Ghanaian Journal of Economics, Vol. 3, December 2015
©The Author(s)
Journal compilation ©2015 African Finance and Economics Consult

Delinquency and default risk modelling of microfinance in Ghana

*Isaac Kwame Baidoo and [#]Dennis Arku
[*#]Department of Statistics, University of Ghana, Box LG 115, Ghana.
Correspondence: denarku.gh@gmail.com

Abstract

This study examines the risk factors of loan defaults in the microfinance industry in Ghana and develops a model for assessing credit worthiness of customers in that sector. A binomial logistic model was fitted to a dataset of 472 customers. The findings identified that the loan officer, personal guarantors, age of customer, type of residence, type of collateral and assessment are key predictors of default. A test of the full model against a baseline model was statistically significant, indicating that the predictors, as a set, reliably distinguished defaulters and non-defaulters. The Receiver Operating Characteristic (ROC) that measures the sensitivity and specificity of the model was significant. Using a hold out sample, the model is able to classify defaulters and non-defaulters with at least 80% accuracy. The model could serve as a tool to manage and improve loan application decisions and ultimately improve enterprise profitability of microfinance institutions.

Keywords: *microfinance in Ghana, default risk, delinquency, logit model.*

1. Introduction

During the past three decades, many developing countries have gone through structural transformation in financial services and regulation due to the emergence of Microfinance Institutions (MFIs). Banks that hitherto did not want to lend to the poor due to the high risk of loan default associated with the informal sector are now doing so under the name of microfinance. Microfinance is defined by Robinson (2001) as, "The small-scale financial services provided to people within the petty businesses in both rural and urban communities". Thus, MFIs are primarily established to provide business and consumer credit to the poor and vulnerable groups in the society with the ultimate goal of improving living standards. The benefits of Microfinance to poverty alleviation and overall improvement in living standards have extensively been studied by many social science researchers. For instance, Vukson (as cited in

Awotwi, 2011) in his study showed that, health, education and self-esteem of the borrowers and their families had significantly improved as a result of emergence of MFIs. Khandkar (2005) and Hashemi et al, (1996) in another study also indicated that microfinance has contributed to women‘s empowerment and overall poverty reduction in most developing countries.

In Ghana, there are number of microfinance companies empowering and supporting the informal sectors of the economy. The number of registered MFIs has grown from 150 in 2008 to 364 in 2011 and continues to surge as the industry appears profitable to prospective market entrants (Ghana Association of Microfinance Companies (GAMC), 2011). The Greater Accra region alone constitutes over 40% of the number of MFIs in Ghana. Also, as at May, 2013 the number of Microfinance Institutions that are registered with Bank of Ghana has increased to 228 (BOG, 2013). This means that even though most MFIs are registered with Ghana Association of Microfinance Companies (GAMC), there are a sizable number (about 37%) of MFIs currently operating without a Bank of Ghana permit. Growth in outreach has closely followed existing high-density areas, leading to these regions becoming more and more saturated. As the number of MFIs continues to grow and competition becomes intense in the industry, concerns for the survival and long-term sustainability of these institutions also grow. Borrowing from more than one institution has become widespread in the absence of a fully-functioning centralized database that records clients by their national identity card numbers.

The need for micro-credit today is at its highest owing to the structure of the Ghanaian economy where majority (80%) of the workforce works in the informal sector (Boateng and Ampratwum, 2011).

The microfinance industry is considered inherently risky mostly due to the segment of population it serves in which incomes are low and credit histories are sparse or non-existent. The sustainability and growth of these institutions largely depends on quality of their loan portfolio. In most MFI’s the methods of deciding whether to grant a loan to an individual are mostly speculative and based on the subjective experience of previous decisions rather than objective analysis. While it may be that all lending decisions are speculative, decisions could be aided greatly by referencing the actual delinquency and underwriting data. Usually, the borrower puts in a request for loan and on that basis three things happen: the client makes a deposit (% requested amount), the client’s home and business is visited (Assessment) and collateral or a personal guarantor is requested. This is done in order to assess the character, capacity and capital of the client. Objective data is collected in the process but may not be sufficient or credible. In the absence of data, the manager thus has little objective basis to make a decision to approve or decline. This usually leads to guided but inadequate decisions resulting in increased rate of default. For instance,

in one MFI, it was found that as at December 2012, the company had disbursed GHC 1,739,041.00 out of which 79% had been in good standing whilst 21% of the portfolio value had been in arrears. In terms of number of clients, the 79% constituted 68% of the clients whilst the 21% arrears constituted 32% of the clients. This result shows poor decision making; the consequence of which is a potentially severe financial loss to the institution. There remains a substantial opportunity to improve MFI rates of loan default through the correct application of data analytics risk assessment.

