AGRICULTURAL LOAN EVALUATION WITH DISCRIMINANT ANALYSIS *

R. Bruce Johnson and Albert R. Hagan

INTRODUCTION

Agricultural lending institutions are faced with a perpetual task of periodically evaluating personal and financial attributes of their borrowers. This examination is necessary in order to determine the present quality of the loans and to assess the current financial position of each borrower. Moreover, analysis of each borrower's financial performance establishes a basis for extending, limiting or withdrawing the present line of credit and for determining the amount and kind of supervision needed.

Presently, most analyses of borrowers' financial position are conducted via personal examination of individual credit files by either trained credit analysts or loan officers of various lending institutions. The initial objective of credit examination is to determine the current financial condition of each borrower and classify his loan into one of two possible categories, namely, acceptable loans (high quality loans requiring only normal supervision) or problem loans (weak loans possessing serious credit deficiencies and requiring more than normal supervision). Considerable time is required for a credit analyst to audit a borrower's loan records and to accurately determine his financial performance rating. For example, credit analysts of the Federal Intermediate Credit Bank of St. Louis estimate that 50 percent of their time is required to determine the financial performance of the problem, vulnerable, and loss loans. Yet loans in these groups comprise only 12 percent of the total loans made by Production Credit Associations in Missouri, Illinois and Arkansas [2].

Both farm operators and credit institutions would benefit from a more efficient use of credit resources if a simple mathematical model could be incorporated to analyze borrowers' credit files. More specifically, combining a quantitative credit scoring model with the speed and accuracy of a digital computer to analyze loans would result in several benefits. First, there would be a significant reduction in the man-hours (and associated costs) required for trained analysts to classify loans into acceptable and problem loan groups. Second, this savings in man-hours could be utilized in a more thorough analysis of the problem loans. Third, there could be a more frequent check on the quality of credit and direction of financial performance.

THE LINEAR DISCRIMINANT MODEL

Discriminant analysis is a statistical tool which lends itself to classifying items into predetermined populations. The linear discriminant model has been used previously to quantify the credit rating of both consumer and agricultural loans [1, 4, 5]. The technique of discriminant analysis is based on the assumption that a linear function $Y = B_1X_1 + B_2X_2 + ... + B_nX_n$ exists which will distinguish between elements of a population. The discriminant model utilizes coefficients $B_1, B_2, ..., B_n$ chosen in such a way that the ratio of between-groups sum of squares is maximized. Therefore, the index Y represents the optimum discriminator between the groups. Factors

R. Bruce Johnson is assistant professor of agricultural economics and agribusiness at Louisiana State University and Albert R. Hagan is professor of agricultural economics at the University of Missouri.

*Missouri Agricultural Experiment Station Journal Series No. 6826.

X_1, ..., X_n represent the quantifiable characteristics of the loans.

With regard to the actual mathematic description of the model, let the n factors X_1, X_2, ..., X_n be the n characteristics of the agricultural loan. If there are h categories of loans, each category having M_k, (j = 1, 2, ..., h) individual loans then the tuple $(X_{1j,k}; X_{2j,k}; X_{3j,k}; ...; X_{nj,k})$ represents the data vector for loan k in category j.

Detailed theoretical and computational procedures for determining the discriminate coefficients are readily available [3]. However, some important facets of the typical analysis are outlined below:

1. Assumption: The data vector is assumed to be multivariate-normal in distribution to facilitate tests of hypotheses and classification routines. The covariance structure among the variables in the data vector is assumed to be constant within each loan category.
2. The discriminate coefficients are chosen to maximize the ratio of among to within groups variance in discriminate scores. These coefficients are dimensionless and their ratio is important, not their value.

Diagrammatically, the expected proportion of correct classifications is illustrated in Figure 1. The diagram denotes the situation where there are two populations and only one variable, i.e., M = 2 and n = 1. The figure assumes that the samples are large enough that all the population parameters can be regarded as known.

