 100 percent for unsold sales and 65.4 percent for sold sales. Figures corresponding to the other equations and gates are presented in appendix A, figures 2-4.

Figure 1 also illustrates how the cutoff points could be varied from gate to gate. As the timber sale nears the auction date, the land manager can become more precise regarding the cutoff point without drastically affecting the percentage correctly classified. At gate 1, which is 5 to 10 years from the auction date, the land manager may want to maintain the standard cutoff, 0.50.

The gate 1 equations produce the largest variation in percentage correctly classified at the various probability levels. The explanation is that gate 1 equations are the least accurate in percentage correctly classified. The gate 1 predicted probabilities are clustered around 0.50, meaning the equations have difficulty making correct predictions. As more and better information is generated (moving from gate 1 to gate 4), the equations are better able to classify sales, and the predicted values do not cluster around 0.50. For the eastside gate 1 discriminant function using a probability of 0.20, the percentage correct varies from 0 to 100; at a probability of 0.80 the percentage correct ranges from 100 to 7.7 for unsold and sold sales, respectively. The eastside gate 4 discriminant function ranges from 64.3 percent to 96.2 percent for

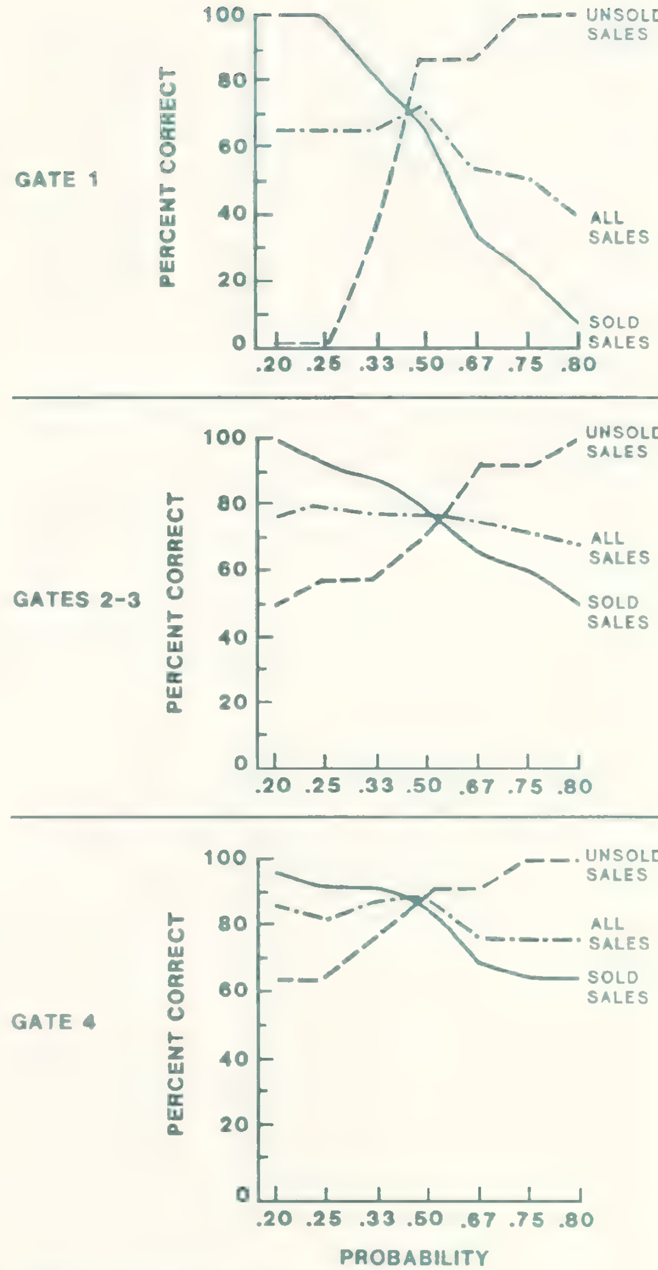


Figure 1—Cutoff point analyses for eastside Discriminant model.

a probability of 0.20, and 100 percent to 65.4 percent for a probability of 0.80. The maximum range for gate 4 is 34.6 percent, in comparison to 100 percent for gate 1.

MANAGEMENT IMPLICATIONS

The Gates timber sale planning system plays an important role in timber management. The Gates process can be viewed as adding structure to the designing of timber sales, where a sale must meet certain requirements before moving to the next gate. Additional information can be generated through the Gates process to help timber sale planners to make sound economic decisions about the direction and extent of development. This is especially important today when government spending is a major issue. Using classification procedures, like those developed in this paper, coupled with the Gates process, the timber sale planner can generate a critical piece of information—likely salability. The planner must be aware of how his planning decisions affect salability.