The question is, can the outcome of the decision to advance a credit facility be successfully predicted with the data available and how can the bad performing customers be isolated for differential treatment? Better decision mechanisms separate good and bad clients. In addition, the growing sophistication of the lending public requires the development of computational approaches to evaluate a potential borrower's credit worthiness in a way that is fair and transparent. This study therefore was inspired by the need to ascertain factors that have a major influence on loan default by considering the attributes of defaulters and non defaulters. The findings of the paper contribute to the existing studies on credit risk modeling and we hope will have an impact on industrial practice.

The rest of the paper is structured as follows: section 2 presents a discussion of past related studies, section 3 provides an overview of the methodology and section 4 analyzes and discusses the results. Section 5 provides the conclusions and recommendations.

2. Literature review

Amilie and Allen (2006) identified three major financial ratios that significantly influence the probability of default. These are leverage, profitability and liquidity. These factors are relevant but in the microfinance setting it is difficult to assess individuals based on the financial ratios as most of the businesses in Ghana do not have the required systems or data. Agarwal et al (2009), assessed the role of individual social capital information characteristics on household default and bankruptcy outcomes. The study used monthly panel data set of more than 170 000 credit card holders over 24 months. The study found distress factors such as riskiness, spending, debt, income, wealth, economic conditions, legal environment and socio-demographical characteristics significantly affect default risk. These are helpful but it is difficult to make an assessment on these factors. According to Mario and Claudio (2010), in United States both socio-demographical and behavioral variables have a notable effect on default. The most significant socio-demographical variables are income, time since last moving, age, possession of credit card, education and nationality. Some behavioural variables seemed to have even more predictive power.

Those are the amount of scores the customer obtained, loan size and the information if customer has been granted a loan earlier from the same company. Interestingly, the results have variation to some extent when excluding few of the variables outside the model. The predictive power of all three models is adequate and thus can be employed as a reliable credit scoring model for the credit institutions.

Updegrave (1987) found that there were eight variables that affected consumer credit risk: the number of variables, the historic repayment record, bankruptcy history, work and resident duration, income, occupation, age and the state of savings account. Similar results were found by Steenackers and Goovaerts (1989) who collected data on personal loans in Belgian credit company and found out that age, resident and work duration, the number and duration of loans, district, occupation, phone ownership, working in the public sector or not, monthly income and housing ownership have a significant relationship with repayment behavior. Gan et al. (2012) did a research on default of mortgage loans in China and observed that a borrower's demographic characteristics impact banks' lending decision process. His results also suggested that the loan amount and interest rate range are positively correlated with the probability of loan default; which means an increase in the loan amount, interest rate level may increase the probability of loan default. Seven variables were identified as having a negative impact on the probability of loan default, good bank rate, occupation such as general manager, general staff, professional employee, whether the borrower resides within the same district with the bank.

Awotwi (2011) used a data from an international bank in Ghana, to study the risk of default. The data comprise of 9939 observations of customers of the bank of which 14% turned out to be defaulters. According to him, six variables were found to be highly significant in the default. These are Marital Status, Number of months the applicant has been in current employment, interest rate, tenure of loan or repayment period, income level and loan amount. However, much as his result is useful, the same response may not apply when it comes to the microfinance sector.

A number of statistical approaches are available for assessing borrowers' credit worthiness. These include discriminant analysis (Dunn and Frey, 1976), linear probability models (Turvey, 1991), probit models (Lufburrow et al., 1984) and logit models (Mortensen et al., 1988). The last three methods estimate the default rate based on the historical data on loan performances and the borrowers' characteristics based assumptions of the data. Linear Probability Model (LPM) assumes there is a linear relationship between the default rate and the factors. The Probit model assumes the probability of default follows the standard cumulative normal distribution function. Discriminant analysis divides borrowers into high and low default-risk classes and the data must be continuous (Mester, 1997).