Since the variance of Y (which is assumed to be the same in the two populations) and the population means are known, the likelihood of an observation being classified into either Population 1 or Population 2 is determined by consulting a table of normal distributions. The likelihood of an observation receiving a classification into either Population 1 or Population 2 are equal at Y_c. One would classify all cases where $Y>Y_c$ in Population 2. Conversely, all observations where $Y<Y_c$ would be classified in Population 1. The shaded area in Figure 1 represents the expected proportion of misclassified cases.

The Cut-Off Point

If one assumed that the two kinds of errors, that is, classifying an acceptable loan into the problem group and classifying a problem loan as acceptable, are of equal significance, the cut-off point would be Y_c on Figure 1. This point can be determined algebraically:

$$Y_c = \frac{\sigma_A \bar{Y}_P + \sigma_P \bar{Y}_A}{\sigma_A + \sigma_P}$$

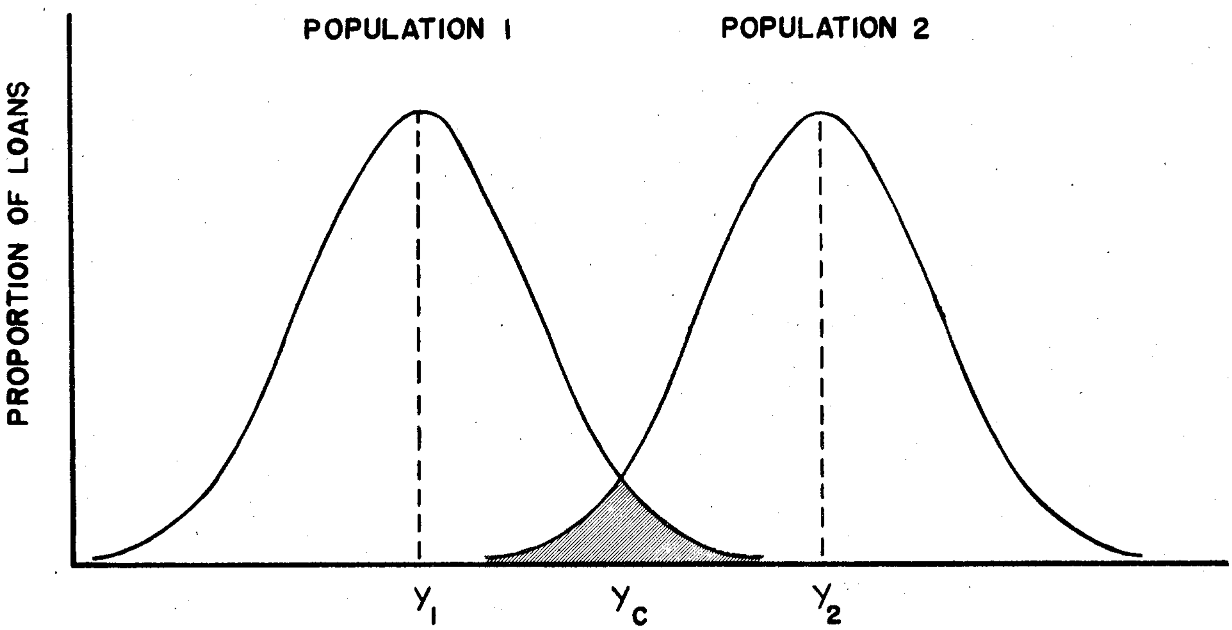

Figure 1.CLASSIFICATION FOR TWO POPULATIONS AND ONE VARIABLE: POPULATION PARAMETERS KNOWN

where

σ_A = the standard deviation of the Y-values for the acceptable loan group

σ_P = the standard deviation of the Y-values for the problem loan group

$\overline{Y}_A$ = the mean Y value for the acceptable loan group

$\overline{Y}_P$ = the mean Y value for the problem loan group.

