Philippine Review of Economics and Business
Volume XVII, Nos. 1 & 2, March & June, 1980

PREDICTION OF AGRICULTURAL LOAN REPAYMENT PERFORMANCE

By Dominador A. Clemente, Jr.*

1. Introduction

Agricultural credit in the Philippines is provided by institutional sources and non-formal channels. Studies (PCAC, 1977) indicate that private money lenders accounted for as high as 90 per cent of total credit during the period 1954-61 despite rates of interest from 50 to 65 per cent per annum. However, there was a marked shift to institutional sources during the period 1975-76 when formal sources supplied 65 per cent of total loans to palay farmers in Central Luzon. This can be attributed to the development of the countryside banking network and the supervised credit program.[1]

Attempts to liberalize credit to farmers go back to the 1950s. In 1952, the Philippine Legislature passed RA 720 creating the rural banks, and RA 820 establishing the Agricultural Credit and Cooperative Financing Administration (ACCFA, now ACA). The rural banks extend credit on reasonable terms while ACCFA extended credit based on the productive capacity of the farmers. However, it was only in crop year 1964-65 that the supervised credit program was initiated by the rural banks. Under this scheme, farmers could borrow without collateral as long as they agreed to be supervised in their farm operations by technicians.

---

*Associate Professor, College of Business Administration, University of the Philippines. This paper is based on the author's D.B.A. dissertation (1980).

[1] Central Bank Circular No. 474 dated June 30, 1975 formally defines supervised credit as "a system of lending wherein the farmer-borrower agrees in writing that he will apply proven farm practices necessary to conserve the land, improve its fertility and increase its production, and abide by the approved farm plan and budget prepared by an accredited supervised credit technician."

One of the programs involving supervised credit is the Masagana 99 Program (M-99) implemented in May 21, 1973 to increase the national average yield from 40 to 99 cavans of palay per hectare. This program involves a package of technology consisting of high yielding varieties, fertilizer, agricultural chemicals, production loans, extension services, mass media coverage, marketing schemes, and general management coordination and evaluation services.

Noncollateralized institutional credit extending for 180 days is provided by the Philippine National Bank (PNB), rural banks and the Agricultural Credit Administration. At the start of the M-99 Program, production loans ranged from ₱700 to ₱900 per hectare and were later raised to ₱1,350 per hectare. Interest on loans has been pegged at 10 per cent per annum with a service fee of 2 per cent. An additional 2 per cent per annum is charged overdue loans.

As a whole, the M-99 credit program, while liberalizing credit to small farmers, also proved to be costly for the lending institutions because of the small size of the loans and the high delinquency rates. The magnitude of the repayment problem can be seen in Table 1.

The first two phases of the program registered 94.0 and 93.7 per cent repayment, respectively. However, the repayment performance in succeeding phases declined to as low as 57.4 per cent in Phase XII. In terms of absolute values, a total of ₱3,985.8 million was granted in 1979, with ₱3,223.3 million representing loans repaid, for an overall repayment rate of 80.9 per cent.

The uncollected amount of ₱761.7 million poses a serious threat to the viability of the supervised credit program. Although the limits to the capacity of the credit institutions are not known, they should not be expected to perform a social welfare function. This does not imply, though, that collateral should be required because this will screen out small farmers with the capacity to repay. Even farms which may be uneconomic units at present should be provided with credit because they can be made viable through diversification of farm projects, processing and marketing of farm products.

To detect the viability of the borrower requires the development of a loan evaluation process. Under the M-99 Program, a bonafide farmer may be granted a loan provided he has an approved farm plan and budget and he is a member of a *selda*.[2] Since these two re-

---

[2] In Phase XII of the M-99 Program, a farmer who can put up the necessary collateral need not join a selda.

Table 1 — Masagana 99 Repayment Status (in million pesos)

Financing Institution	Loans granted (pesos)	Loans repaid (pesos)	Percent repayment rate
Phase I (May-October 1973)			
RB*	152.9	150.2	98.2
PNB*	195.4	178.1	91.1
ACA**	21.2	19.1	90.1
Total	369.5	347.4	94.0
Phase II (November 1973-April 1974)			
RB	117.2	114.9	98.0
PNB	101.9	93.2	91.5
ACA	11.6	8.1	69.8
Total	230.7	216.2	93.7
Phase III (May-October 1974)			
RB	303.5	285.3	94.0
PNB	382.0	293.4	76.8
ACA	30.6	21.9	71.6
Total	716.1	600.6	83.9
Phase IV (November 1974-April 1975)			
RB	333.3	287.1	86.1
PNB	225.3	173.2	76.9
ACA	13.7	9.5	69.3
Total	572.3	469.8	82.1
Phase V (May-October 1975)			
RB	235.4	191.4	81.3
PNB	319.1	230.7	72.3
ACA	18.5	14.9	80.5
Total	573.0	437.0	76.3
Phase VI (November 1975-April 1976)			
RB	127.2	106.1	83.4
PNB	121.7	98.5	80.9
ACA	6.7	4.9	73.1
Total	255.6	209.5	82.0

* As of July 31, 1979

** As of September 30, 1979

Table 1 (Continued)