In the 1980's a new method for classifying was introduced by Breiman et al. by splitting data into smaller and smaller pieces. According to the paper, the classification and regression tree approach is an appropriate method for classification of good and bad loans.

Apilado et al. (1974) applied discriminant analysis to construct their credit scoring models and state that "discriminant analysis firstly distinguishes among group and identifies group differences; secondly, it classifies existing and new observation into predetermined groups, and finally it identifies the key variables that contribute the most to the discrimination among groups" Discriminant analysis was used as a credit scoring tool first by Durand (1941) to produce good predictions of credit repayment. Extensive use of discriminant analysis, to build credit scoring models for general banks and credit card sectors have been carried out by Eisenbeis (1983), Martel and Fitts (1981), Grablowsky and Talley (1981), Reichart et al., (1983), Titterington (1992), Desai et al., (1996), and Lee et al., (1999).

According to Collins and Green (1982), the linear probability model could present reasonable prediction results compared to discriminant analysis and logit models. However, Pyndick and Rubinfeld (1998), Greene (1997), and Judge et al. (1985) indicated that the linear probability model could predict the default rate, but the predictive value might not necessary lie between zero and one. Moreover, because the variance of the models is generally heteroskedastic, it leads to inconsistent estimation problems and invalid conventional measure of fit such as the R^2. According to Hand and Henley (1997), the logistic approach is a more appropriate statistical tool than linear regression, when there are two discrete classes (good and bad risks) defined in the model. This gives the logistic approach superior classification rate. The probit model is very similar to the logit model. The logit model is generally preferred to the probit model because of its simplicity (Barney et al., 1999; Novak and LaDue, 1999; Lee and Jung, 1999).

Laitinen and Kankaanpää (1999), assessed six alternative methods (LDA, LR, RPA, survival analysis, and HIP) that have been applied to financial failure prediction. The main objective was to study whether the results stemming from the use of alternative methods differ from each other. They used only three financial ratios (total debt to total assets, the ratio of cash to current liabilities and the operating income to total assets) due to methodological issues. The results of 76 randomly selected from Finnish small and medium sized failed firms indicate that no superior method has been found but the predictive power of logistic analysis was best resulting 89.5% versus other methods. A systematic review of articles that have used logistic regression not only promotes the learning about this method, but also helps suggest new guidelines for principled applications of this versatile technique.

Logistic regression, being one special class of regression models, is well suited for

the study of categorical outcome variables, such as staying in or dropping out from college. Clarke, (2005) explained that, logistic modeling approach is commonly used to model bank's lending decisions. According to Collins and Green (1982), the logit model can increase the overall classification rate, and substantially reduce the error rate. The logistic approach also gives superior classification compare to discriminant analysis (Wiginto, 1980).

In this study logistic regression model approach is considered appropriate giving the data structure and purpose of the study. Logistic regression is special case of generalized linear models which is well suited for studies involving binary categorical variables.

3. Research methodology

This section outlines the methodology employed in the study.

3.1 Research design and data collection strategy

The design of the study was strictly quantitative. The study made use of historical data that contains the profile of customers. The data, and to a degree the inferences drawn from the analysis, are restricted to the Greater Accra region. While not randomly selected, the three MFIs selected are representative of the Greater Accra market in size, client base, and product offerings of MFIs. Most notably, the sample institutions have risk decisioning processes that are standard to the industry. The three MFIs were chosen due to access granted by management.

Given the sampling frame, all the files of loans disbursed under the period of the study were grouped into two – non default and default. The categorization was necessary for analytical purposes. Non-default: clients who were able to discharge their obligation within the given timeframe for payment or within 30 days after contract expiry. Default: clients who finished payment beyond one month or have not finished payment yet within the period given. According to Bank of Ghana Credit Act, default occurs if the obligor is unable to make payment beyond thirty days after contract period expires. On this basis the profiles of these clients (defaulters and non-defaulters) with their loan characteristics were then captured. Only those whose contracts were within the study period (2012-2014) were captured comprising a total of 472 clients. In all sixteen (16) variables were considered for this study. The variables were grouped into two main categories: borrower and loan characteristics namely; age, number of dependents, sex, type of business, number of years in residence, guarantor, type of collateral, assessment conducted, residential status, marital status, place of residence, amount disbursed, loan duration, loan cycle, loan officer, purpose of loan, date of disbursement. These variables were considered as the explanatory variables.