After determining the cut-off point (Y_C) a Z statistic[1] can be computed for both $\overline{Y}_A$ and $\overline{Y}_P$. The Z statistic for $\overline{Y}_A$ is determined according to the following formula:

$$Z_A = \frac{Y_C - \overline{Y}_A}{\sigma_A}$$

Similarly, a Z statistic can be computed for $\overline{Y}_P$:

$$Z_P = \frac{Y_C - \overline{Y}_P}{\sigma_P}$$

Referring to a Z table we can determine what percent of acceptable and problem loans will be misclassified.

The F-Test

The null hypothesis that the discriminant function does not discriminate between acceptable and problem loans can be tested by an analysis of the variance of Y. The F value is computed from the following ratio:

$$F = \frac{\text{Sum of Squares}/n \text{ (between loan groups)}}{\text{Sum of Squares}/M_A + M_P - n - 1 \text{ (within loan groups)}}$$

where:

n = number of X's,

M_A = number of acceptable loans, and

M_P = number of problem loans.

Given the appropriate probability level, if the computed value of F is greater than the tabled value of F with n and $M_A + M_P - n - 1$ degrees of freedom, the discriminant function effectively discriminates between the two groups of loans.

Data

Data for the study were collected from loan applications of borrowers at three Production Credit Associations located in central and northwestern Missouri. The president of each association provided a list of all borrowers who had current loans from the PCA. Each loan had been examined recently (by credit analysts of the Federal Intermediate Credit Bank of St. Louis) and classified as problem or acceptable. Data used in the analysis were from 204 acceptable and 68 problem loans.

The Variables

Several variables were considered for analysis and subsequent inclusion in the discriminant model. Following extensive testing and evaluation of various measures of financial performance, the following three variables were selected for use in the final model:

X_1 = Repayment index. The amount of the loan actually repaid each year plus the value of marketable crops and market livestock not sold during the year was expressed as a percentage of the amount expected to be repaid. This index was computed for the current year only.

X_2 = Current ratio. The ratio of current assets to current liabilities computed for the most recent financial statement.

X_3 = Debt-asset ratio. Total debts divided by total assets for the most recent financial statement.

Turning now to an examination of financial ratios computed for each group of sample borrowers, data in Table 1 reveal some sizable differences in the values of the three measures of financial strength of the 272 sample borrowers. The ratio of repayment made (plus marketable crops and market livestock) to repayment expected is a measure of the repayment performance of the borrower. The mean values for this ratio clearly indicate that the acceptable loan group has a better repayment performance than the problem loan group. The difference for the repayment index between loan groups is .46.

The current ratio is often referred to as a measure of a borrower's liquidity. Data in Table 1 show the acceptable borrowers are in a more liquid position than those in the problem loan group. The numerical difference in the average current ratio between the acceptable and problem loan groups is 3.76.

The debt-asset ratio is a measure of longer term financial strength. Since the debts comprise only 27 percent of the assets of the acceptable group and since the corresponding figure for the problem loan group is 56 percent, it appears that borrowers in the acceptable loan group are in a stronger financial

[1] The Z statistic is distributed almost normally with variance $\sigma_Z^2 = \frac{1}{n-3}$.

Table 1. MEAN VALUES OF VARIABLES INCLUDED IN THE LINEAR DISCRIMINANT FUNCTION, ACCEPTABLE AND PROBLEM LOANS, PRODUCTION CREDIT ASSOCIATIONS, MISSOURI, 1969

Variable	Loan Classification Group		Difference
	Acceptable	Problem	Between Groups
Repayment index	1.12	0.66	0.46
Current ratio	5.24	1.48	3.76
Debt-asset ratio	0.27	0.56	-0.29

position than those in the problem loan group.

Development of the Discriminant Model

The discriminant model was developed on the basis of the application of discriminant analysis to data from the 204 acceptable and 68 problem loans. Applying the estimated coefficients, the specific linear discriminant function for the 272 loan observations was:

$$Y = 0.02525\, X_1 + 0.0091\, X_2 - 0.04502\, X_3$$

where:

X_1 = one year repayment index,

X_2 = the ratio of current assets to current debts, and

X_3 = the ratio of total debts to total assets.