Financing Institution	Loans granted (pesos)	Loans repaid (pesos)	Percent repayment rate
Phase VII (May-October 1976)			
RB	139.1	115.4	83.0
PNB	110.8	88.1	79.5
ACA	24.2	15.5	64.0
Total	274.1	219.0	79.9
Phase VIII (November 1976-April 1977)			
RB	78.5	68.7	87.5
PNB	73.7	58.5	79.4
ACA	12.0	6.7	55.8
Total	164.2	133.9	81.5
Phase IX (May-October 1977)			
RB	114.1	96.0	84.1
PNB	117.3	87.6	74.7
ACA	18.6	9.7	52.2
Total	250.0	193.3	77.3
Phase X (November 1977-April 1978)			
RB	95.3	75.1	78.8
PNB	70.3	62.1	88.3
ACA	9.5	5.1	53.7
Total	175.1	142.3	81.3
Phase XI (May-October 1978)			
RB	118.9	73.5	61.8
PNB	103.9	76.3	73.4
ACA	14.2	7.8	55.6
Total	237.0	157.7	66.5
Phase XII (November 1978-April 1979)			
RB	85.6	39.5	46.1
PNB	67.8	49.1	72.4
ACA	14.8	8.0	54.1
Total	168.2	96.6	57.4
Phase I-XII			
RB	1,901.0	1,603.2	84.3
PNB	1,889.2	1,488.8	78.8
ACA	195.8	131.3	67.1
Total	3,985.8	3,223.3	80.9

Source of Data: National Food and Agriculture Council.

quirements are easy to comply with, they do not indicate the lenders' risks arising from defaults on loan payment.

Even before the inception of the M-99 Program, studies have been conducted to determine the causes of delinquency among farmers. But the factors found to have significant effects on repayment were never used to classify loan applicants as good or bad risks. The reluctance of policymakers to impose a credit scoring process could be due to the fact that small farmers will be the ones adversely affected. The credit scoring process proposed in this study is not intended to reject applicants but to classify the good from the bad risks so that corrective measures can be instituted. Those classified as good risks may be exempted from the *selda* membership or eventually graduated from the program. On the other hand, bad risk applicants may be supervised more closely or extended assistance in marketing their products.

Objectives of the Study

The purpose of this study is to develop a reliable credit scoring system using multiple discriminant analysis (MDA) to evaluate the credit worthiness of M-99 loan applicants. The specific objectives of the study are as follows:

1. to identify the variables associated with low and high repayment performance and to determine the relative contributions of these variables to repayment; and

2. to evaluate the usefulness of the MDA in practical decision-making situations.

Hypothesis of the Study

The hypothesis to be tested is that a set of attributes can discriminate well between high and low repayment performance of M-99 borrowers.

To test the hypothesis, 23 independent variables were identified from the data as possible discriminants between good and bad borrowers. The dependent variable corresponds to the classification categories, i.e., high and low repayment performance. Assuming the good risks will have higher discriminant scores, the independent variables and their hypothesized effects on the scores are identified as follows:

Variable	Hypothesized Increase (+) or decrease (—) in discriminant scores
Household	
Age of farmer	+
Number of years in school	+
Number of years in farming	+
Size of farm family	—
Dependency ratio	—
Total consumption	—
Total income	+
Attitudinal	
Attitude towards credit	—
Attitude towards saving	—
Attitude towards repayment	—
Farm Business	
Effective crop area	+
Total palay production	+
Fixed assets	+
Current assets	+
Total Liabilities	—
Net Worth	+
Value of palay sold	+
Value of other crops sold	+
Value of livestock sold	+
Production cost of palay	+
Production cost of other crops	—
Production cost of livestock	—
Number of techicians' visits	+

2. Methodology

This section describes the procedures employed in the study. The variables that can discriminate between low and high repayment borrowers were obtained through discriminant analysis. Also discussed are the criteria for classifying the repayment groups, the validation methods, and the ramifications of the "ambiguity zone" or "gray area".

The Statistical Procedure

Although not as popular as regression analysis, MDA has been used in a number of disciplines mainly in the biological and behavioral sciences since it was first introduced in the 1930s. It was only recently, however, that this technique was applied to financial problems such as credit evaluation and investment classification. In fact, one of the first successful business applications of MDA was in credit selection.

MDA is a statistical technique designed to classify an observation by means of a set of independent variables into one or more mutually exclusive and exhaustive *a priori* groupings. This objective is achieved by the statistical decision rule of maximizing the ratio of among-group to within-group variance-covariances on the profile of discriminating (predictor) variables. MDA is appropriate if the single dependent variable is either dichotomous (e.g. male-female) or multichotomous (e.g., high-medium-low) and therefore nonmetric (Frank, et. al., 1965; Massy, 1962; Robertson and Kennedy, 1968; and Sheth, 1971). When the dependent variable is a random variate, regression analysis is used (Robertson and Kennedy, 1962).

The first step, therefore, is to determine explicit group classifications. The number of original groups can be two or more. MDA then attempts to derive a linear combination of these characteristics which "best" discriminate between the groups. The use of a linear classification procedure allows a clear interpretation of the effect of the independent variables. If a more complex nonlinear discriminant function is used, it is more difficult to isolate the effect of each variable (Morrison, 1969).

The MDA technique has the advantage of considering an entire profile of characteristics common to the sample under consideration as well as the interaction of these properties. A univariate study can only consider the measurements used for group assignments one at a time. Another advantage of MDA is the reduction of the analyst's space dimensionality. If G is the number of original *a priori* groups, the number of different independent variables reduces to G-1 dimension(s). Since this study is concerned with only two groups, i.e., high and low repayment borrowers, the analysis is transformed into its simplest form: one dimension. (The details of this technique are discussed in Avery and Eisenbeis, 1972).

The Classification of Groups

A common problem encountered in discriminant analysis is the definition of groups. An ideal classification scheme exists if the groups are truly distinct and non-overlapping as in Fisher's problem for separating species of Iris. Unfortunately, there are no rules for classifying the respondents when the groups overlap.