3.2 Model specification

3.2.1 Latent dependent variable creation

In this study a binary dependent variable default t (Y_i) was created. $i = 0,1$

$$\begin{cases} 1 \text{ defaulted; if delayed payment is} > 30 \text{ days} \\ 0 \text{ non-default; if delayed payment} \leq 30 \text{ days} \end{cases}$$

3.2.2 The logistic equation

Logistic regression predicts an outcome from a set of predictor variables. The model is more appropriate when the dependent variable is binary and categorical. It is also useful when the predictor variables are mixture of categorical or continuous variables. There are two types of logistic regression; binary regression is useful when the dependent variables is dichotomous and multinomial logistic regression is used when the dependent variables are more than two categories (Leech, Barret and Morgan,2005).

In this study, Binary logistic regression model was used. The logistic regression model predicts the logit of Y from X. The logit is the natural logarithm of odds of Y and odds are ratios of probabilities $(\hat{p})$ of default $(Y = 1)$ to probabilities (1 - $\hat{p}$) of non-default $(Y = 0)$.

Thus, generally, logit model is a derivative of the odds function. The odd of a function is the ratio of the probability of success to that of failure. Thus

$$odds(Y = 1) = \frac{P(Y = 1 / X = x)}{P(Y = 0 / X = x)} \tag{1}$$

Where '$odds(Y = 1)$' is the odds of default

$P(Y = 1 / X)$ is the probability that a default occurs given a set of explanatory variables X.

$P(Y = 0 / X)$, is the probability of non-default given set of explanatory variables X

The simplest Logistic Regression model is presented in equation

$$\log it(Y) = \ln(odds) = \ln(\tfrac{\hat{p}}{1-\hat{p}}) = \beta_o + \beta_1 X \tag{2}$$

Thus

$$\ln[\frac{P(Y = 1)}{1 - P(Y = 1)}] = \beta_o + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 \ldots, + X_n\beta_n$$

Where β is the regression coefficient of X, and α is the Y intercept, and $\hat{p}$ is the probability of default. By rearranging equation (2) one derives equation (3) to predict the probability of default.

$$P(Y=1)=\frac{1}{1+e^{-(\beta_o+X_1\beta_1+X_2\beta_2+X_3\beta_3\ldots,+X_n\beta_n)}} \tag{3}$$

According to equation 2 the relationship between logit (Y) and X is linear but according to equation (3) the relationship between the probabilities of Y given X is nonlinear.

Typically, $\beta_o,\beta_1,\ldots,\beta_n$ are estimated by Maximum Likelihood (ML) method, which is preferred over the weighted least squares approach by several authors, such as (Haberman,1978 and Schlesselman, 1992). The maximum likelihood method is designed to maximize the likelihood of reproducing the data given the parameter estimates. Data are entered into the analysis as 0 or 1 coding for the dichotomous dependent variable. The general likelihood of the data set can be represented as the product across all cases of the probabilities as presented in equation (4)

$$l=\prod_{i=1}^{n}P(Y_i/X_{i1,\ldots,}X_{ik})$$

$$l=\prod_{i=1}^{n}[(\frac{1}{1+e^{-\sum_{i=0}^{n}X_i,\beta_i}})^{Yi}(1-\frac{1}{1+e^{-\sum_{i=0}^{n}X_i,\beta_i}})^{1-Yi}] \tag{4}$$

Where Y is the 0 or 1 outcome for the ith case and $X_{i1\ldots}$, X_{ik} are the values of the predictor variables for the ith case base on a sample of n cases. The use of the exponents Y_i and 1 - Y_i is an indication that the appropriate probability estimate depends on whether $Y_i = 1$ or $Y_i = 0$.

The study made use of Logistic regression technique to develop a default model based on the fact that the dependent variable is categorical (ie. default or non-default).

4. Analysis of results

4.1 Preliminary analysis

Preliminary analysis was conducted to examine default classification by sector and other important categories. Table 1 below reports the findings.

Table 1: Default rate classification
The table presents the default rate for each variable holding others constant.

