Methods in Credit Scoring: A Review with Applications to the Canadian Farm Credit Corporation

Calcum Greig Turvey and Reginald Brown

Proceedings of
Regional Research Committee NC-161

FINANCING AGRICULTURE IN A CHANGING ENVIRONMENT: MACRO, MARKET, POLICY AND MANAGEMENT ISSUES

Kansas City, Missouri
September 24-25
1990

Department of Agricultural Economics and Rural Sociology
The Pennsylvania State University
University Park, Pennsylvania 16802
May 1991

Methods in Credit Scoring: A Review with Applications to the Canadian Farm Credit Corporation

by

Calum Greig Turvey and Reginald Brown*

Parametric approaches to evaluating the credit worthiness of agricultural loans has, in recent years, become an important topical issue. This is primarily due to the large amount of farm failures and loan defaults among borrowers and, in the United States, bank failures. Models used by lenders to assess credit worthiness are called credit scoring models. These use statistical analyses of economic and qualitative variables to objectively rank loan applications in terms of their probability of default, to price loans in terms of default risk, and in some cases to establish loan loss provisions for accounting and tax purposes.

While a host of parametric approaches to credit scoring are available, (see Chhikara for a non-parametric approach) the four most common are the linear probability model (LPM), discriminant analysis (DA), LOGIT regression and PROBIT regression. The appropriateness of which method should be used is not clear from the current body of literature. For example, many academics would recommend one method over the other on the basis of statistical and econometric properties alone (e.g. Lo; McFadden; Madalla; and others) while others are willing to weigh the statistical properties in light of more pragmatic issues such as ease of use (Fischer and Moore; Collins and Green).

The purpose of this paper is to provide a review of the four alternative parametric approaches to credit scoring. The intent of this research is to provide academics and lenders with a base from which credit scoring models can be evaluated both in a qualitative sense and empirical sense. Consequently, the paper does not conclude with a solid recommendation of practice, because the results cannot, in light of other studies, be deemed conclusive. Rather, the conclusions are drawn from a framework of analysis which provides a comparative review of alternatives, which can be used as a first step in model selection and development.

The paper is outlined as follows: in the next section the LPM, DA, LOGIT and PROBIT models are introduced, with special emphasis on their statistical properties and interrelationships, and applications. The empirical section follows using 1981 loan applications data from Canada's Farm Credit Corporation and loan status as of March 31, 1990. The paper then concludes with a discussion of the empirical results.

TECHNIQUES IN CREDIT SCORING

This section presents the econometric models for credit scoring. Namely, the linear probability model, discriminate analysis, LOGIT regression and PROBIT regression are presented in that order.

---

* Cal Turvey is an Assistant Professor and Reg Brown is a former graduate student in the Department of Agricultural Economics and Business, University of Guelph.
Paper presented to the NC-161 Finance Conference, Kansas City, Missouri. September 24-25, 1990.

Linear Probability Model

The linear probability model uses ordinary least squares or weighted least squares to regress substantive quantitative and qualitative independent variables against a dichotomous dependent variable which takes a value 1 if the loan is in default and zero otherwise. The model is specified as

$$(1) \quad Z_i = \sum_{j=1}^{n} B_j X_{ij} + e_i$$

where Z_i is the (0,1) dependent variable, B_j is the coefficient on the j^{th} variable, X_{ij} is the i^{th} observation on variable j, and e_i is the residual error term, with $E[e_i] = 0$. The expected value of (1), i.e. $E[Z_i \; X_{ij}] = \sum_{j=1}^{n} B_j X_{ij}$ is interpreted as a probability, ρ_i, that a particular loan will go into default. The probability of the loan being current is $1 - \sum_{j=1}^{n} B_j X_{ij}$.

Estimates of the LPM are not efficient since the variance of e_i is heteroscedastic, i.e.

$$\sigma^2(Z_i) = \rho_i(1-\rho_i) = E[Z_i] \; (1-E[Z_i]).$$

Correction for heteroscedasity may be obtained using weighted least squares (WLS) with weights defined by the predicted OLS values for Z_i, i.e.

$$W_i = [\hat{Z}_i \; (1-\hat{Z}_i)]^{1/2} .$$

However, since there is no guarantee that $\hat{Z}_i$ will lie in the unit interval (0,1) these weights may not be applicable to all observations. While it is possible to redefine the predicted probabilities outside of the unit interval to be near 0 or 1, both Johnson, and Pyndick and Rubinfeld urge against it, with Pyndick and Rubinfeld recommending OLS and Johnson recommending neither.