The sale planning process starts with a timber management specialist inspecting a proposed sale area. Based on initial site and volume estimates, the forester determines if the area can support a timber sale. From a modeling standpoint, the volume estimates and site information can be used to assess salability.

At gates 2 and 3, the sale takes on more definite form, and more specific information is known about the prospective sale. It is now possible to more reliably evaluate salability. Using the more specific information, prediction capability of classification procedures has been improved. The ability to quantify sale design decisions, such as adding another mile of road or decreasing the volume per acre harvested, is essential to predicting salability. The prediction of salability is most important at this point. If salability can be accurately predicted at this point, the timber sale planner can implement the necessary changes to produce a salable offering.

Gate 4 offers little time to make necessary changes to produce a viable timber sale, with the auction only 3 months away. If selling the offering is questionable, the sale could be delayed and the necessary changes implemented to produce a salable timber offering.

The logistic regression and discriminant function were evaluated as salability classification tools. The categorical analysis of variance results verified that there was no statistically significant difference in prediction results with respect to the classification procedure.

But the logistic regression approach may have an added advantage over discriminant analysis. The logistic regression produces an estimated probability of a sale selling. This prediction supplies more information to the user about the degree of salability. For example, if the logistic model predicts a 10 percent chance of selling, this should indicate to the planner that major renovations are needed. The above sale is accurately classified as an unsold sale, but it is very different from a sale that might have an estimated probability of 0.45 (45 percent chance of selling). Given our 50-50 decision rule, they would both be classified as unsalable sales, even though the sale with an estimated probability of 0.10 should probably not be offered, while the other could be offered under favorable market conditions. The predicted probability provides the planner with a flexible decision rule. This flexibility could lead to a rule that allows for a zone of indecision. The zone could be defined as any sale having a probability less than 0.30 will be deferred, greater than 0.30 will be revised, and any sale having a probability greater than 0.70 will be advertised for sale. Any sale having an estimated probability falling in the 0.30 to 0.70 range of indecision will be withheld, revised, or advertised based on the professional judgment of the planning staff.

The equations described here were based on a sample of sold and unsold timber sales in the Northern Region of the Forest Service, during 1980 to 1985. The results should not be used to predict sold and unsold sales in any other region of the Forest Service nor for any other seller of stumpage.

REFERENCES

Artley, D. 1986. DLOGPRICE economic model. Unpublished draft supplied to author by D. Artley.

Bishop, Y. M. M.; Fienberg, S. E.; Holland, P. W. 1975. Discrete multivariate analysis: theory and practice. Cambridge, MA: MIT Press. 348 p.

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

Johnson, R. A.; Wichern, D. W. 1982. Applied multivariate statistical analysis. Englewood Cliffs, NJ: Prentice-Hall. 594 p.

Johnson, R. N. 1979. Oral auction versus sealed bids: an empirical investigation. Natural Research Journal. 19(1): 315-335.

Lachenbruch, P. A. 1975. Discriminant analysis. New York: Hafner. 128 p.

Maddala, G. S. 1983. Limited-dependent and qualitative variables in econometrics. Cambridge, MA: Cambridge University Press. 401 p.

Merzenich, J. P. 1985. Transaction evidence timber appraisals in the Northern Rockies. Missoula, MT: U.S. Department of Agriculture, Forest Service, Northern Region. 19 p. [Mimeo].

Morrison, D. F. 1976. Multivariate statistical methods. New York: McGraw-Hill. 415 p.

Nie, N. H. 1983. SPSSX statistical software. New York: McGraw-Hill. 806 p.

Peterson, T. 1980. Timber sale feasibility analysis. Unpublished paper on file at: U.S. Department of Agriculture, Forest Service, Bitterroot National Forest, Hamilton, MT. 15 p.

Pindyck, R. S.; Rubinfeld, D. L. 1981. Econometric models and economic forecasts. New York: McGraw-Hill. 630 p.

Press, S. J.; Wilson, S. 1978. Choosing between logistic regression and discriminant analysis. Journal of American Statistical Association. 73(364): 669-704.

U.S. Department of Agriculture, Forest Service. 1985. Forest Service Manual 2400 Timber Management. Washington, DC: U.S. Government Printing Office: 2414.27, 2431.2.

