Agricultural Economics Miscellaneous Report No. 119 October 1988

Loan Default Prediction Using Logistic Regression and a Loan Pricing Model

Timothy Mortensen
David L. Watt
F. Larry Leistritz

Department of Agricultural Economics • North Dakota State University
Fargo, ND 58105

Acknowledgements

This publication is based on an M.S. Thesis completed by Timothy Mortensen in August 1987. The authors wish to thank Roger Johnson, Cole Gustafson, David Saxowsky, and Brenda Ekstrom for their valuable editorial comments and to Mary Altepeter and Marna Unterseher who did an excellent job of typing the document.

The authors accept sole responsibility for errors in the text.

Table of Contents

List of Tables

List of Figures

Highlights

A logistic regression model is developed that demonstrates the estimation of crucial financial characteristics of farmers that determine loan delinquency. The resulting model is then used in conjunction with a loan pricing model to determine the optimum loan default breaking point where lender revenue is maximized. Results indicate that debt-to-asset ratio and the operating ratio are most indicative of loan default for the population surveyed.

LOAN DEFAULT PREDICTION USING LOGISTIC REGRESSION AND A LOAN PRICING MODEL

Timothy Mortensen, David L. Watt, and F. Larry Leistritz[1]

Substantial increases in the real price of farm commodities during 1972 and 1973 benefited producers. Some farmers made investments thinking that continued high prices would make it possible to pay principal and interest with current earnings and capital gains. During the 1970s, farm prices advanced with inflation, but the boom that began nearly a decade earlier ended in 1980 when commodity prices failed to advance along with general price inflation (Melichar, 1984). Deterioration of the financial condition of North Dakota farm operators has been the topic of news reports and research efforts since then (Leholm et al. 1985; Watt et al. 1986; Leistritz et al. 1987).

Other research has focused on the effects that the farm crisis has on lenders in the state. Saxowsky et al. (1987) report that 13.5 percent ($466.1 million) of North Dakota's farm loan volume was delinquent as of July 1, 1986. This has impacted the financial status of most North Dakota lenders. These lenders were faced with growth in agricultural credit demand, increased use of national credit markets (which exposed them to volatile interest rates in the 1980s), then deteriorating credit worthiness of some farm borrowers.

Several studies have focused on the problem of classifying farm borrowers as good or poor risks (see Luftburrow, 1984; Pederson, 1983; and Fiske, 1986). Luftburrow and Pederson used probit and discriminant analysis, respectively, in their research. Fiske employed logistic regression to model factors influencing currentness of debt payments of Ohio farmers. The purpose of this paper is to demonstrate a model estimating crucial predictors of loan defaults. Sample data are used to identify significant determinants of the debt payment status of North Dakota farmers. Selected determinants are then used to model the probability of loan default. In conjunction with a loan-pricing model specific to the lender, the data are used to demonstrate the calculation of a breaking point (PPROB) that maximizes revenue to the lending institutions.

DATA AND STUDY PROCEDURES

Data from longitudinal, cross-sectional surveys of 763 farm operators conducted by the North Dakota Agricultural Experiment Station in 1985 and 1986 were used in this analysis. Respondents to the surveys were initially screened to include only those who (1) were less than 65 years old, (2) considered farming to be their primary occupation, and (3) sold at least $2500 of farm products in the past year. Additional information and comparisons of the survey respondents with 1982 Census of Agriculture data

---

[1]The authors are Research Assistant, Associate Professor, and Professor, respectively, Department of Agricultural Economics, North Dakota State University.

are found in Leistritz et al. (1986). Comparisons indicate that the survey is representative of the state's farmers who were less than 65 years old and considered farming their primary occupation.

Measures taken from farmers' responses on the surveys were used as determinants of solvency and performance of the farm business and were thought to be directly or indirectly related to farm financial stress.

A farm operator is current on annual debt obligations if all principal and interest have been paid (or debt has been renegotiated). Survey respondents were asked whether they were current on debt payments on January 1, 1986. The dependent variable in the regression model, CURRENT, equals 0 if a farm operator was current on debt payments, and equals 1 if the operator was delinquent.

Debt payment status on January 1, 1986, is hypothesized to be the result of the solvency position of the business on January 1, 1985, performance during 1985, and other socioeconomic measures of the operator. Most explanatory variables are ratios representing profitability and liquidity during 1985 and leverage on January 1, 1985. Absolute dollar measures and other socioeconomic variables thought to be useful as determinants of debt payment status also are tested in the analysis. Table 1 lists the variables used in the analysis. These variables were drawn from the many listed in popular and research literature as being important indicators of factors in farm financial stress.