Other problems related to the LPM are that the estimated standard errors are not consistent thereby invalidating R^2 and statistical tests on the coefficients, and the fitted relationship are sensitive to bunching in the explanatory variables. Also, even if the in-sample conditional probabilities fall within the unit interval there is no guarantee that out-of-sample predictions would also do so (Judge <u>et al</u>). Thus the predictive ability of the LPM is tenuous at least. Finally, a major criticism of the LPM is that there exists alternative procedures which are efficient and consistent (e.g. LOGIT and PROBIT), while providing a more tractable prediction response.

Discriminant Analysis

An alternative linear approach to credit scoring is the widely used discriminant analysis. Discriminant analysis differs from the LPM in one important way. Whereas the LPM estimates are based on the distribution of Z_i conditional on the X's, DA uses X conditional on Z. Thus, the objective of DA is not to provide the user with a probability of loan default per se, but rather a numerical range over which classes are defined.

The objective of DA is therefore to find a linear function.

$$(4) \quad Z_i = \sum_{j=1}^{n} B_j X_i$$

which discriminates between the two loan classifications (0 or 1). This requires analysis of variance which maximizes the between group variance while minimizing the within group variance. These dual objectives will maximize the discriminating power. Thus, the objective is to choose the B_j which maximizes the ratio of the between group variance to within group variance (see Madalla page 17), i.e.

$$(5) \quad \underset{B}{\text{Max}}\ \phi = \frac{B'(\bar{X}_o - \bar{X}_1)^2}{B'S\ B}$$

where $\bar{X}_o$ and $\bar{X}_1$ represent the vectors of expected values for the explanatory variables for each of the two subgroups, B is a vector of coefficients, and S is the covariance matrix assumed to be equal for each subgroup.

Maximization of (5) gives

$$(6) \quad B = S^{-1}(\bar{X}_o - \bar{X}_2)\ .$$

Each of the two classifications can thus be defined by

$$\bar{Z}_o = \hat{B}'\bar{X}_o\ , \quad \text{and}$$

$$\bar{Z}_1 = \hat{B}'\bar{X}_1\ .$$

so that any observation can be discriminated if its value Z^* is closer to one of the classification values ($\bar{Z}_o$ or $\bar{Z}_1$) than the other.

The appropriate test of the discriminant function is against the null hypothesis that there are no significant differences in the means of the two groups. This is accomplished using Hotellings T^2 test which is distributed as an F test,

$$(7) \quad T^2 = \frac{N_o N_1\ (N_o + N_1 - K - 1)}{(N_o + N_1)(N_o + N_1 - 2)\ K}\ D^2\ ,$$

where N_0 and N_1 are the number of observations with dichotomous priors 0 and 1 respectively, k is the number of explanatory variables, and D^2 is the Mahalanobis generalized distance ($D^2 = (\bar{Z}_0 - \bar{Z}_1)^2$)). The T^2 statistic is evaluated with k, and (N_0+N_1-k-1) degrees of freedom.

The biggest problem with discriminant analysis is in the assumption that the explanatory variables in the two groups come from normal populations. If they do not come from normal populations, and or, the variance-covariance matrices are not equal then the estimator is not consistent (Madalla; Lo). Collins and Green state that for credit scoring models it is unlikely that the distribution of financial ratios is normal, and Press and Wilson indicate that the introduction of dummy variables as explanatory variables automatically violates the assumption of normality. Moreover, Collins and Green argue that when dichotomous relationships for bankruptcy models are defined it would be difficult in most cases to successfully argue equal variances.

An alternative approach to credit scoring, which is applicable for a wider range of distributions (Lo) is logistic regression analysis, or LOGIT. This is discussed in the next section.

The LOGIT Probability Model

Judging from the substantial literature which uses logistic regression in credit scoring it would appear to have merits, statistical and otherwise, which makes it more favourable than either LPM or DA.

The dichotomous logit model, assumes a logistic cumulative distribution function of the form

$$(8) \qquad F(Z_i) = \frac{1}{1+\exp(-Z_i)}$$

which is the probability of being non current and

$$(9) \qquad 1-F(Z_i) = \frac{\exp(-Z_i)}{1+\exp(-Z_i)}$$

is the probability of being current

where $Z_i = \sum_{j=1}^{n} B_j X_{ij} + \mu_i$.