APPENDIX A: CUTOFF POINTS

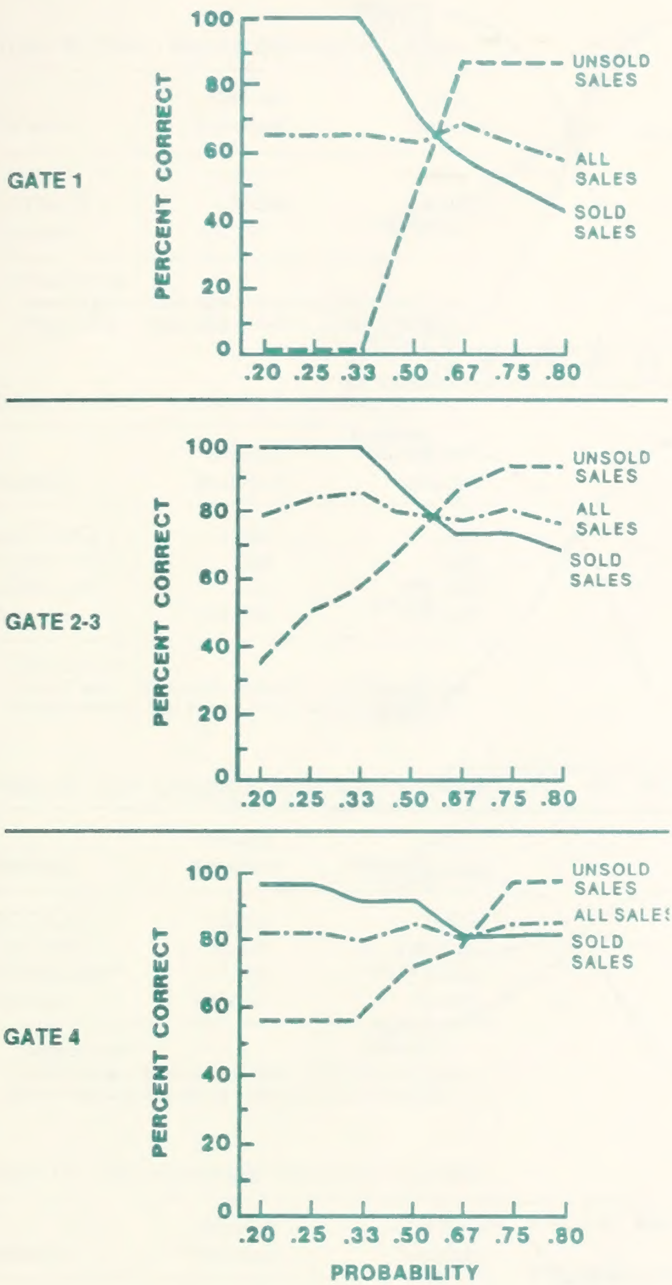


Figure 2—Eastside logistic regression model.