A decision to lend money to a farm operator is usually based on solvency of the business at the time of loan application and on other factors such as farm profitability in past years and expected performance in the upcoming crop year. Returns deviate from their expected value due to variations in yields and product prices derived from production activities. This means that operators, in years of low returns, may not generate sufficient cash flow to meet debt service obligations. A model able to predict the probability of loan default would be useful to lenders in making loan decisions.

The intent of the model is to verify significant 1985 variables that, when applied in combination with the operator's solvency status on January 1, 1985, will be good indicators of expected debt payment status at the end of the operating year. Ideally, to test for good indicators at the time a loan decision is made, previous years data from a lending institution are necessary.

TABLE 1 . VARIABLE DEFINITIONS.

CURRENT	= 0 if current on debts; 1 otherwise (on January 1, 1986).
AGE86	= age of the operator on January 1, 1986.
LONGFARM	= number of years the farmer has operated the farm.
RATACRES	= the ratio of acres rented to total acres operated.
GROSIN85A	= gross cash farm income during 1985.
NETIN85A	= net cash farm income for 1985.
GPCR	= gross profit on cash revenue or net cash farm income before depreciation divided by gross cash farm income.
OPERAT5	= operating ratio which consists of production expenses divided by gross cash farm income in 1985.
NETCAP5	= net cash farm income divided by persons in the household.
VIAB85	= the viability of farm operators during 1985.[2]
NETAST5	= net cash farm income in 1985 divided by total assets.
INTGRS	= interest paid in 1985 divided by gross cash farm income.
CURAT5	= a ratio of current debts divided by current assets.
OFFDBT	= a ratio of non-farm income to total debt in 1985.
DBTAST85	= debt to asset ratio of the operator on January 1, 1985.
DBTACR5	= total debt divided by acres farmed on January 1, 1985.

METHOD

Regression Model

Application of linear regression models when the dependent variable is binary (0 or 1) in nature is complex and involves use of binary choice models. While direct continuous relationships between various independent variables or attributes and behavior of a farm operator are possible, survey data lacks consistency in measurement for the purpose of testing attributes associated with farm loan delinquency. A desirable

[2]See Appendix A.

characteristic of a model is its ability to predict the likelihood of a farm operator being current or delinquent given attributes of that particular operator.

A logit model is useful in estimating the occurrence of events that are bivariate in nature (they can occur or not occur). Logit is based on a cumulative probability function. The function estimates the probability of, in this case, the farmer being delinquent on debt repayment at the end of the year. Data are analyzed using the LOGIST procedure (Harrel, 1983) on Statistical Analysis System (SAS) software.

Optimization of the Classification Table

Figure 1 shows a classification of the number of actual and predicted current and delinquent observations. As can be seen in the example shown, the actual percentage of correct classifications is very high (e.g., 80 percent). The actual delinquent observations that are predicted current are, however, troublesome. In this example no actual delinquent observations were predicted as delinquent. Because delinquent loans are usually very expensive to banks, the percent of correctly classified predictions may not be the best criterion for achieving profit maximization. This is demonstrated later in this section.

Fiske et al. (1985) reported that their two model specifications correctly classified 78 and 76 percent of the observations, respectively. They did not indicate how many actual delinquent observations were predicted delinquent. If a lender were to use the model to classify potential borrowers, every borrower would be granted a loan; and an 80 percent accuracy rate could cause financial stress for the bank.

The classification for predictions can be adjusted by setting the probability of default on loan payments at different levels. This level is called the breaking point. Observations having a probability greater than the breaking point are classified predicted delinquent and observations having a linear probability are classified predicted current.

Adjustment of the breaking point does not affect estimation of the model, or statistics associated with the procedure. The breaking point affects only the predicted classification of the observations.

From a lender's perspective, an institution would be exposed to the greatest potential loss by farm operators who were predicted to remain current on debt obligations that actually default. The objective, then, is to maximize economic return to a lending institution.

		PREDICTED	
		CURRENT	DELINQUENT
ACTUAL	CURRENT	200 QUADRANT I	0 QUADRANT II
	DELINQUENT	50 QUADRANT III	0 QUADRANT IV

PERCENT CORRECT = 80.0

Figure 1. Classification Table.