In practice Z itself is not observable so that $\sum_{j=1}^{n} B_j X_{ij} + \mu_i$ is an instrument of Z.

The likelihood function is the non linear estimator to be used and it takes the form (Madalla)

$$(10) \quad L = \prod_{i=1}^{N} (1-F(-Z_i))^{Y_i} F(-Z_i)^{1-Y_i}$$

where Y_i takes on the value of the dichotomous (0,1) variable, and N is the number of observations.

The log of the likelihood functions is

$$\ln L = \sum_{i=1}^{N} (Y_i \ln(1-FL(-Z_i)) + (1-Y_i) \ln F(-Z_i))$$

$$= \sum_{i=1}^{N} (Y_i \ln \left[\frac{\exp(-Z_j)}{1+\exp(-Z_i)}\right] + (1-Y_i) \ln \left[\frac{1}{1+\exp(-Z_i)}\right]$$

The log of this likelihood function after substituting for $F(Z_i)$ and $(1-F(Z_i))$ is

$$(11) \quad \ln L = \sum_{i=1}^{n} (Y_i \ln(\exp(-Z_i) - \ln(1+\exp(-Z_i)),$$

and the LOGIT coefficients are obtained by simultaneous solution (through non linear iterative techniques) of

$$(12) \quad \frac{d\ln L}{dB_j} = \sum_{i=1}^{N} \left[Y_i X_{ji} + X_{ji} \frac{\exp(-Z_i)}{1+\exp(-Z_i)} \right], \; j = 1,n .$$

The solution of the $\hat{B}_j$ coefficients then permits solving that the probability of observation i being non current as

$$(13) \quad \rho_i = 1 / 1+\exp (-\sum_{j=1}^{n} \hat{B}_j X_{ij}) .$$

The LOGIT model is appealing on many accounts. First, having solved for B_j, the probability estimate is easily solved for. However, both LPM and DA also have this advantage. But in addition to simplicity in form and implementation, LOGIT is asymptotically efficient and consistent. Moreover, by its very nature probability estimates are guaranteed to fall within the unit interval. Hence, it does not have the problems of efficiency indicated with the LPM. Nor does the LOGIT model need the strict assumption of multivariate normality and equal covariance matrices as required by DA. It is interesting that if the assumption of normality is satisfied, DA is asymptotically more efficient than LOGIT, but if the assumption fails DA is not consistent, whereas this property is retained for LOGIT (Madalla; Lo; McFadden).

The PROBIT Probability Model

An alternative to LOGIT, is to assume a cumulative normal distribution rather than a cumulative logistic distribution. Both logistic and normal distributions intersect at 50% probability but differ at the tails.

The solution procedure for solving the PROBIT likelihood function is similar to that of LOGIT except in the definition of the density, but is substantially more complex (see Madalla, p.26). Having solved iteratively for the PROBIT coefficients, B_j, the value for $Z_i = \sum_{j=1}^{n} B_j X_{ij}$ is substituted into the cumulative normal density function to obtain (usually from a polynomial approximation of the normal density function) an estimate of the probability of being non-current, ρ_i.

Because the normal density resembles the logistic density the parameter estimates are usually quite close such that the LOGIT estimates can be multiplied by $3^{.5}/\pi$ to obtain comparable PROBIT estimates (Madalla, page 23). (Madalla also notes that Ameniya suggests using .625 rather than .551).[1]

Both PROBIT and LOGIT are asymptotically efficient and consistent so that standard log likelihood test statistics can be used. Both models provide probability estimates which are between 0 and 1.

DISCUSSION OF CREDIT SCORING MODELS

The above discussion emphasized primarily the econometric estimation and statistical properties of the alternative models. In many cases however LPM and DA perform quite well in large samples even when some properties are violated. For example, Press and Wilson show that when multivariate normality is violated, LOGIT performs only slightly better than DA. Similar findings are reported by Collins and Green, and Lo. In fact Collins and Green find that the prediction accuracies of LOGIT, LPM and DA are so close that they question whether or not the extra computational effort is worth it. One might extrapolate from this a similar query about PROBIT models.