Variable	Category	Frequency		Total	Default rate
		Non-Default	Default		
SECTOR	Food industry	72	11	83	13%
	Manufacturing	19	8	27	30%
	Trade industry	109	97	206	47%
	Service industry	120	36	156	23%
RESIDENCE	Rented	197	119	316	38%
	Own house	91	26	117	22%
	Family house	32	7	39	18%
AMOUNT DISBURSED	Below 2000	224	108	332	33%
	2000-2900	73	25	98	26%
	3000-3900	8	2	10	20%
	4000-4900	10	7	17	41%
	5000+	5	10	15	67%
AGE	<30 Years	43	4	47	9%
	30-39 Years	179	69	248	28%
	40-49 Years	78	57	135	42%
	50+ Years	20	22	42	52%
MARITAL STATUS	Married	232	120	352	34%
	Single	80	23	103	22%
	Divorced	1	5	6	83%
	Widow/Widower	7	4	11	36%
SEX	Male	152	57	209	27%
	Female	168	95	263	36%
COLLATERAL	Stock	162	36	198	18%
	Household Items	65	102	167	61%
	Vehicle	54	12	66	18%
	Cheque	39	2	41	5%
GUARANTOR	Yes	119	16	135	12%
	No	201	136	337	40%

Source: fieldwork (2014)

The findings in Table 1 showed that default risk is lower with food vendors (13%) compared with manufacturing and trade sectors which recorded highest risk of default. In terms of residential status, those living in rented houses have a higher probability of default. With regard to loan size, it was observed that the loan size exceeding 4000 recorded a high risk of default. It was also observed that the elderly have a high default risk. In terms of marital status, divorced individuals tend to have higher risk of default than single. The findings also indicate that male customers perform better in their loan obligations than the female counterpart. With collateral, the study observed that those who use household items as collateral have higher risk of default (61%), whiles those who use cheque as a form guarantee has lower risk of default. Customers with personal guarantors tend to perform better than those without personal guarantors.

4.2 The MFI credit risk model

Sixteen (16) variables were initially considered as having influence on loan default as it has been the number of factors that credit decisions are made in most MFIs. Data on 472 customers were collected for analysis. Consistent with model development principles, about 332 (70%) customers were randomly selected as the sample, whilst 140 (30%) customers were isolated to validate the model. The dependent variable (default) was created by the researcher. Thus it was categorised 1 when default occurs and 0 if default did not.

Table 2 presents the baseline default risk model, thus, the model with only the constant included before any coefficients were added or entered into the equation. This is done to ascertain the usefulness or otherwise of considering other covariates to build a model.

Table 2: Classification Table for the baseline model

		Predicted					
		Model data			Validation data results		
Observed		Non-default	Default	Percentage Correct	Non-default	Default	Percentage Correct
Defaulted	Non-default	217	0	100.0	103	0	100.0
	Default	115	0	.0	37	0	.0
Overall Percentage				65.4			73.6

The study compares the outcome of the baseline model with that including all the predictors to determine which gives efficient result. The information in the Table 2

suggests that if one knew nothing about the client requesting for a loan and guessed that a person will default, one will be correct 65.4% of the time. This means that the baseline model has some predictive power but only 65.4% on average. In microfinance, 30% incorrect predictions could be very disastrous considering the nature of the business.

When the 16 covariates were added as part of explanatory variables to default, only Six (6) variables were found significant. Table 4 below summarizes the result.

Table 3: Significant predictive variables

Variable	B	S.E.	Wald	Df	Sig.	Exp(B)
Loan Officer			41.56	5	.00	
Residential status			7.75	2	.02	
Guarantor	-2.212	.62	12.91	1	.00	.110
Assessment	-1.761	.89	3.91	1	.04	.172
Collateral type			33.15	3	.00	
Age	.084	.02	12.99	1	.00	1.088
Constant	-25.028	7307.66	.00	1	.99	.000

Loan officer, residential status, guarantor, assessment, and age. Intuitively, these variables play key roles in loan portfolio management. The Exp(B) column represents the odds ratio. It represents the extent to which alternating the dichotomous variable or raising the continuous measure by one unit influences the odds ratio. If the value exceeds 1 then the odds of an outcome occurring increase; but if it is less than 1 any increase in the predictor leads to a drop in the odds of the outcome, if it is equal to 1, both groups have the same odds. For instance, the odds ratio for 'age' is 1.088. Since age is a continuous variable, the interpretation is that a client whose is 31 years old is 1.088 times more odds of default when compared with a client who is 30 years old.