<u>Loan Pricing Model</u>

Saxowsky et al. (1987) report a loan pricing model where interest income from loans is equated with costs. The equation is shown as:

(1) $LY = I + L + O + E - S - F - DY + r$

Where: LY = interest income from loans not in default.

I = interest costs

L = provisions for loan losses

O = operating expenses

E = earnings for building reserves

S = income from securities

F = fees and other income

DY = interest income from loans in default

r = risk premium.

Lending institutions derive interest income from loans not in default (LY), interest from loans in default (DY), income from securities and government instruments (S), and service fees (F). Costs are incurred by interest paid to depositors (I), provisions for loan principal losses (L), and operating expenses (O). These may be expressed as cost of funds (COF) shown as:

(2) COF = I (average interest paid) + O + L.

Moving income items to and subtracting COF from the left side results in the following equation:

(3) LY + S + F + DY - COF = E + r.

Equation (3) is used for illustrative purposes in optimizing the classification table. Simplifying assumptions concerning a lending institution include:

1. An institution has two alternatives for investment of deposits: (a) lend money to individuals or businesses for various purposes, or (b) invest in government or other securities. No preference is shown as to the proportion of funds used for either choice. Securities are considered risk-free investments (for simplicity); therefore, (r) is zero when this choice is made.

2. Principal, but no interest, is recovered from loans granted that eventually default (DY is zero).

3. Service fees (F) are assumed constant and so are not included in calculations.

Funds invested in securities (S) less cost of funds (COF) equals earnings for building reserves (E). This may be considered "normal" profit. Funds loaned to individuals or businesses (LY) less cost of funds (COF) equals (E) plus a premium for added default and business risk (r).

(4) S - COF = E and

(5) LY - COF = E + r.

Variables may be expressed in percentage terms by dividing through each its respective total.

Saxowsky et al. (1987) reported an estimated average interest rate of approximately 10.5 percent on outstanding loans. An interest rate of 11 percent and a cost of funds (COF) of 9.5 percent are used for illustration in calculating an optimum loan decision breaking point. Return from loaned funds (E plus r) is 1.5 percent, and considering (r) as 1.0 percent, return on security investment is .5 percent. The values are used in the classification table to illustrate setting a breaking point for optimizing returns. Values used in the illustration should be adjusted to fit individual lender situations.

Observations properly classified in quadrant one (Figure 1) are borrowers that were correctly predicted as current and actually did not default. Revenue derived from loans in this quadrant includes E and r or 1.5 percent over cost of funds (COF).

Observations in quadrant two are actual current borrowers which the model predicted as delinquent. The model incorrectly classified potential borrowers as delinquent in this quadrant. Funds not loaned to individuals or businesses are invested in securities and earn a positive return of .5 percent.

Quadrant four contains delinquent borrowers predicted to be delinquent. They are correctly classified and a lender also would invest funds in securities, earning .5 percent above the cost of funds.

Quadrant three contains the borrowers who were predicted current that actually defaulted. If collateral is sufficient to cover principal and recover cost, a lender could lose interest income from borrowers who defaulted. In this example, losing interest (not principal) on loans is the cost of loan default. Delinquent loan losses vary among institutions, and the average loss is unknown. An institution using this model could use either their historic values or an estimate of future impacts of loan default. In this example, if the interest rate charged on a loan is 11 percent and overall cost of funds (COF) is 9.5 percent, the lender is penalized and will lose 9.5 percent interest income on misclassified borrowers in quadrant three.

The objective, then, is to find the breaking point (PPROB) that maximizes lender earnings. This is accomplished by comparing earnings as the breaking point is varied. An institution is considered to have $10 million (a fixed amount is used for simplicity) that may either be used for loans or invested in securities. Potential borrowers in quadrants one and three receive loans based on their respective percent of total customers. The remaining potential loan applicants are rejected and funds are invested in securities, assuming that no other customers are available for loan processing.

As the breaking point is reduced, potential borrowers classified in quadrants one and three are reclassified into quadrants two and four, respectively. The optimum breaking point is reached when the sum of revenue from the four quadrants is maximized. Results derived from this analysis are presented below.

RESULTS

Selection of Variables

A combination of step-wise and regular logistic regression was used to analyze multivariate models for the most significant determinants of financial stress. Variables that were logically correlated (e.g., interest paid and interest to gross income ratio) were not included simultaneously. Debt-to-asset ratio, a variable commonly used in other studies, was included in all models tested.

Criteria used to select the best model were (1) the model with the highest overall chi-square statistic significant at the 10 percent level of significance, (2) the model with the highest rho-square (pseudo-R square), and (3) the model with the highest correctly classified observations.