In practice, criteria other than econometric considerations should be adopted when evaluating alternative credit scoring model. Fischer and Moore provide the following reasonable criteria; the credit scoring model should be able to contribute to the bank's loan classification system in screening loan applicants, diagnose credit weaknesses and price loans based on credit quality; it must be accurate enough to contribute to sound lending decisions and valid loan

---

[1] And a great many studies have appeared in the non agricultural literature. As well as those cited in the text see Aziz and Lawson; Altman (1986); Collins; Altman _et al_; Ohlson; and Scott.

classification; objective in its ability to price loans[2]; simple enough for loan officers to compute and interpret credit scores for screening applicants; and statistically valid. In addition Turvey and Brown urge that the model should take into account and discriminate across a heterogenous farm population.

However, credit scoring models in themselves will not be successful in assessing the success of a particular loan. Studies have shown that there is a great deal of subjectivity involved on the part of lenders which lead to variations in the amounts of loans awarded (Sonka, Dixon and Jones; Stover, Teas and Gardner). Nor can it be expected that credit scoring models be used in isolation of institutional restrictions, the bank loan portfolio, macroeconomic policy, and competition, if differential pricing according to credit risk is to be practiced (Barry and Calert; Gustafson). Moreover, credit scoring models are not necessarily derived from probabilistic statistical models. Some lenders develop a credit scoring function using subjective weights on key financial variables with the weights summing to 1. We have proprietary information on one major Canadian chartered bank which provides branch lenders with such a model and guidelines for use on a voluntary basis. Barry and Ellinger illustrate the use of these models based upon scoring models used by the St. Louis and Louisville Farm Credit banks. There is no evidence that objectively based credit scoring models perform any better or worse than subjectively based models.

Credit Scoring for Agricultural Lending

A substantial literature is appearing which applies credit scoring models to agricultural lending (Dunn and Frey; Lufburrow, Barry and Dixon; Mortensen, Watt and Listritz; Fischer and Moore; Turvey and Brown; Miller and LaDue (and citations therein)). Most of these studies use measures of liquidity, profitability, leverage, efficiency, and repayment ability as the explanatory variables, (Lufburrow et al; Barry and Ellinger; Miller and LaDue (see their Table 1, page 27); Turvey and Brown).

This research too uses the above measures to obtain estimates for the alternative credit scoring models. The data were obtained from actual 1981 FCC loan applications for which loans were made. The dichotomous dependent variable is 1 if the loan is noncurrent and 0 otherwise. These are based on the status of the loan as of March 31, 1990. Using this data, the following variables were defined; liquidity is measured by the current or liquidity ratio (CR); profitability by the return on assets (ROA); leverage is measured by 3 variables, 1) the debt to asset ratio (DA), 2) a dummy variable (DLOAN) which takes on a value of 1 if the loan is required for refinancing and zero otherwise, and 3) the loan to security ratio (LS); efficiency is measured by the gross ratio (GR); and repayment ability is measured by 1) the interest coverage ratio (IC) and 2) the ratio of off-farm income to cash income before interest payments (DOF). As well,

---

[2] Similar relationships can also be derived for comparison of LPM, DA, LOGIT and PROBIT. These are (using Ameniya's approximation of .625); $B_{LP} \approx .25B_L$ for coefficient; $B_{LP} \approx .25B_L + .5$ for intercept; $B_{LP} \approx .4B_P$ for coefficient; $B_{LP} \approx .4B_L + .5$ for intercept; $B_{DA} = \frac{B_{LP}(N_o+N_1-2)}{RSS}$ excluding intercept, where N_o, N_1 are proportions as 0's and 1, respectively, and RSS is the residual sum of squares. The subscripts LP, DA, L and P are linear probability, Discriminant Analysis, LOGIT and PROBIT, respectively.

binary dummy variables on province and farm type are introduced to capture covariance relationships. This procedure is reported in Turvey and Brown, and is intended to capture regional and farm type differences which may affect the probability of being non-current. The procedure recognizes that the FCC is a federal lending institution with a very heterogenous loan portfolio.

In total 2,798 loans are used in the estimating procedure, of which 1,746 are current (type = 0) and 1,052 non-current (type = 1)[3]. The summary statistics for the continuous variables are reported in Table 1.