Also the odds ratio for 'Assessment' is 0.172. This suggests that a client who is assessed before loan is 0.172 times less odds of default when compared with someone who was not assessed.

To design an appropriate model the categories for each significant variable were re-classified 1 and 0. The summary of the result is presented in table 4 below;

Table 4: The Predictor Variables

Description	Model variable	B: (Coefficients)	Exp(B): Odds ratio
Constant	X_o	-25.03	0.00
1 if client was handled by officer A and 0 otherwise	X_1	4.51	90.99
1 if client was handled by officer B and 0 otherwise	X_2	1.08	2.95
1 if client was handled by officer C and 0 otherwise	X_3	1.62	5.04
1 if client was handled by officer D and 0 otherwise	X_4	1.61	5.00
1 if client was handled by officer E and 0 otherwise	X_5	3.96	52.30
1 if client lives in rent and 0 otherwise	X_6	1.00	2.69
1 if client lives in own house and 0 otherwise	X_7	-0.36	0.70
1 if client has a guarantor and 0 otherwise	X_8	-2.21	0.11
1 if client was assessed and 0 otherwise	X_9	-1.76	0.17
1 if client used stock and 0 otherwise	X_{10}	19.57	3.16E8
1 if client used Household items and 0 otherwise	X_{11}	21.51	2.21E9
1 if client used Vehicle and 0 otherwise	X_{12}	18.31	8.98E7
Age of client	X_{13}	0.08	1.09

Source: fieldwork (2014)

The default model is then given as;

$$P(Y=1)=\frac{1}{1+e^{-(-25.03+4.51X_1+1.08X_2+1.62X_3+1.61X_4+3.96X_5+0.99X_6-0.36X_7-2.21X_8-1.76X_9+19.57X_{10}+21.51X_{11}+18.31X_{12}+0.08X_{13})}} \quad ...(5)$$

The predictive ability of the model was assessed and the result presented in Table 5.

Table 5: Predicted and observed probabilities of default

		Model data			Validation data		
		Non-default	Default	Percentage Correct	Non-default	Default	Percentage Correct
Defaulted	Non-default	199	18	91.7	93	10	90.3
	Default	32	83	72.2	13	24	64.9
Overall Percentage				84.9			83.6

Source: fieldwork (2014)

In Table 5, the prediction and validation results are displayed. The table indicate that 199 out of the 217 (91.7%) applicants who did not defaulted are classified correctly. Also 83 out of the 115 (72.2%) defaulters were classified correctly. Overall, 84.9% of the applicants were classified correctly. The result of the validation data also showed that 83.6% were correctly classified by the model. This suggest that the model is able to correctly predict the delinquency outcome four times out of five.

5. Conclusions and recommendations

5.1 The key findings in the analysis are:

- Loan officer, residential status, guarantor, assessment, age and type of collateral are significant in predicting default.
- Customers in the trading and manufacturing sectors have high risk of default and practitioners should take extra precaution to mitigate the risk
- A credit default risk model developed on portfolio data could assist MFIs in loan management.

5.2 Recommendations

1. It is recommended that on the basis of the findings that Microfinance Institutions use the credit default model as an additional tool to ascertain the level of risk of its clients. After calibration of the model to their own experience, which would involve a reweighting of the factors as opposed to a full model development, the applicants score should be used as an additional screening mechanism. Results of the trial could then be assessed against the current business practice in a champion-challenger comparison. It expected that the 80% validity of the development model should hold up in industrial application.

2. Another factor that should be taken into consideration is the influence of the loan officer in determining default. If MFIs are to achieve the scale necessary to spread risk across both good and bad subpopulations of applicants, there needs to be uniformity in decisioning. This can be achieved through improved training of personnel, adequate documentation of risk criteria to be used, and adoption of a risk scoring tool.
3. Bank of Ghana should consider enacting a law that will ensure that all MFIs in Ghana are put on a centralized database to check multiple borrowing and also serve as an internal control measure for sustainable MFIs in Ghana.