Model Results

The model selected includes the debt-to-asset ratio on January 1, 1985 [DBTAST85] and the operating ratio for 1985 [OPRAT5] to predict the probability of loan default and is shown as

$$\ln P/1-P = \underset{(52.4)}{-5.67} + \underset{(31.4)}{3.86}\ \text{DBTAST85} + \underset{(10.2)}{1.99}\ \text{OPRAT5}.$$

This model had a chi-square statistic of 51.5 which was significant at the 1 percent level. Individual chi-square statistics are shown in parentheses. All are statistically significant at the one percent level. Signs associated with individual parameters are correct according to economic theory. As both DBTAST85 and OPRAT5 increase, the probability of an individual being delinquent also increases.

The model correctly classified 88.2 percent of the observations when the breaking point was set at 0.5 (Figure 2).
Attempts to specify a model with more than two variables were not successful at the 10 percent level of significance.

Maximizing Lender Revenue

The model selected was evaluated for sensitivity of the percentage correctly classified and the number of loans written when the level of breaking point was varied from .095 to .5 (Figure 3). Percent of loans granted declines rapidly at first as PPROB is adjusted downward.

Calculations for maximizing net income are shown in Figure 4. Maximum net revenue occurred at a breaking point of .18. Slightly more than 80 percent of the observations were correctly classified at that breaking point.

The resulting values for the classification matrix are shown in parentheses in Figure 2; 280 borrowers (71.6 percent) are classified correctly as remaining current on debt payments. Thirty-three borrowers (8.4 percent) are correctly classified as delinquent (i.e., they actually defaulted on loans). Nearly 16 percent (62 observations) were incorrectly predicted as delinquent and only 4.1 percent (16) were incorrectly classified as current.

Solution of Model

The specified model may be used to calculate the probability of a farm operator being delinquent on debt payments. Evaluating both variables at their means (.354 for DBTAST85 and .940 for OPRAT5), the solution is as follows:

$$\text{Ln } P/1-P = -5.67 + 3.86\ (.354) + 1.99\ (.940)$$

$$\text{Ln } P/1-P = -2.433.$$

		PREDICTED	
		CURRENT	DELINQUENT
ACTUAL	CURRENT	I 338 (280)	II 4 (62)
	DELINQUENT	III 42 (16)	IV 7 (33)

PERCENT CORRECT = 88.2 (80.1)

Figure 2. Classification Table Results with PPROB at .5.
(Results when PPROB is .18 are in parenthesis).

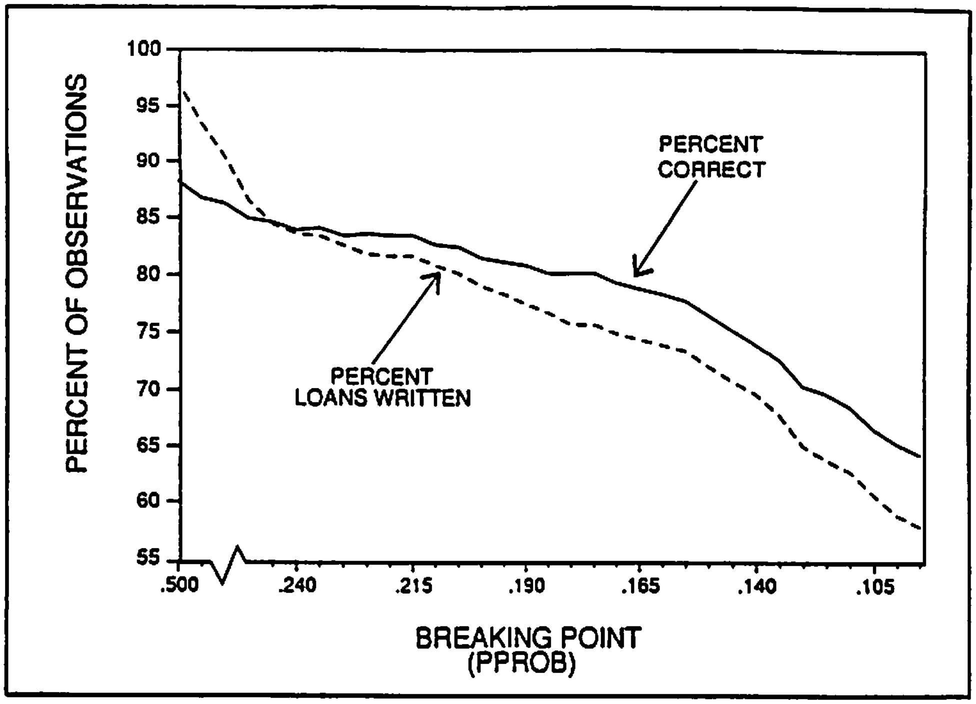

Figure 3. Percent Observations Classified Correct and
Percent Loans Written as PPROB is Adjusted.