Model Results

The results of the 4 credit scoring models are presented in Tables 2 through 5. Table 2 presents the estimated equations and coefficients values on the independent variables. General consistency was found for the signs on the parameters. The positive sign indicates that the probability of being non-current increases with the value of the variable while a negative sign indicates a decrease. The variables DA, LS, and DLOAN were all expected to have positive signs since these measure financial risk. The loan to security ratio (LS) was not found to be significant for the LPM but was for DA, LOGIT, and PROBIT. The coefficient on the remaining financial variables, ROA, CR, IC, and GR all had negative signs, and these were consistent across the alternative models. Thus, profitability, liquidity, efficiency, and repayment capacity do interact to reduce the probability of default. The interest coverage ratio, IC, was not significantly different from zero in any of the models.

The results support hypotheses that measures of leverage, liquidity, profitability, efficiency, and repayment capacity are important variables in the model.

Elasticities for the key financial variables (Table 3) indicate that financial leverage (DA) and efficiency (GR) have the greatest impact on the probability of default. For example with the LOGIT model a 1% increase in the leverage ratio increases the probability of default by 1.52% and a 1% increase in efficiency decreases the probability of default by .93%. The weights placed on these values are consistent with the studies by Lufburrow et al, Mortenson et al and Miller and LaDue. The interest coverage ratio, a measure of repayments ability has the lowest absolute elasticity value. All models consistently rank DA, GR, and ROA as the top 3 most important determinants of credit risk, and IC is consistently the least important. Inconsistent ranking appear relative to the loan to security ratio (LS) and current ratio (CR) with the former being ranked higher with LOGIT and PROBIT, and the latter being ranked higher with LPM and DA.

The interactions of farm type and province are presented in an analysis of covariance framework in Table 4. The absolute values of the interactions differ

---

[3] Boyes, Hoffman and Low point out that the population of borrowers is actually censored since only those that obtained loans are observed. This censoring may lead to biased estimates if lenders use other than objective measures to evaluate loans. This, selectivity bias, is troublesome since a purely objective measure of loan classification is virtually impossible. Moreover, it is not clear that credit scoring is intended to fully substitute for lenders' judgement.

across the models. The LPM has sign consistency with DA, and LOGIT has sign consistency with PROBIT. However, the inconsistencies are found only in the dairy enterprise and even then for only 1 province (Alberta). In general then these covariance relationships can be used to adjust the probabilities. For example, all other things being equal, the LPM model indicates that a dairy farm in Quebec (-.188) is less likely to default on a loan than a dairy farm in Saskatchewan (-051), and a cash crop farm in Alberta (.276) is more likely to default on a loan than a beef farmer in Ontario (-.066). In fact since the LPM is linear, it can be stated that, <u>ceterus paribus</u>, the Saskatchewan dairy farmer has a 13.7% greater probability of default than the Quebec dairy farmer, and the Alberta cash crop farmer has a 34.2% greater probability of being non-current than the Ontario beef farmer.

Of the 5 crop types and 9 province dummy variable, 7 were significantly different than zero. While not reported here, Turvey and Brown show that both the crop and regional dummy variables contribute significantly to the LOGIT regression results. (Due to problems of efficiency and consistency with the LPM and DA models, testing the LOGIT model is appropriate. The LOGIT results can, confidently, be applied to the PROBIT results.)

Prediction success tables for the four models are presented in Table 5. These tables were compiled using prior probabilities of 63% for current and 37% for non-current loans (see Hensler and Johnson). The results indicate that the greatest prediction accuracy of 71.3% was found for LOGIT, followed by DA (71.19%), PROBIT (68.9%) and LPM (68.2%). These results are consistent with other studies. In fact the LOGIT-DA results are remarkably consistent with the findings of Collins and Green and Press and Wilson, who found that DA performed only slightly less than LOGIT even when the assumption of multivariate normality is violated.

The main difference between the models is in the Type I and Type II errors. Type I errors arise from classifying problem borrowers as being acceptable and Type II errors arise from classifying acceptable borrowers as a problem (Miller and LaDue). These errors are evaluated relative to the percent predicted to actual in the prediction success tables. For example the DA model predicts correctly 82.5% of current and 52.28% of non-current, whereas LOGIT predicts correctly 92% of current but only 30.8% of non-current borrowers. These results imply that LOGIT has lower Type I error than DA, but higher Type II error. One might argue that the opportunity cost of a Type II error (in terms of revenue foregone) is less than the opportunity cost of Type I errors (in terms of loss of principle, interest income and additional administrative and legal costs). Hence a low Type I error is preferable to a low Type II error. This result may lead one to prefer DA over LOGIT, regardless of the obvious loss in some of the required statistical properties. However, whereas the relative overall prediction accuracy of the two models is similar to that of other studies, Collins and Green find lower Type I errors for LOGIT than DA. Thus nothing general can really be said on this aspect of model comparison.