The classification of groups in this study follows the sampling scheme which drew samples from towns classified as having either a high or a low repayment performance. The advantage of this method is that the classification of the borrowers into high and low groups was determined by averaging their repayment performance over phases. This would minimize classification errors due to changes in the status of borrowers from high to low repayment or vice versa over time.

To develop a credit scoring device for M-99 borrowers, it is more practical to initially classify applicants as either good (high repayment) or bad (low repayment) credit risks. Good risks will include fully paid and partially paid borrowers who paid at least 60 per cent or more of their loans. Bad risks will include totally delinquent and partially paid borrowers who paid less than 60 per cent of their loans.

Estimating the Discriminant Function

In developing the final discriminant function, the initial step was to group separately the low and the high repayment borrowers. Then each group was split into analysis and validation subsamples with five replications presented in matrix form as follows:

		Validation subsample			
	1st half	2nd half	all odd	all even	random
Analysis Subsample: 1st half		x			
2nd half	x				
all odd				x	
all even			x		
random					x

For each replication, several computer runs were made based on several criteria. Initially, five discriminant functions were generated through step-wise MDA with the following options:

a. minimum Wilk's lambda
b. minimum Mahalanobis distance between groups
c. largest minimum between-groups F
d. minimum residual variation
e. largest increase in Rao's V.

In subsequent iterations, the replicate that had the highest predictive accuracy was used as the analysis subsample.

A second discriminant function was computed using variables employed in other studies. Towards this end, a listing of variables found to be significant in other studies was made. A variable that was found to be significant in at least two studies was considered with the proviso that the variables included in the analysis must appear in the list of variables used in this study.

In addition, a discriminant function was computed using the variables selected based on the judgment of personnel involved in the M-99 Program. To obtain the required data, a survey was undertaken. A list of variables used in this study was shown to the respondents who were asked to indicate the variables deemed important in affecting repayment. The respondents then ranked the variables according to their effects on repayment. A variable that was considered important by at least half of the respondents was included in the analysis. Finally, separate discriminant functions were computed based on household and attitudinal variables, and farm business variables.

The predictive results of the various iterations were then compared to determine which discriminant function provided the greatest predictive accuracy. The "best" model was then subjected to further refinements through logarithmic transformation and standardization of outliers to ± 2 SD. The final discriminant function was then selected considering the various iterations and refinements. Since the process is essentially iterative, there is no claim as to the optimality of the final discriminant function. The various iterations above are necessary to reduce the number of variables. Including all the variables in the list will not only be impractical but some of the data required might be difficult or expensive to obtain.

The usual approach in reducing the number of variables to a manageable size is either through univariate significance tests or stepwise procedures. This approach would be appropriate if the objective is to maximize the separation among groups while minimizing the number of variables. On the other hand, if the goal is to construct a classification scheme, elimination of seemingly insignificant variables may affect the classification results. In a study made by Altman, et. al. (1974), reduction in the number of variables decreased the accuracy of the model from 81 to 62 per cent correct classification.

In this study, the overall classification results incorporating all variables was examined first as suggested by Eisenbies (1977) before reducing the number of variables. By this method, the effects or "costs" of variable reduction can be ascertained.

Validation of the Discriminant Model

After the values of the discriminant coefficients were obtained, each observation in the sample was assigned to one of the groups based on the calculated discriminant scores. A high degree of correct classification is expected because the discriminant function is based on the individual measurements of the same borrowers being classified. The high accuracy or upward bias of the classifications may be due to: (a) sampling errors in the original sample and (b) search bias. The latter results from the reduction of the original set of variables to the best variable profile. This bias is inherent in any empirical study.

Suggested methods to test the accuracy of the model are to use holdout and/or secondary samples (Altman, 1968; Deakin, 1972; Pinches and Mingo, 1973; Altman, et. al., 1974; and Sinkey, 1975)[3].

In the absence of a secondary sample, the validation procedure using holdout samples described by Frank, et. al. (1965) was used in this study. The essence of this method is to estimate the parameters of the model using only a subset of the original sample and then classifying the remainder of the sample based on the computed parameters. A simple t-test is then applied to test the significance of the results. Five replications of this method were employed :

1. Random sampling
2. Choosing every other borrower starting with borrower number one
3. Starting with borrower number two
4. Choosing the first half of the sample
5. Choosing the second half

[3] A holdout sample shall mean a subset of the original sample while a secondary sample involves new observations.

To further assess the validity of the model, a subjective classification scheme was used to classify a subset of the farmer-borrowers with given characteristics into good or bad credit risks. This "naive" classification procedure is as follows:

1. For each respondent, the raw values of the variables identified by MDA are compared with the mean values for the province.
2. If the raw value of a variable is equal or greater than the mean value, the respondent is classified as a good risk. Otherwise, he is classified as a bad risk. This process is continued until the respondent has been classified on all the variables.
3. The average rating of the respondent is then determined if on the whole, he is either a good or a bad risk.

Since we have prior knowledge of the membership of the borrowers in either group, the accuracy of this "naive" classification scheme can be determined. It would be of interest to know if the results obtained from this "naive" procedure are better than those obtained with the use of a computer.

"Ambiguity Zone" or "Gray Area"

As pointed out earlier, the classification of the borrowers into high and low repayment followed the sampling scheme. We would, therefore, expect minimum overlaps between the two groups. Nevertheless, it is still possible that a large shaded area will occur which will undermine the predictive ability of the model.

Although a cutoff point can be established for the respondents being studied, we would be uncertain about the classification of a new applicant whose score falls within the "ambiguity zone." Following Massy (1982) and Altman (1968), the process begins by identifying all the sample observations that fall within the overlapping range. Then the range of scores that results in the minimum number of misclassifications is determined. The best critical value is the midpoint of the interval corresponding to the minimum number of misclassifications.