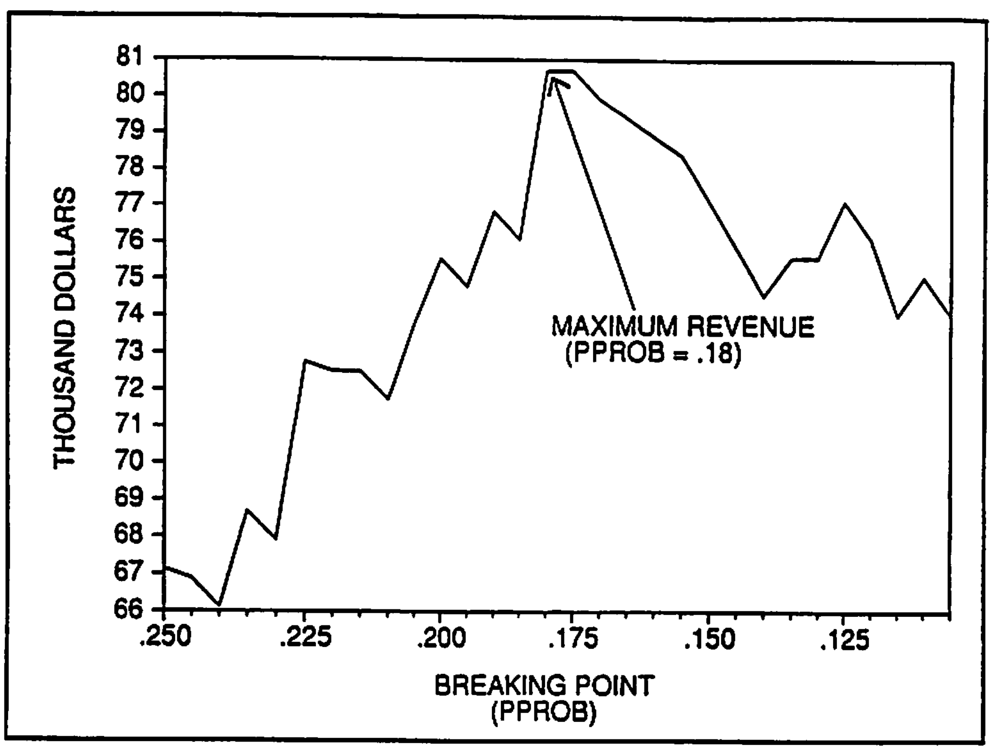

Figure 4. Interest Revenue to a Lender as PPROB is Adjusted.

Taking the antilog of both sides and solving, P is equal to .081. This indicates that for an average farm operator with a debt-to-asset ratio of .354 and an operating ratio of .940 the probability of being delinquent is 8.1 percent.

Varying the debt-to-asset ratio while holding the operating ratio at its mean is shown as the dashed line in Figure 5. The function exhibits the theoretical shape of the cumulative logistic probability function. Similar results can be derived for the operating ratio when debt-to-asset ratio is held at its mean.

The model may be solved to reveal the debt-to-asset ratio required to reach the breaking point of .18 by holding OPRAT5 at its mean and solving for DBTAST85. The left side is solved first using P = .18 and taking the log of both sides.