CONCLUSIONS

The purpose of this paper was to review, and empirically estimate four alternative credit scoring models. These were the linear probability model, discriminant analysis, LOGIT regression and PROBIT regression.

The results indicate, as expected, that the coefficients of the LPM and DA, were more closely related to each other than LOGIT and PROBIT, and LOGIT and PROBIT were more closely related to each other from LPM and DA. The DA and LOGIT models showed the highest prediction accuracy when a .37 prior probability of being non current was used. The PROBIT and LPM models had slightly less predictive accuracy. In general the models were more likely to accurately predict a current loan as being current, than a non-current loan as being non-current. In fact Type I error (of misclassifying non-current loans) is quite high. The lowest Type I error occurred with DA, which would imply (putting statistical issues aside) that it is an appropriate estimator for classifying FCC loans. However, DA cannot in general be deemed superior since other research has shown higher Type I errors.

The high Type I error can be attributed to the fact that 10 years of volatile commodity markets, drought, and low farm incomes has passed since the loans were actually made. Given the history, the financial crisis in the 1980's the overall predictive accuracy of over 70% is quite good and is not out of line with other studies. Moreover, in personal conversations with FCC lenders, it was pointed out that between 10% and 20% of bad loans were due to personal misfortune such as health, injury, death, divorce and in-family legal conflicts. Thus, substantial Type I error can arise from uncertainties which no probability based model can predict.

Finally, statistical accuracy and predictive ability, may not be sufficient measures of model selection. Ease of use, and purpose should also be considered. PROBIT models, for example, are neither simple in form or use. If the lender is appealing to a probability based measure then DA may not be appropriate. These considerations should be given to the appropriate selection of a credit scoring model.

Table 1: Summary Statistics of Financial Variables for Current and Non-current Loans

Variable	Current		Non-current	
	Mean	Standard Deviation	Mean	Standard Deviation
DA	.481	.194	.602	.178
RGA	.115	.106	.110	.095
LS	.523	.224	.574	.227
LR	1.964	.226	1.379	1.996
IC	.741	.507	.95	.522
GR	.894	.156	.86	.188
DOF[a]	.134	.37	.16	.466
DLOAN[a]	589	-	453	-

[a] Number of observations, not mean.

Table 2: Alternative Parameter Estimates of Credit Scoring Models[a]

Variable	Linear Probability Model	Discriminant Analysis	LOGIT	PROBIT
Constant	.243*	-.275*	-1.386*	-.873*
DA	.81*	4.353*	4.331*	2.581*
ROA	-.407*	-2.187**	-2.225*	-1.285*
LS	.072	.385*	0.459*	.275*
CR	-.023*	-.122*	-.129*	-.077*
IC	-.028	-.151	-.186**	-.103**
GR	-.396*	-2.126*	-2.035*	-1.16*
DLOAN	.108*	.58	.585*	.342*
OFI	.005	.029*	.013	.021
Cash Crop	.043	.230	.189	.129
Dairy	-.191*	-1.025*	-1.234*	-.711*
Beef	-.097*	-.52*	-.605*	-.342
Hogs	.069	.369	.328	.207
Poultry	-.014	-.0746**	-.085	-.046
Newfoundland	-.287*	-1.541*	-2.295*	-1.365*
British Columbia	.191*	1.025*	0.981*	.575*
Alberta	.233*	1.252*	1.195*	.703*
Saskatchewan	.158*	.850*	0.85*	.488*
Manitoba	.14*	.751*	0.71*	.412*
Ontario	.031	.169	0.15	.075
Quebec	.003	.018	-.071	-.047
New Brunswick	.009	.051	-.065	-.031
Nova Scotia	.0514	.276	.230	.292
Likelihood Ratio	-	-	661.19	662.19
F	33.06	33.06	-	-
R^2	.20	-	.21	.21

* Indicates significance at 5% level

** Indicates significance at 10% level. These use t tests except for DA which is F-Test.

Table 3: Estimated Elasticities on Key Financial Variables for Alternative Credit Scoring Models[a]

Variable	Linear Probability Model	Discriminant[b] Analysis	LOGIT	PROBIT
DA	1.134	6.095	1.515	1.460
ROA	-.123	-.661	-.167	-.156
LS	.103	.554	.165	.160
CR	-.105	-.564	-.149	-.143
IC	-.061	-.328	-.101	-.091
GR	-.929	-4.993	-1.193	-1.100

[a] Elasticity at means.