3. Empirical Results of the Discriminant Analysis

This section presents the results of the discriminant analysis conducted for the Nueva Ecija borrowers. In essence, the analysis attempts to separate the two groups (low and high repayment) by clustering the borrowers in each group around the group's centroid and at the same time to separate the centroids as far apart from each other as possible to minimize the overlapping range.

Development of the Discriminant Models

The results of the stepwise selection method are presented in Table 2. The model with the highest percentage of correct classification (78.3 per cent) was generated by using the first half of the observations as analysis subsample.

Table 2 — Stepwise Prediction Results

Analysis Subsample	Classification Accuracy
First Half	78.3%
Second Half	73.1%
All Even	72.9%
All Odd	76.0%
Random	68.7%

Using the first half analysis subsamples, several sets of discriminant functions were computed based on variables employed in other studies, judgment of personnel involved in the M-99 Program, household and attitudinal variables, and farm business variables. The predictive accuracy of these various iterations are given in Table 3. The highest percentage of correct classification (76.7 per cent) was generated by the function incorporating household and attitudinal variables.

Table 3 — Prediction Results of the Direct Method

Method of Variable Selection	Classification Accuracy
Based on Other Studies	71.7%
Judgment of Personnel Involved in the M-99 Program	68.6%

Table 3 (Continued)

Method of Variable Selection	Classification Accuracy
Household and Attitudinal Variables	76.7%
Farm Business Variables	70.8%

A comparison of the results in Tables 2 and 3 indicates that the stepwise method is superior in generating discriminant functions that have a higher predictive accuracy. As the last step in searching for the final model, the function that had the highest predictive accuracy was further refined through logarithmic transformation and standardization of outliers to ± 2 SD[4]. The models improved from 77.5 per cent for the log transformation to 79.5 percent when the variables were transformed and standardized.

The Final Model

The MDA model that performed best among those tested incorporated seven variables: dependency ratio, attitude towards saving, total income, current assets, production cost of palay, total palay production, and number of technicians' visits.

The variables' means and the resulting F statistics are presented in Table 4. As shown by the F-ratios, significant differences were observed with respect to two variables. Attitude towards saving and total income are significant at the .01 level indicating high significant differences in these variables between groups. The rest of the variables do not show any significant difference between groups. On a strictly univariate basis, the order of importance of the variables is indicated by the F-ratios. Thus, we can conclude that attitude towards saving is the most important variable.

---

[4] A common log transformation was applied to all the variables to improve normality and reduce the heteroscedasticity of the distributions. On the other hand, extreme values were standardized because they can influence the results of the analysis.

Table 4 — Variable Mean Values and Test of Significance, 120 Borrowers in Nueva Ecija, 1977

Variable	Low Group Mean n = 75	High Group Mean n = 45	F-ratio
Dependency Ratio	106.8	132.6	2.6770
Attitude Towards Saving	1.5	2.6	11.7210***
Total Income	1606.0	3629.0	7.4055***
Current Assets	2174.0	2012.0	.0053
Production Cost of Palay	2403.0	3155.0	3.7346
Total Palay Production	50.0	54.0	.2586
Number of Technician's Visits	6.0	4.0	3.9456

*** Significant at the .01 level.

Table 5 — Discriminant Weights and Relative Contribution of the Variables, 120 Borrowers in Nueva Ecija, 1977

Variable	Unstandardized	Standardized	Ranking
Dependency Ratio	.58468	.30089	5
Attitude Towards Saving	2.36928	.40875	1
Total Income	.45982	.32568	4
Current Assets	− .69207	− .36780	3
Production Cost of Palay	.13077	.37223	2
Total Palay Production	− .76322	− .21398	7
Number of Technicians Visits	− .44057	− .25700	6
Constant	− 3.55335		

The standardized coefficients shown in Table 5 enable us to evaluate each variable's contribution on a relative basis. Attitude towards saving, production cost of palay, current assets and total income are the four highest ranked variables. Together, they contributed 66 per cent of the discriminating power while the rest were accounted for by the remaining variables.

As to the signs of the coefficients, five of the seven variables exhibited relationships with repayment contrary to expectations. The dependency ratio was hypothesized to have a negative effect on repayment. The positive sign of the coefficient indicates that the more dependents in the family, the higher is the probability of repayment. Although the dependents were assumed not to have contributed to non-farm income, their labor could have been used in the production of other crops and livestock thereby increasing farm income. The cash receipts from other crops and livestock amounted to ₱1,344 with a cash expense of ₱310, or a cash income of ₱1,034. This income could be used to partly service outstanding obligations.

Attitude towards saving should have a negative sign owing to the scaling system used where the lower magnitude numbers signify favorable attitude. The positive sign indicates that farmers have reasons for saving other than to repay the M-99 loans. Their reasons to save show that personal and social needs have higher priorities.

The negative sign for total palay production can be attributed to a number of factors, the most important of which is the farmers' low productivity which led to a yield of only 42 cavans per hectare. This problem was compounded by the high cost of production, averaging ₱1,373 per hectare.

The negative sign of current assets indicate that below a given level or threshold limit, an increase in its value does not contribute positively to repayment. Nueva Ecija borrowers had current assets amounting to only ₱3,389 on the average. The same line of reasoning could be applied to the negative sign for the number of technicians' visits. For a one-year period, the farmers were visited 16 times or slightly more than once a month. The low frequency of visits was not very helpful to the farmers considering the biological nature of production where technical problems crop up occasionally.