$$P/1-P = e^{Z_i}$$

$$.18/.82 = e^{Z_i}$$

$$.2195 = e^{Z_i}$$

$$\text{Ln } .2195 = Z_i.$$

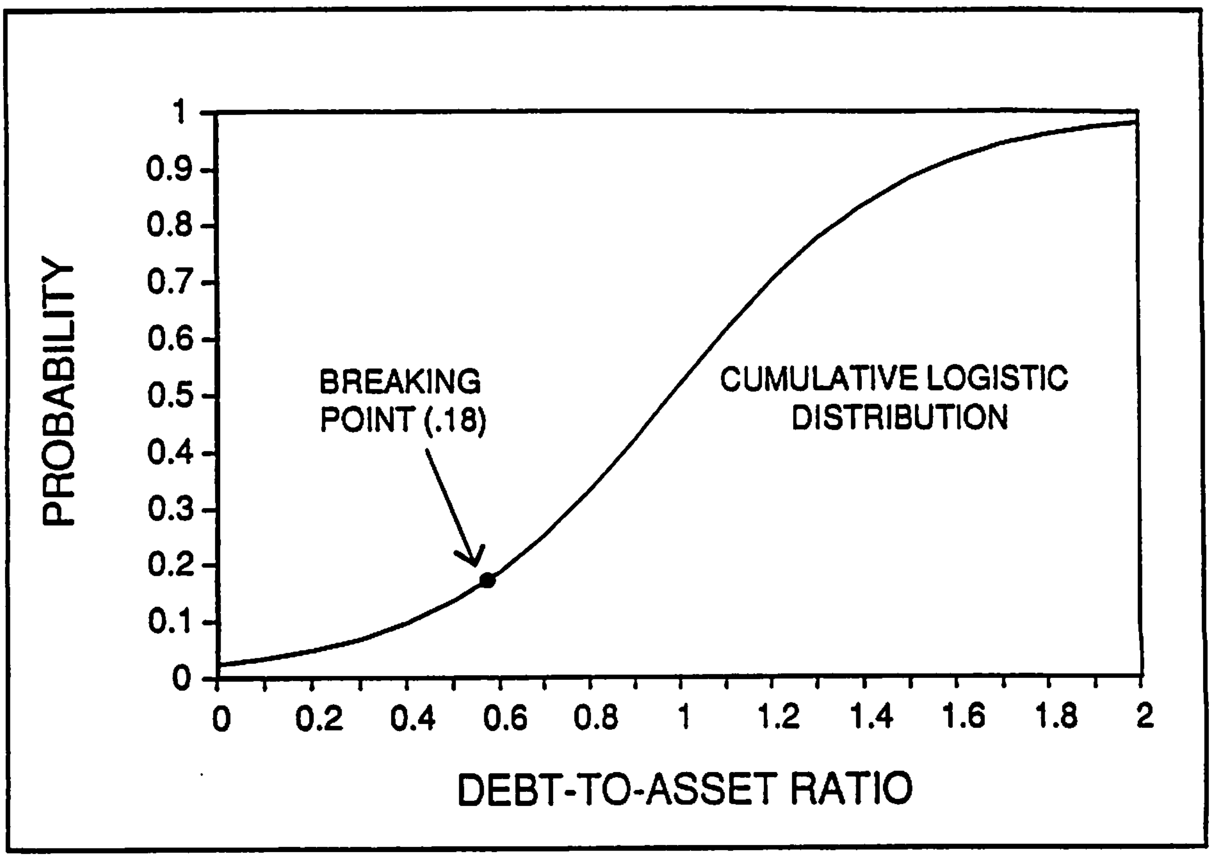

Figure 5. Probability of Being Delinquent Holding Operating Ratio at its Mean.

The model is then solved for DBTAST85.

$$\ln .2195 = -5.67 + 3.86\ \text{DBTAST85} + 1.99\ (.940)$$

$$\text{DBTAST85} = .591.$$

A debt-to-asset ratio of .59 is the threshold value where an individual has an 18 percent probability of being delinquent, given an operating ratio of .940.

Figure 6 shows the limit of debt-to-asset ratio and operating ratio possible to reach the breaking point frontier of .18. It is obvious that more profitable operations are able to carry more debt. This graph explicitly illustrates the possible combinations before the frontier is reached.

Models containing other variables also were significant, but less than for the one discussed above. Viability [VIAB85] and return on assets [NETAST5], when combined separately with debt-to-asset ratio, revealed model chi-square statistics of 38.2 and 41.1, respectively.

SUMMARY

The results of using logistic regression as a tool for classifying farm loan applications as current or delinquent appear promising. The model correctly classified 88.2 percent of the borrowers at a breaking point of .5. The classification of borrowers can be further enhanced by applying the loan pricing model and adjusting the breaking point to maximize profit to a lending institution.