In order to test the null hypothesis that the discriminant function does not discriminate between acceptable and problem loans, analysis of the variance of Y was conducted. As illustrated in Table 2, the calculated F value (128.65) was found to be significant at the .01 level of significance.

The means, variances, and standard deviations were computed for the discriminant functions of the two borrower groups. Values of these estimated parameters are illustrated in Table 3. These estimates, based on large samples, will be treated as population parameters in the discussion that follows.

To classify agricultural loans with the discriminant model, a critical value for Y must be established. If we assume that the two kinds of errors in misclassification are of equal significance, the critical or cut-off value can be calculated by the previously discussed method where

Table 2. ANALYSIS OF VARIANCE OF ESTIMATED DISCRIMINANT FUNCTION, ACCEPTABLE AND PROBLEM LOANS, PRODUCTION CREDIT ASSOCIATIONS, MISSOURI, 1969

Source	d.f.	S.S.	M.S.	F
Between Groups	3	0.040528	0.013509	128.65*
Within Groups	268	0.028190	0.000105	

*F-value significant at .01 level of significance.

Table 3. THE MEANS, VARIANCES, AND STANDARD DEVIATIONS FOR THE DISCRIMINANT MODEL, ACCEPTABLE AND PROBLEM LOANS, PRODUCTION CREDIT ASSOCIATIONS, MISSOURI, 1969

Loan Classification Group	Sample Size	Mean Discriminant Value	Variance	Standard Deviation
Acceptable	204	0.02075	0.00011	0.01050
Problem	68	-0.00744	0.00009	0.00931

$$Y_C = \frac{\sigma_A \overline{Y}_P + \sigma_P \overline{Y}_A}{\sigma_A + \sigma_P}$$

$$= \frac{(0.01050)(-0.00744) + (0.00931)(0.02075)}{(0.01050) + (0.00931)}$$

$$= 0.00581$$

After calculating this cut-off point, Z_A and Z_P would be

$$Z_A = \frac{Y_C - \overline{Y}_A}{\sigma_A} \qquad Z_P = \frac{Y_C - \overline{Y}_P}{\sigma_P}$$

$$= \frac{(0.00581) - (0.02075)}{(0.01050)} \qquad = \frac{(0.00581) - (-0.00744)}{(0.00931)}$$

$$= -1.42 \qquad = 1.42$$

Referring to a table of values for "cumulative normal frequency distribution," the computed Z values indicate that the discriminant function would correctly classify 92 percent of the borrowers.

When applying the cut-off score to computed Y values for a group of agricultural loans, those loans with Y values equal to or greater than 0.00581 would be classified acceptable while those with Y values less than 0.00581 would be classified into the problem group. There is an eight percent probability that loans in each group would be misclassified.

This method of loan classification would be suitable if the consequences of the two possible classification errors were of equal significance. However, since the computer credit scoring model will replace the credit analysts' personal examination and since all problem loans need to be reviewed annually, a more precise classification scheme is needed. Thus, the probability of misclassifying problem loans into the acceptable loan group must be reduced to a more tolerable level. Credit analysts of the FICB indicated a one percent misclassification level could be tolerated. Therefore, a cut-off score which has a .01 probability for misclassification of problem loans was calculated. This alternative cut-off score was specified as the critical Y value (CV) for classifying loans.

Consulting a table of cumulative normal frequency distribution, the appropriate critical value was derived through the following calculation:

$$Y_{CV} = \overline{Y}_P + (Z)\sigma_P$$

where:

Y_{CV} = critical Y value,
$\overline{Y}_P$ = mean Y value for the problem loan group,
Z = standard measure, and
σ_P = standard deviation of $\overline{Y}_P$.

The appropriate value of Z which allows a one percent misclassification tolerance was 2.33. Thus, multiplying the standard measure times the standard deviation of the sample mean (σ_P) and adding this product to the sample mean ($\overline{Y}_P$) results in a critical Y value. Assuming the sample mean ($\overline{Y}_P$) score approximates the population mean, there is only one chance out 100 of misclassifying a problem loan into the acceptable loan group.