Based on the results of the analysis of variance given in Table 6,

Table 6 — Analysis of Variance for the Discriminant Function in Nueva Ecija

Source	SS	DF	MS	F
Discriminant	36.08655	7	5.15521	509.91196
Residual Error	1.13273	112	.01011	

$F_{.99}$ (7,112) = 2.79 < 509.91196

it is concluded that the model can discriminate between the low and high repayment borrowers. Although the F-value is significant at the .01 level, further evaluation of the model will be made since the computed measure of association between the model and the groups as given by the canonical correlation was only .551. The square of this figure is .304 which means that the model can explain only 30.4 per cent of the variation in the seven variables describing the borrowers.

Evaluation of the Predictive Accuracy of the Model

In evaluating the predictive accuracy of the models, the basic format for presenting the results can be illustrated in terms of a confusion matrix or classification chart. Following Altman (1968), the chart is set up as follows:

	Predicted Group Membership	
Actual Group Membership	Low	High
Low	H	M_1
High	M_2	H

The Hs stand for correct classifications (hits) and the Ms stand for misclassifications (Misses). M_1 represents a Type I error (low repayment borrower classified as high) while M_2 is a Type II error (high repayment borrower classified as low). The sum of the figures along the diagonal equals the total correct "hits" and when divided into the total number of borrowers classified, gives the per cent of observations correctly classified. This percentage is analogous to the coefficient of determination (R^2) in regression analysis.

Classification Performance. The model was accurate in classifying correctly 79 per cent of the analysis subsample consisting of 120 borrowers (Table 7). This result is higher than the results obtained

Table 7 — Prediction Results of Discriminant Function, 120 Borrowers in Nueva Ecija, 1977

Actual Group Membership	No. of Cases	Predicted Group Membership Low		Predicted Group Membership High	
Low	75	65	(87%)	10	(13%)
High	45	15	(33%)	30	(67%)

Per cent of cases correctly classified : 79 %

by Cigaral (1977) and Matienzo (1978). The model developed by Cigaral correctly classified 67 per cent of the borrowers while the classification accuracy of Matienzo's models ranged from 59 to 70 per cent. The performance of the model in classifying borrowers in Nueva Ecija is quite encouraging. Type I error was 13 per cent while Type II error was higher at 33 per cent. Since a lower percentage of low repayment borrowers will be classified as high, the incidence of potential delinquent borrowers will be minimized. On the other hand, interest income from high repayment borrowers classified as low will be foregone.

Validation Results. The results of the first validation procedure suggested by Frank, et al. (1965) are listed in Table 8. As a whole, the test results reject the hypothesis that there is no difference between the groups. Therefore we can conclude that the model possesses dicriminant power and can be applied to the population in general. Furthermore, any search bias does not appear significant.

The second validation procedure compares the classification accuracy of the computer and a "naive" system. For this purpose, a subset of 20 farmers divided equally into low and high repayment borrowers was used. The comparative results of the two classification procedures are presented in Table 9. On the average, the computer correctly classified 90 per cent of the respondents compared to 50 per cent for the "naive" procedure. An appropriate test to determine significant differences in two procedures is the chi^2 statistic. However, this test was not conducted because more than 20 per cent of the theoretical frequencies were less than 5.

Based on the analysis of variance, classification performance

Table 8 — Accuracy of Classifying a Secondary Sample in Nueva Ecija

Replication	Per Cent Correct Classifications	t — values
1	72.9	3.5146***
2	55.9	.9109
3	62.7	1.9524*
4	66.1	2.4731**
5	66.1	2.4731**

* Significant at the .10 level.
** Significant at the .02 level.
*** Significant at the .001 level.

Table 9 — Comparative Classification Accuracy of the Computer and a "Naive" System for Nueva Ecija

Replication	Computer (%)	"Naive" System (%)
First Half	80	50
Second Half	100	50
All Odd	90	60
All Even	90	40
Random	90	50
Average	90	50

and validation results, the derived discriminant function for the Nueva Ecija borrowers is potentially useful in classifying M-99 loan applicants into low and high repayment borrowers. However, a number of qualifications must be considered. First, reducing the number of variables decreased the classification accuracy from 82 per cent when all 23 variables were included in the analysis to 78 per cent when only seven variables were considered. The four percentage decrease in accuracy, however, may not be that "costly" considering that the number of variables decreased by 70 per cent.

Second, there is evidence that a quadratic classification procedure may result in a higher classification accuracy as shown by the significant difference in the dispersion matrices. And third, there is a substantial overlapping of the groups as evidenced by the insignificant differences in the group means.

4. Application of the Model

The purpose of this last section is to translate the results of the discriminant analysis into guidelines that can be applied directly by financial institutions to classify borrowers into potentially low and high repayment borrowers. The topics to be discussed include the ambiguity zone or gray area, the screening process and an appraisal of the variable profiles.

"Ambiguity Zone" or "Gray Area"

The ambiguity zone can be defined as the range of Y scores where misclassifications of borrowers occur. The ambiguity zones for the province are as follows:

low repayment borrowers	
lower limit	-2.059
upper limit	2.150
high repayment borrowers	

Borrowers with Y scores below the lower limit can be concluded to fall into the low repayment sector while those above the upper limit are all high repayment borrowers. Within the ambiguity zone, the classification of a new applicant would be subject to uncertainty.

The process starts by identifying sample observations which fall within the overlapping range. Then the range of Y values that results in the minimum number of misclassifications is found. Table 10 shows the indicated Y values and the number of misclassifications.

The Screening Process

Two alternative screening methods are proposed. The first is a direct application of the discriminant analysis using the discriminant scores while the second is based on the rankings or relative contributions of the variables.