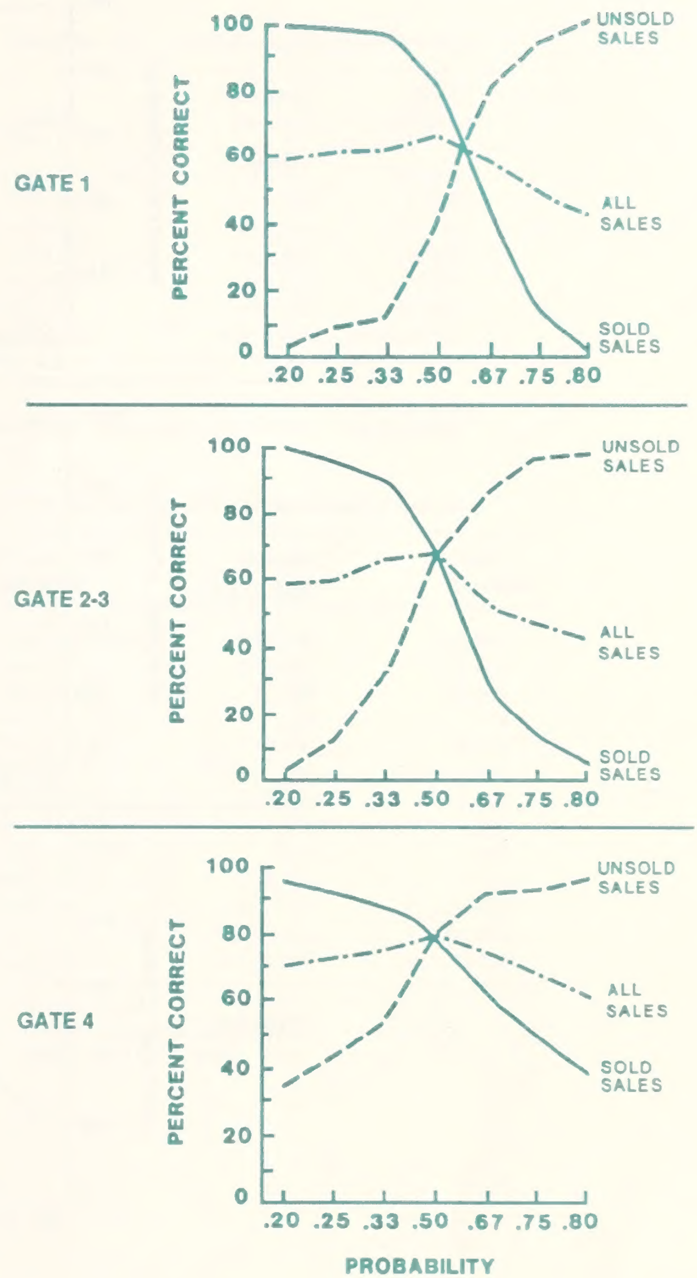


Figure 3—Westside discriminant model.

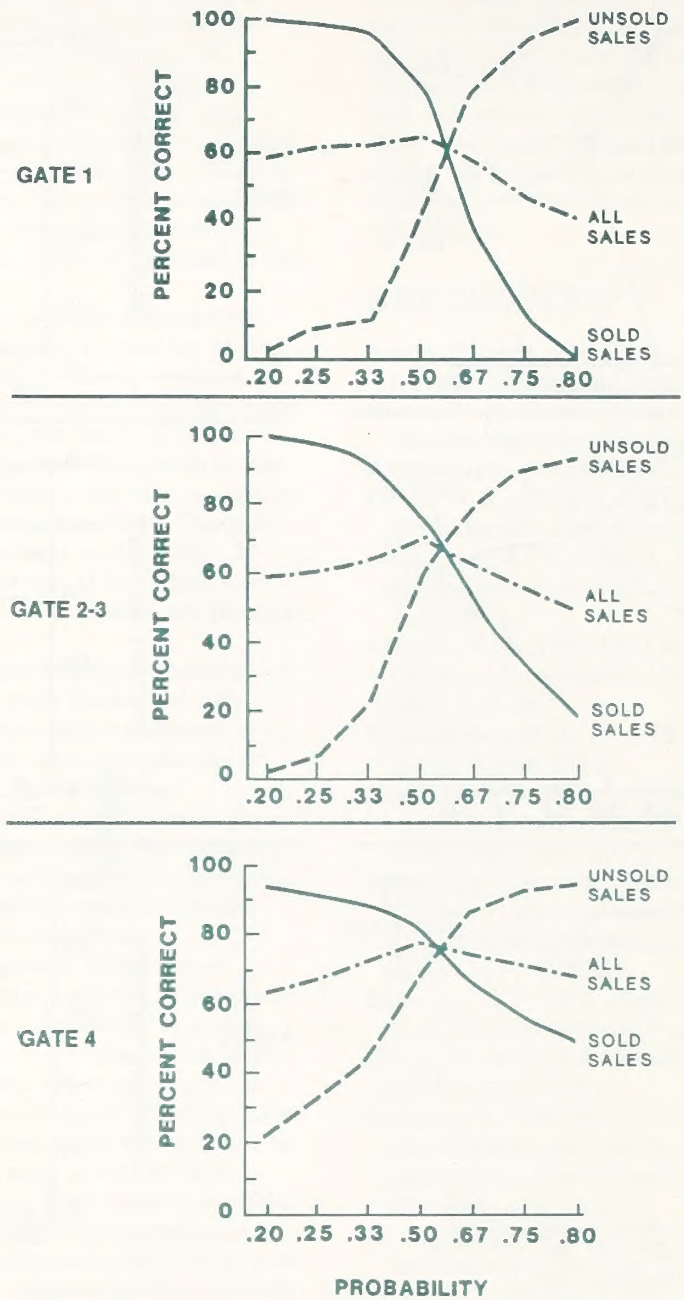


Figure 4—Westside logistic regression model.

APPENDIX B: DISCRIMINANT ANALYSIS CLASSIFICATION EQUATIONS

Table 10—Gate 1 eastside classification functions

Variable	Unsold function[1]	Sold function[1]
$(TOTVOL)^{1/2}$	0.125	0.147
TOTSALE	–.00096	–.00072
Constant	–2.207	–3.379

[1]Decision rule:
Unsold value > Sold value —then—> Unsold prediction
Unsold value < Sold value —then—> Sold prediction.