[b] Since $B_{DA} = 5.375\ B_{LPM}$, these elasticities are obtained by multiplying the LPM elasticities by 5.375.

Table 4: Prediction Success Tables for Alternative Credit Scoring Models

	Predicted Current(1)	Predicted Arrears(2)	Observed Count	Observed Score
		Linear Probability Model		
Actual Current (1)	1653	93	1746	.63
Actual Arrears (2)	797	255	1052	.37
Percent Correctly Predicted	67.5	73.3	68.2	
Percent Predicted to Actual	67.5	73.3	68.2	
Predicted Share (%)	87.56	12.44		
		Discriminant Analysis		
Actual Current (1)	1442	304	1746	.63
Actual Arrears (2)	502	550	1052	.37
Percent Correctly Predicted	74.18	64.40	71.19	
Percent Predicted to Actual	82.59	50.28		
Predicted Share (%)	69.48	30.52		
		LOGIT Regression		
Actual Current (1)	1607	139	1746	.63
Actual Arrears (2)	728	324	1052	.37
Percent Correctly Predicted	68.8	70.0	71.3	
Percent Predicted to Actual	92.0	30.8	71.3	
Predicted Share (%)	83.45	16.55		
		PROBIT Regression		
Actual Current (1)	1617	129	1746	.63
Actual Arrears (2)	740	312	1052	.37
Percent Correctly Predicted	68.6	70.8	68.9	
Percent Predicted to Actual	92.61	29.7	68.9	
Predicted Share (%)	84.23	15.76		

Table 5: Farm Type and Province Covariance Matrix for Alternative Credit Scoring Models

Variable	Cash Crop	Dairy	Beef	Hogs	Poultry
			Linear Probability Model		
Newfoundland	-.244	-.478	-.384	-.218	-.301
British Columbia	.234	.000	.094	.26	.177
Alberta	.276	.042	.136	.302	.219
Saskatchewan	.201	-.033	.061	.227	.144
Manitoba	.183	-.051	.043	.209	.126
Ontario	.074	-.16	-.066	.100	.017
Quebec	.046	-.188	-.094	.072	-.011
New Brunswick	.009	-.182	-.088	.078	-.005
Nova Scotia	.094	-.14	-.046	.120	.037
			Discriminant Analysis		
Newfoundland	-1.311	-2.566	-2.061	-1.172	-1.616
British Columbia	1.255	0	.505	1.394	.950
Alberta	1.482	.227	.732	.883	1.177
Saskatchewan	1.08	-.175	.330	1.219	.775
Manitoba	.981	-.274	.231	1.120	.676
Ontario	.399	-.856	-.351	.538	.094
Quebec	.248	-1.007	-.502	.387	-.057
New Brunswick	.281	-.974	-.469	.420	-.024
Nova Scotia	.506	-.749	-.244	.645	.201
			LOGIT Regression		
Newfoundland	-2.11	-3.529	-2.899	-1.967	-2.380
British Columbia	1.171	-.253	.377	1.309	.896
Alberta	1.384	-.040	.589	1.522	1.109
Saskatchewan	1.039	-.385	.245	1.177	.765
Manitoba	.899	-.524	.105	1.038	.625
Ontario	.339	-1.084	-.455	.477	.065
Quebec	.260	-1.305	-.676	.257	-.156
New Brunswick	.124	-1.299	-.670	.262	-.115
Nova Scotia	.419	-1.004	-.374	.558	.146
			PROBIT Regression		
Newfoundland	-1.236	-2.076	-1.707	-1.158	-1.411
British Columbia	.704	-.136	.233	.782	.529
Alberta	.832	-.008	.361	.910	.657
Saskatchewan	.617	-.223	.146	.695	.445
Manitoba	.541	-.299	.070	.619	.366
Ontario	.204	-.636	-.267	.282	.029
Quebec	.082	-.758	-.389	.160	-.093
New Brunswick	.098	-.742	-.373	.176	-.077
Nova Scotia	.421	-.419	-.05	.499	.246

References

Altman, E.I., "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy", J. of Finance. 23(1968):589-609.