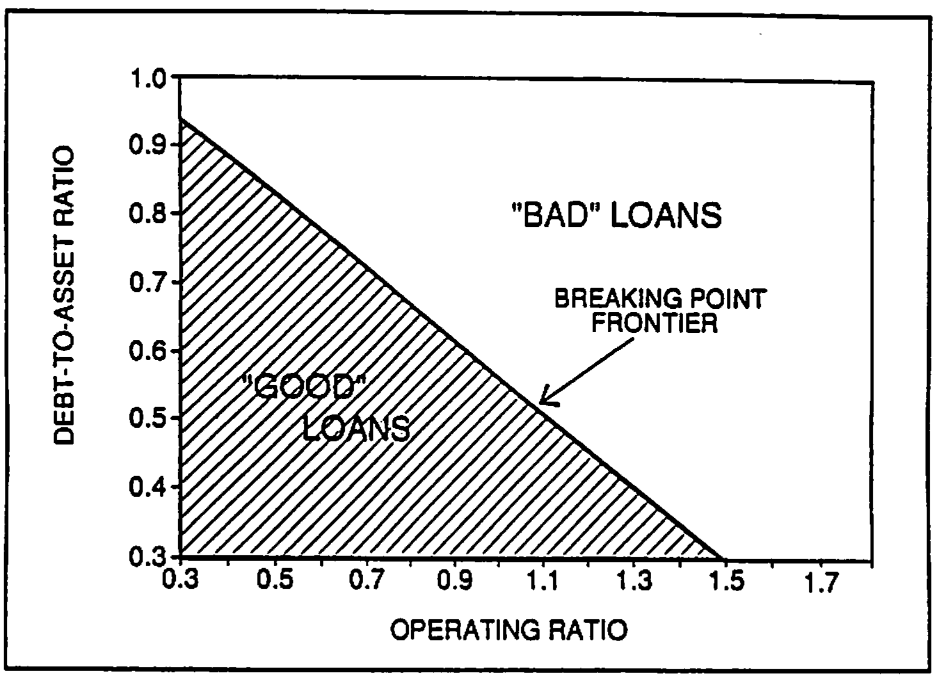

Figure 6. Values of Debt-to-Asset Ratio and Operating Ratio to Reach the Breaking Point Frontier.

When the breaking point was adjusted downward, borrowers were reclassified among quadrants in the classification table. The resulting 80 percent accuracy rate maximized profit to a lending institution. Some borrowers (16 percent) were predicted delinquent even though they did not eventually default on their obligations. The rejection of "good" borrowers is unavoidable in most situations, even using subjective analysis of loan applicants. The intent is to combine significant variables in an objective analysis of loan applications and use this as at least a partial guide for basing loan decisions.

Perhaps the most serious limitation of this study was that farms are not homogenous. Farm types vary widely across the state from farm operations exclusively growing small grain crops to various livestock operations.

Geographic location also is a problem when considering homogeneity of farms. Climate and soil characteristics vary widely across North Dakota and are factors that should be considered when farm characteristics are analyzed. Performance of farms can vary greatly under dissimilar conditions.

Profitability, management ability, and efficiency all vary not only among farm operators but among farm types and geographic regions as well. This makes identification of characteristics difficult at best as averages are for all farm types and geographic areas.

Asset valuations may be inconsistent among operators answering survey questions. Inventory changes and value of livestock and feed purchased are not specifically asked on questionnaires; therefore, return-to-equity and return-to-asset ratios cannot be calculated with accuracy.

A potential user of the model described here could use historic data from individual borrowers. This would localize weather, soil properties, and local farming practices. An institution would be able to tailor a model to its individual situation. Under a controlled local environment, a greater number of borrower attributes would become apparent, and could be combined with the subjective judgement of the loan officer in the loan decision process.

REFERENCES

Fiske, John R., Marvin T. Batte and Steven L. Richenbacker. Factors Influencing the Currentness of Debt Payments for Ohio Commercial Farmers. Ohio State University. Department of Agricultural Economics and Rural Sociology. ESO Number 1291. 1986.

Harrel, Frank E., Jr. "The LOGIST Procedure. "SUGI-Supplemental Library Users Guide. Version 5 Edition. SAS Institute Inc. Cary, North Carolina. pp 269-292. 1983.

Leholm, Arlen G., F. Larry Leistritz, Brenda L. Ekstrom, and Harvey G. Vreugdenhil. Selected Financial and Other Socioeconomic Characteristics of North Dakota Farm and Ranch Operators. North Dakota State University. Department of Agricultural Economics. Report Number 199. 1985.

Leistritz, F. Larry, Arlen G. Leholm, Brenda L. Ekstrom, and Harvey G. Vreugdenhil. "North Dakota Farmer's Views on Financial Assistance Policies. "North Dakota Farm Research. Volume 44, No. 1. July-August, 1986.