Table 11—Gate 1 westside classification functions

Variable	Unsold function[1]	Sold function[1]
Ln(TOTVOL)	2.554	2.841
Ln(TOTSALE)	.553	.128
AVGSLOPE	.140	.109
Constant	–14.486	–12.972

[1]Decision rule:
Unsold value > Sold value —then—> Unsold prediction
Unsold value < Sold value —then—> Sold prediction.

Table 12—Gate 2-3 eastside classification functions

Variable	Unsold function[1]	Sold function[1]
$(TOTVOL)^{1/2}$	–0.056	0.062
Ln(ALPM)	48.414	46.825
$(TOTROAD)^{1/2}$	–1.139	–2.965
Constant	–81.956	–78.007

[1]Decision rule:
Unsold value > Sold value —then—> Unsold prediction
Unsold value < Sold value —then—> Sold prediction.

Table 13—Gate 2-3 westside classification functions

Variable	Unsold function[1]	Sold function[1]
$(DENSE)^{1/2}$	19.399	20.392
$(VPA)^{1/2}$	3.019	3.398
$(NEW)^{1/2}$	–1.139	–2.965
$(RECON)^{1/2}$	.560	.523
Ln(DEAD)	.036	–.077
$(ALPM)^{1/2}$	4.611	4.398
(%TRAC)	.018	.035
Constant	–22.847	–24.577

[1]Decision rule:
Unsold value > Sold value —then—> Unsold prediction
Unsold value < Sold value —then—> Sold prediction.

Table 14—Gate 4 eastside classification functions

Variable	Unsold function[1]	Sold function[1]
Ln(STUMPMILL)	46.431	33.289
$(TOTVOL)^{1/2}$	–.136	–.056
Ln(SPLT)	317.393	331.734
Ln(ALPM)	–81.956	–78.007
Ln(DEAD)	2.579	3.197
PMETH	–30.205	–31.641
Constant	–1,011.769	–1,016.721

[1]Decision rule:
Unsold value > Sold value —then—> Unsold prediction
Unsold value < Sold value —then—> Sold prediction.

Table 15—Gate 4 westside classification functions

Variable	Unsold function[1]	Sold function[1]
Ln(STUMPMILL)	105.432	100.952
Ln(SPLT)	155.338	156.482
HAULRAT	17.254	18.292
DEAD	.366	.334
Ln(DENSE)	–4.533	–4.288
$(VPA)^{1/2}$	8.346	8.498
Ln(ACRES)	5.491	5.335
$(\%CABLE)^2$	–.00091	–.0011
$UNCUT_{t-3}$	.049	.048
COMPMILL	–8.184	–7.315
PMETH	–27.697	–26.897
$EXCH_{t-3}$	568.198	559.344
Constant	–1,170.096	–1,139.000

[1]Decision rule:
Unsold value > Sold value —then—> Unsold prediction
Unsold value < Sold value —then—> Sold prediction.

Niccolucci, Michael J. 1989. Predicting salability of timber sale offerings in the Forest Service Northern Region. Res. Pap. INT-418. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 13 p.

Some 389 sold and unsold timber offerings in the Northern Region of the Forest Service were used to develop classification equations. Equations were designed to be used at three points in the Gates timber sale planning process. Classification results ranged from 59 to 90 percent correct classification.

KEYWORDS: unsold timber sales, classification, Gates timber sale planning process, logistic regression, discriminant analysis

INTERMOUNTAIN RESEARCH STATION

The Intermountain Research Station provides scientific knowledge and technology to improve management, protection, and use of the forests and rangelands of the Intermountain West. Research is designed to meet the needs of National Forest managers, Federal and State agencies, industry, academic institutions, public and private organizations, and individuals. Results of research are made available through publications, symposia, workshops, training sessions, and personal contacts.

The Intermountain Research Station territory includes Montana, Idaho, Utah, Nevada, and western Wyoming. Eighty-five percent of the lands in the Station area, about 231 million acres, are classified as forest or rangeland. They include grasslands, deserts, shrublands, alpine areas, and forests. They provide fiber for forest industries, minerals and fossil fuels for energy and industrial development, water for domestic and industrial consumption, forage for livestock and wildlife, and recreation opportunities for millions of visitors.

Several Station units conduct research in additional western States, or have missions that are national or international in scope.