Table 10 — Number of Misclassifications Using Various Ranges of the Y Scores

Range of Y	Number Misclassified	Critical Value of Y*
–2.059 – –.654	2	
– .654 – .163	4	
.163 .292	2	
.292 – .380	1	.336
.380 – .680	5	
.680 – 2.150	4	

*The critical value of Y is the midpoint of the interval where the minimum number of misclassifications is observed.

Regarding the first method, the discriminant score of the applicant to be classified will be computed by substituting the values of his characteristics in the discriminant function. The resulting discriminant score will now be compared with the critical value as shown in Table 10. Thus, if the discriminant score of the applicant being classified is greater than .336, he is classified as a high repayment borrower. If his score is less than .336, he is classified as a low repayment borrower.

In the second method, the initial step is to compute the percentage contribution of the variables based on the size of the standardized discriminant coefficients shown in Table 5. For purposes of illustration, the percentage contribution of the variables are computed and presented in Table 11. The greater the size of the coefficient, with the signs disregarded, the greater is the percentage contribution. For example, the coefficient of attitude towards saving is the largest at .40875. Its percentage contribution is 18 per cent. Since the total palay production has the smallest coefficient of only .25700, the percentage contribution of this variable is the lowest at 10 per cent.

With the percentage contributions as weights, the perfect score for an applicant will be 100 points. To arrive at the score for an applicant, the raw values of his variables are compared with the mean values of the variables for the province and the corresponding points given, depending on whether the raw value is equal to, greater or

Table 11 — Percentage Contribution of the Variables for Nueva Ecija Borrowers

Variable	Standardized Coefficient*	Percentage Contribution
Attitude Towards Saving	.40875	18
Production Cost of Palay	.37223	17
Current Assets	.36780	16
Total Income	.32568	15
Dependency Ratio	.30089	13
Number of Technicians' Visits	.25700	11
Total Palay Production	.21398	10
Total	2.24633	100

*The negative signs have been omitted.

less than the mean value. For example, an applicant has current assets of ₱5,000. This value is then compared with ₱3,389 which is the mean value for all Nueva Ecija borrowers. Since the raw value is greater than the mean value, the applicant is given 16 points for current assets. On the other hand, if the raw value is less than the mean value, lesser points ranging from 0 to 15 will be given to current assets. The same procedure for assigning points is followed for the other variables.

A raw value equal or greater than the mean value is deemed desirable and, therefore, the applicant should be given the full points. The problem of assigning lesser points is encountered if the raw value is lesser than the mean value. Two alternative methods of assigning lower points were tried. First, if the raw value of a variable was lower than the mean value, the applicant was given zero points for that particular variable. Second, one-half of the full points was given if the raw value was between the mean value of the low or high repayment borrowers, whichever was lower, and the mean value of all farmers for the province. The predictive ability of the two methods were compared using a subsample of 10 low and 10 high

Table 12 — Determination of the Cutoff Score for Nueva Ecija Based on a Subsample of 20 Borrowers

Points	Interest Income [b]	Principal Plus Interest [c]	Profit (Loss) [d]
20	1080	8910	(7830)
30	810	5940	(5130)
40[a]	405	4455	(4050)
50	135	4455	(4320)
60	135	4455	(4320)

a/ Cutoff score that minimizes losses.

b/ Computed by multiplying total number of high repayment borrowers correctly classified by ₱135 (10% of ₱1350).

c/ Computed by multiplying total number of low repayment borrowers classified as high repayment by ₱1485 (principal of ₱135).

d/ The difference between interest income and principal plus interest.

repayment borrowers in each province. Since the results were comparable, the first method of giving 0 points was chosen because of its simplicity.

The cutoff score to separate the low from the high repayment borrowers was determined by comparing the interest income from high repayment borrowers if classified correctly, and the loss of the principal plus interest if a low repayment borrower was incorrectly classified as high repayment. For this purpose, the same subsample of 20 borrowers mentioned was utilized.

Table 12 gives the profit or loss at different levels of cutoff scores. The loan obtained by each borrower is assumed to be ₱1,350 with an interest of 10 per cent per annum with no service charges. Cutoff scores below 20 and greater than 60 points were omitted to avoid classification results of zero for either high or low repayment borrowers. The cutoff score that minimizes losses is 40 points.

A comparison of the predictive accuracy of the scoring system using discriminant scores and percentage contribution of the variables indicated that the former is a more reliable method. Based on the values of the same 20 subsamples, the percentage classification accuracy of the two methods is shown as follows:

Classification Using Discriminant Scores	90%
Classification Based on Percentage Contribution	50%
Type I Error	
First Method	20%
Second Method	30%
Type II Error	
First Method	0%
Second Method	70%

While the scoring system based on the percentage contribution is both simple and practical, it should be applied with some reservations. For one, the classification accuracy is low and furthermore, the incidence of both types of errors is generally higher.

An Appraisal of the Variable Profiles

Seven out of 15 variables appear or can be derived from the M-99 loan application form. An additional five variables can be added without difficulty while three variables must be substituted with surrogate variables. The listing is as follows:

Variables in the Application Form	Variables to be Added	Variables to be Substituted
1. Number of years in school	1. Total palay production	1. Attitude towards credit
2. Size of farm family	2. Value of other crops sold	2. Attitude towards saving
3. Dependency ratio	3. Production cost of palay	3. Attitude towards repayment

Variables in the Application Form	Variables to be Added
4. Total income	4. Production cost of other crops
5. Fixed assets	5. Number of technicians' visits
6. Current assets	
7. Total liabilities	

If only a selected number of variables will be used to screen the applicants, the predictive accuracy of the model will decline considerably as shown in Table 13. Any attempt therefore, to decrease the number of variables must consider the effects on the classification accuracy of the models.