Altman, E.I., R. Haldeman and P. Narayanan, "Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations", J. of Banking and Finance. 1(1977):29-54.

Amemiya, T., "Qualitative Response Models: A Survey", J. of Econ. Lit. 19(1981):483-536.

Aziz, A. and G.H. Lawson, "Cash Flow Reporting and Financial Distress Models: Testing of Hypotheses", Financial Management. 18(1989):55-63.

Barry, P.J. and P.N. Ellinger, "Credit Scoring, Loan Pricing, and Farm Business Performance", Western J. of Agric. Econ. 14(1989):45-55.

Barry, P.J. and J.D. Culvert, :Loan Pricing and Customer Profitability Analysis by Agricultural Banks", Agricultural Finance Review. 43(1983):21-29.

Boyes, W.J., D.L. Hoffman, and S.A. Low, "An Econometric Analysis of the Bank Credit Scoring Problem", J. of Econometrics. 40(1989):3-14.

Chhikara, R., "The State of the Art in Credit Evaluation", American Journal of Agricultural Economics. 71(1989):1138-1144.

Collins, R.A., "An Empirical Comparison of Bankruptcy Prediction Models", Financial Management. 9(1980):52.57.

Collins, R.A. and R.D. Green, "Statistical Methods for Bankruptcy Forecasting", Journal of Economics and Business. 34(1982):349-354.

Dunn, D.J. and T.L. Frey, "Discriminant Analysis of Loans for Cash-Grain Farms", Agricultural Finance Review. 36(1976):60-66.

Fischer, M.L. and K. Moore, "An Improved Credit Scoring Function for the St. Paul Bank for Cooperatives", Journal of Agricultural Cooperation. 1(1986):11-21.

Gustafson, C.R., "Credit Evaluation: Monitoring the Financial Health of Agriculture", American J. Agricultural Economics. 71(1989):1145-1151.

Hensher, D.A. and L.W. Johnson, Applied Discrete Choice Modelling", John Wiley and Sons, New York, N.Y. 1981.

Johnson, J., Econometric Methods, McGraw-Hill Inc., New York. (1984)

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl and T. Lee, The Theory and Practice of Econometrics, John Wiley and Sons, Inc., Toronto. 1980.

Lo, A.W., "LOGIT versus Discriminant Analysis: A Specification Test and Applications to Corporate Bankruptcy", J. of Econometrics. 31(1986):151-178.

References continued......

Lufburrow, J., P.J. Barry and B.L. Dixon, "Credit Scoring for Farm Loan Pricing", Agricultural Finance Review. 44(1984):8-14.

Madalla, G.S., Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge. (1983).

McFadden, D. "A Comment on Discriminant Analysis versus LOGIT Analysis", Annals of Economics and Social Measurement. 5(1976):511-523.

Miller, L.H. and E.L. LaDue, "Credit Assessment Models for Farm Borrowers: A LOGIT Analysis", Agricultural Finance Review. 49(1989):22-36.

Mortenson, T., D.L. Watt, and F.L. Leistritz, "Predicting Probability of Loan Default" Agricultural Finance Review. 48(1988):60-67.

Ohlson, J., "Financial Ratios and the Probabilistic Prediction of Bankruptcy", J. of Account. Res. (Spring 1980):109-131.

Press, S.J. and S. Wilson, "Choosing Between Logistic Regression and Discriminant Analysis", Journal of the American Statistical Association. 73(1978):699-705.

Pyndick, R.S. and D.L. Rubinfeld, Econometric Models and Economic Forecasts, 2nd ed., McGraw Hill Book Co., N.Y., N.Y. 1981.

Scott, J., "The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models", J. of Bank. and Fin. (Sept. 1981):317-344.

Sonka, S.T., B.L. Dixon and B.L. Jones, "Impact of Farm Financial Structure on the Credit Reserve of the Farm Business", American J. Agricultural Econ.", 62(1980):565-570.

Stover, R.D., R.K. Teas, and R.J. Gardner, "Agricultural Lending Decisions: A Multivariate Analysis", American J. Agricultural Economics. 67(1985):513-520.

Turvey, C.G. and R. Brown, "Credit Scoring for a Federal Lending Institution: The Case of Canada's Farm Credit Corporation", Agricultural Finance Review. 50(1990):47-57.