In order to test the discriminant function on borrowers' loan data, the following critical Y value was calculated:

$$Y_{CV} = -0.00744 + 2.33\,(0.00931)$$
$$= 0.01425$$

Thus all loans receiving Y scores equal to or greater than 0.01425 were classified into the acceptable loan group. Conversely, loans with Y scores less than the CV were categorized into the problem loan group.

Application of the Discriminant Model

To verify the effectiveness of the discriminant function, the coefficients were applied to appropriate financial data of borrowers of the Mississippi Valley Production Credit Association, Pittsfield, Illinois. A total of 378 loans were selected for analysis (Table 4). Credit analysts of the Federal Intermediate Credit Bank of St. Louis previously had analyzed the loans and classified them as either acceptable or problem. Three hundred of the loans rated the acceptable loan classification while 78 were classified as problem loans (including two loss loans and 24 loans which were rated acceptable with significant credit weaknesses).

Application of the discriminant model to the Mississippi Valley PCA borrowers' financial data resulted in the correct classification of 156 of the 300 acceptable loans (Table 4). In addition, only one of the 78 problem loans was misclassified into the acceptable loan group. Thus, 61.6 percent of the 378 loans were accurately categorized into their respective loan classification groups.

Following intensive testing of the discriminant model on financial data of borrowers of Production Credit Associations in Missouri, Illinois and Arkansas the Federal Intermediate Credit Bank of St. Louis asked for and received approval from the Farm Credit Administration to utilize the model for credit scoring Production Credit Association loans in the Sixth Farm Credit District. This program was implemented in October, 1971.

Staff members of the St. Louis FICB developed detailed instructions for use by PCA members for the credit scoring program, using the discriminant model. A "Credit Scoring Form" was prepared for transmitting pertinent data for each borrower

Table 4. **NUMBER OF ACCEPTABLE AND PROBLEM LOANS CLASSIFIED BY FICB CREDIT ANALYSTS AND THE LINEAR DISCRIMINANT MODEL, MISSISSIPPI VALLEY PRODUCTION CREDIT ASSOCIATION, PITTSFIELD, ILLINOIS, 1970**

Method of Classification	Loan Classification Group	
	Acceptable	Problem
	---- number of loans ----	
Credit analyst	300	78
Discriminant function	156	222

included in the computerized program. Some loans in each association are omitted from the computerized scoring and are evaluated personally by the credit representatives (examiners) during visits to the association. These omissions include extremely large loans, small loans of a routine nature, and special loans with unique features. Actually, the purpose is to identify and classify a large percentage of the "acceptable" loans in order to achieve the following benefits, as explained in a memorandum to the PCA presidents.

1. Reduce credit examination costs.
2. Reduce the man hours needed to classify the obviously acceptable loans, thereby allowing more time for those loans requiring more attention and in-depth analysis.
3. Create greater opportunity for credit representatives to assist the associations in credit training and specialized loan handling.
4. Provide credit scoring index information that will be useful to the PCA's in their credit administration.

Staff members of the FICB report that these objectives are being achieved and they are quite pleased with the performance of the new credit scoring program. Since instituting the program, credit representatives in the three-state district now have more time to assist association personnel with the improvement of lending procedures.

REFERENCES

[1] Durand, David, *Risk Elements in Consumer Installment Financing,* National Bureau of Economic Research, New York, 1941, p. 125.

[2] Federal Intermediate Credit Bank of St. Louis, *Summary of Operations – Production Credit Associations,* Sixth Farm Credit District, 1969 (St. Louis) p. 8.

[3] Rao, C. R., *Advanced Statistical Methods in Biometric Research,* John Wiley & Sons, New York, 1952.

[4] Reinsel, Edward Ignatius, *Discrimination of Agricultural Credit Risks from Loan Application Data,* unpublished Ph.D. dissertation, Michigan State University, 1963, pp. 1-122.

[5] Zellner, Robert E., Jr., *Determining the Quality of Consumer Credit,* unpublished Master's Thesis, University of Florida, 1968, p. 1.