Table 13 — Decline in the Predictive Accuracy of the Model by Using Selected Variables

Number of Variables	Classification Accuracy
All Variables Included in the Models	79%
Variables to be Added in the Application Form	65%
Variables in the Application Form	35%

While it is desirable to include as many variables as possible, some variables are quite difficult to measure directly. This is especially true of the attitudinal variables. Instead of a lengthy interview with the applicants to arrive at the ratings for their attitudes on credit, saving, and repayment, it is more practical to look for surrogate variables that will approximate the needed ratings.

Based on the results of the University of the Philippines Business Research Foundation (UPBRF, 1979) study, the rating for the savings attitude can be indicated by either one or a combination of the following variables: 1) number of organizations the applicant is a member of, 2) number of radio programs the applicant listens to daily, 3) number of household members who received medical attention the previous year and 4) the number of related families residing in the barrio. For the first two variables, the greater the frequency, the more favorable the savings attitude. The converse is true for the last two variables. Insofar as the credit attitude is concerned, only one surrogate variable was identified. This is the number of radio programs the applicant listens to daily.

As to the repayment attitude, there are three suggested variables: 1) number of organizations the applicant is a member of, 2) distance of applicant's residence from the Samahang Nayon, and 3) average age of household members. Variables 1 and 3 would have a favorable effect on repayment with an increase in the number and age, respectively.

It should be noted that some variables can be used to measure indirectly different attitudes. For example, the number of organizations the applicant is a member of can apply to both savings and repayment attitudes. This is not surprising because these attitudinal variables are correlated with each other.

The coefficients calculated in this study can be postulated to be quite stable over a period of time. Nevertheless, the stability of the coefficients should be scrutinized periodically. In studies made by Deakin (1972) and Sinkey (1975), it was noted that changes in the relative importance of the variables occur through the years. It is also possible that some variables over time may only insignificantly contribute to the discriminant power of the function. A deterioration in the classification accuracy of the model indicates that the coefficients are unstable and new computations are needed, possibly with the inclusion of new variables.

Management must, therefore, set up a formal system to monitor the stability of the coefficients. A flow chart for this purpose is depicted in Figure 1. At the outset, it should be pointed out that the initial cost of generating the final model to screen applicants would be high. However, as long as the model retains its predictive accuracy, the cost of using the model will be minimal. When the accuracy of the model deteriorates, a new set of computations must be made

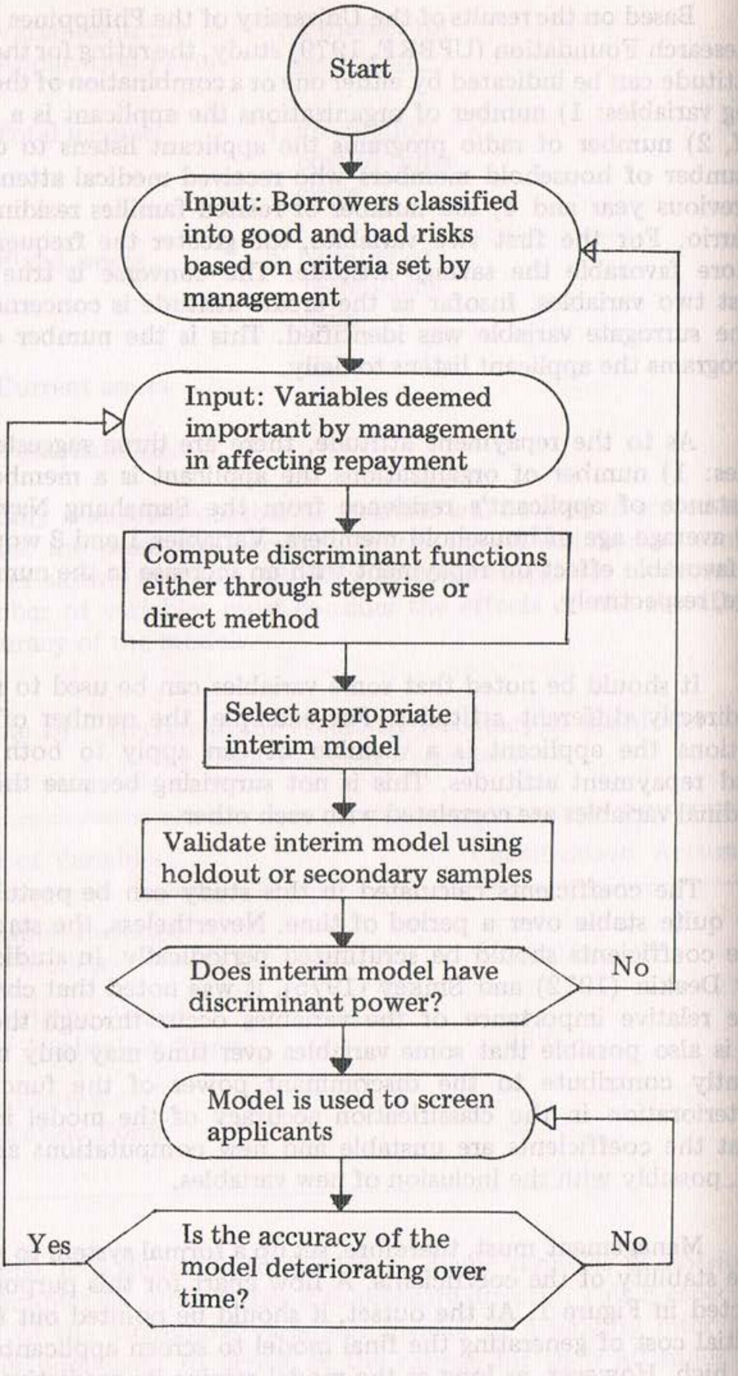

Figure 1 — Flow Chart for Monitoring Stability of the Coefficients

entailing processing time and cost that approximate the outlay in the initial set of calculations.