Leistritz, F. Larry, Wallie G. Hardie, Brenda L. Ekstrom, Arlen G. Leholm and Harvey Vreugdenhil. Financial, Managerial, and Attitudinal Characteristics of North Dakota Farm Families: Results of the 1986 Farm Survey. North Dakota State University. Department of Agricultural Economics. Agr. Econ. Report Number 222. 1987.

Luftburrow, Jean, Peter J. Barry, and Bruce Dixon. "Credit Scoring for Farm Loan Pricing." Agricultural Finance Review. Volume 44:8-14. 1984.

Melichar, Emmanuel. "A Farm Financial Perspective on Agriculture." Federal Reserve Bulletin. Federal Reserve Board. Washington, D.C. pp. 1-13. January, 1984.

Pindyck, Robert S. and Daniel L. Rubenfeld. Econometric Models and Economic Forecasts. 2nd Edition. McGraw-Hill, New York. pp. 273-312. 1980.

Press, James and Sandra Wilson. "Choosing Between Logistic and Discriminant Analysis." Journal of the American Statistical Association. Volume 73:699-705. 1978.

Salant, Priscilla, Melinda Smale and William Saupe. Farm Viability: Results of the USDA Family Farm Surveys. United States Department of Agriculture, Economic Research Service. Washington, D.C. July 1986.

Saxowsky, David M., Cole R. Gustafson, Mir B. Ali, Joan M. Braaten, and Jeff Rotering. Economic Impacts of North Dakota Laws that Permit Delayed or Partial Repayment of Agricultural Debt. North Dakota State University, Department of Agricultural Economics. Agr. Econ. Report Number 216. 1987.

Statistical Analysis System User's Guide: STATISTICS. 1982 Edition. SAS Institute, Cary, North Carolina.

United States Department of Agriculture, Statistical Reporting Service/North Dakota State University. Farm Finance Survey, January 1986. Fargo, North Dakota. February, 1986.

United States Department of Commerce, Bureau of the Census. Characteristics of the Population Below the Poverty Level. Government Printing Office. Series P-60 Number 152. Washington, D.C. 1984.

Watt, David L., James A Larson, Glen D. Pederson and Brenda L. Ekstrom. The Financial Status of North Dakota Farmers and Ranchers: January 1, 1985, Survey Results. North Dakota State University, Agricultural Economics Report Number 207. 1986.

APPENDIX

APPENDIX A

VIAB85 is an encompassing measure reflecting the well-being of the farm household. The ratio, based on a similar ratio used by Salant et al. (1986), is calculated as

$$VIAB85 = \frac{NETIN85 + DEPREC - ITDPAY + UNEARN + OFFRM + OTHER}{CONS85 + LTDPAY}$$

Where: NETIN85 = Net cash farm income in 1985;

DEPREC = Depreciation declared in 1985;

ITDPAY = Intermediate-term assets X .144;

OFFRM = Earned nonfarm income for 1985;

UNEARN = Unearned nonfarm income for 1985;

OTHER = Other nonfarm income for 1985;

CONS85 = Family living allowance for 1985;

LTDPAY = Long-term debt/20.

Net cash farm income including depreciation was reported in the survey in 1985. Depreciation is added to net cash farm income, and an allowance for asset replacement is made by subtracting 14.4 percent of reported intermediate term assets in the numerator [ITDPAY]. This places all farmers in a similar position concerning asset replacement allowance regardless of the reported amount of their intermediate term debt.

OFFRM includes off-farm income earned by the farm operator or spouse as wages during 1985. UNEARN includes income from oil and gas leases during 1985. Other non-farm income, such as interest from savings, income from investments, or net cash income from nonfarm businesses is included in OTHER. CONS85 is a minimum consumption allowance determined by the Bureau of Census (U. S. Department of Commerce, 1984) based on the number of persons living in the household. Allowances used for 1985 were adjusted for inflation using the consumer price index.

LTDPAY is the estimated principal payment on reported long-term debt (primarily land and buildings) in 1985. It is calculated as long-term debt divided by 20 years.

A viability ratio of 1.0 indicates a farm operator is able to meet family and farm business obligations for the year with all sources of income considered. A ratio less than one indicates insufficient cash flow to meet obligations. Ratios greater than one indicate cash flow sufficient to meet obligations, purchase additional assets, or to use as cash reserves. In the short run, operators may be spending dollars intended for asset replacement for consumption and debt payments. Increasing viability implies an increase in gross cash income or nonfarm income or a decrease in production expenses, interest expenses (due to both interest rate and debt level), or family living expenses.