For the model that is used to screen M-99 borrowers, the stability of the coefficients must be examined at least after two phases, or about one year. The accuracy of the model should not be judged after only one cropping season because of uncertainties in the production process.

5. Conclusions

The overall results of the study suggest with a certain degree of confidence that it is possible to discriminate between the low and high repayment groups. However, the discriminant function computed for Nueva Ecija is not recommended for application in other provinces unless the profiles of farmers outside of the study area are similar to the characteristics of the farmers studied. Evidences in the literature indicate that different provinces may have unique discriminant functions.

Based on the results of the study, a number of policies could be adopted on the program level. These policies, already mentioned in the past, were confirmed in this study and need only further elaboration.

1. Separate credit packages should be developed depending on the income levels of the farmers. For instance, those with high incomes can be graduated from the M-99 Program and charged the prevailing interest rates on ordinary loans. Farmers with low incomes should not only continue to avail of the subsidized rate but should also receive the full complement of support services like assistance in marketing their products and closer technical supervision. For the marginal farmers, interest rates should be waived, or they can be charged only a token rate. As a rough rule of thumb, farmers with total incomes of over ₱6,000 are considered as having high incomes. Low income farmers would be those having total incomes of ₱4,000 to ₱6,000 while the marginal farmers would have incomes lower than ₱4,000. The income levels used to classify the farmers were based on the values obtained in the study and should be modified accordingly as income levels change.

2. The credit education campaign being waged at present should aim at improving the farmers' attitude towards saving. Based on the UPBRF study (1979), this can be done by encouraging the farmers to be members of various organizations and to listen more frequently to radio programs.

3. The number of farmers a technician will supervise should be reduced to increase the frequency of visits. On the average, two visits by the technician per farmer a month appear ideal.

At the institutional level, the credit scoring system recommended by this study is not meant to replace human skill and judgment in evaluating loan applicants. Rather, the use of discriminant analysis should be construed as an input to the final goal of developing an early-warning system to detect future problem borrowers.

The advantages offered by the discriminant model are:

a. The processing time for loans can be shortened at lower cost. This is particularly relevant to the M-99 Program where there are many loans with low income per loan. Those with high discriminant scores can have their loans approved immediately.

b. The financial institutions can identify beforehand potential problem borrowers and can direct the efforts of technicians toward these types of borrowers.

c. It serves as a means of evaluating the performance of loan examiners and production technicians.

d. It enables the financial institution to monitor the progress of the farmer from a bad risk to a good risk status or vice versa over time.

REFERENCES

Altman, Edward I. (September 1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance*, pp. 589-609.

Altman, Edward I., Margaine, Schlosser, Michel Micel, and Vernimmen, Pierre (March 1974) "Financial and Statistical Analysis for Commercial Loan Evaluation: A French Experience," *Journal of Financial and Quantitative Analysis*, pp. 195-214.

Avery, Robert B., and Eisenbeis, Robert A. (1972), *Discriminant Analysis and Classification Procedures, Theory and Applications*, Lexington, Massachusetts: D.C. Heath and Company.

Cigaral, Domingo E. (1977), "Short-term Agricultural Credit Repayment in Central Luzon, 1974," M.S. Thesis, University of the Philippines.

Deakin, Edward B. (Spring, 1972), "A Discriminant Analysis of the Predictors of Business Failure," *Journal of Accounting Research*, pp. 167-179.

Eisenbeis, Robert A. (June 1977), "Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics," *Journal of Finance*, pp. 875-900.

Frank, Ronald E., Massy, William F., and Morrison, Donald G. (August 1965). "Bias in Multiple Discriminant Analysis," *Journal of Marketing Research*, pp. 250-258.

Massy, William F. (1962), "Statistical Analysis of Relations Between Variables," in Frank, Ronald E., Kuehn, Alfred A., and Massy, William F., *Quantitative Techniques in Marketing Analysis*, Homewood, Ill: Richard D. Irwin, Inc., pp. 95-100.

Matienzo, Rodolfo M. (1978), "Repayment and Group Lending in the Province of Camarines Sur, Philippines," Ph.D. Thesis, The Ohio State University.

Morrison, Donald G. (May 1969), "On the Interpretation of

Discriminant Analysis," *Journal of Marketing Research*, pp. 156-163.

Pinches, George E., and Mingo, Kent A. (March, 1973), "A Multivariate Analysis of Industrial Bond Ratings," *Journal of Finance*, pp. 1-18.

Presidential Committee on Agricultural Credit (January 26, 1977), "Financing Agricultural Development: The Action Program," p. 13.

Robertson, Thomas S., and Kennedy, James N. (February 1968), "Prediction of Consumer Innovators: Application of Multiple Discriminant Analysis," *Journal of Marketing Research*, pp. 64-69.

Sheth, Jagdish N. (January 1971), "The Multivariate Revolution in Marketing Research," *Journal of Marketing Research*, pp. 13-19.

Sinkey, Joseph, Jr. F. (March 1975), "A Multivariate Statistical Analysis of the Characteristics of Problem Banks," *Journal of Finance*, pp. 21-36.

University of the Philippines Business Research Foundation (April 1979), "Financing Integrated Development in the Rural Community: Focus on Masagana 99," Report submitted to the Technical Board for Agricultural Credit.