5.3 Limitations of the study and further research

1. The study was restricted to MFIs in the Greater Accra region. While we feel that the resulting model is structurally representative of the national MFI experience, this needs to be validated on a geographically diverse sample of institutions
2. The model weights would need to be calibrated before the score be deployed in an industrial application. This would allow for idiosyncratic differences associated with specific through-the-door populations to be captured in the scoring calculation.
3. Further research into modelling the severity of loan loss in addition to the incidence work contained in this paper.

References

Agarwal, S., Chomsisengphet, S. and Liu, C. (2008). Consumer Bankruptcy and Default: The Role of Individual Social Capital. *Working Paper*. Available at SSRN: http://ssrn.com/abstract=1408757.

Allen, M.F and Amelie, J. (2006), Determining the probability of default of Agricultural Loans in French Bank. *Journal of Agricultural Economics*. 67:513-20.

Apilado, V.P., Warner, D.C. and Dauten, J.J.(1974). Evaluative technologies in consumer finance-experimental results and policy implications. *J. Finance Quant.Anal, Mar*., 275-283.

Awotwi, E.(2011). Estimation of probability of default of consumer credit in Ghana: A case study of an international bank. Kumasi:KNUST. Bank of Ghana quarterly report (2013). Microfinance in Ghana. retrieved from http//:www.modernghana.com/news/317528/1/loan-default on 15th May 2013.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). Classification and regression Trees. Wadsworth Belmont: California.

Clarke, J.A.(2005). Implications of stratified sampling for fair lenders binary logit models. *Journal of Real Estate Finance and Economics*, 30(1), 5.

Collins, R.A and Green, R.D.(1992). Statistical Methods for Bankruptcy Forecasting. *Journal of Economics and Business*,34,349-354.

Crook, J. N., Hamilton, R. and Thomas, L. C. (1992). A Comparison of a credit Scoring model with a credit performance Model. *The Service Industries Journal*, Volume 12, Issue 4, 558-579.

Desai, V.S., Crook, J.N. and Overstreet, G.A. (1996). A comparison of neural networks and linear scoring models in the credit environment. *European J. Operational Res*.95, 24-37.

Desai, V. S., Crook, J. N. and Overstreet, G. A. (1996).A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, Vol. 95, Issue 1, p. 23-37.

Durand D., (1941). Risk elements in consumer instalment financing. National Bureau of Economic Research: New York.

Edelman, D.B. (1997). Credit scoring for lending to small businesses, Proceedings of Credit Scoring and Credit Control (V). Credit Research Centre: University of Edinburgh.

Eisenbeis, R.A. (1977). Pitfalls in the application of discriminant analysis in business, finance and economics. *Journal of Finance* 32, 875-900.

Eisenbeis, R.A. (1978). Problems in applying discriminant analysis in credit scoring models. *Journal Banking and Finance* 2, 205-219.

Gan, C., Li, Z., Wang, W., & Kao, B. (2012). Credit scoring in mortgage lending: evidence from China. *International Journal of Housing Markets and Analysis*, *5*(4), 334-350.

Ghana Association of Microfinance Companies, GAMC. Retrieved from http://ghana-microfinance.com, on 20th March.2012.

Grablowsky, B.J. and Talley, W.W(1981). Probit and significant functions for classifying credit applicants; a comparison. *J. of Economics and Business*, 33, 254-261.

Greene, W.(1997). Economic Analysis, (3rd ed): Prentice Hall.

Hand, D. J. and Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *160*(3), 523-541.

Haberman, S. (1978). Analysis of quanlitative data (vol.1). Academic press, New York.

Hashemi, S. M. (1990). Rural Credit program and Women Empowerment; World Development 24(4):635-53.

Judge G.G., Griffiths, W.E., Hill, R.C., Helmut, L. and Lee,T.C (1985). The Theory and Practice of Econometrics. 2nd ed, New York,NY: John Wiley and Sons.

Laitinen, E. and Laitinen, T. (2000). Bankruptcy prediction: Application of the Taylor's Expansion in logistic regression. International Review of Financial Analysis, Vol. 9, 327-349.

Laitinen, T. and Kankaanpää, M. (1999). Comparative analysis of failure prediction methods: The Finnish case'. European Accounting Review, Vol. 8, Issue 1, 67-92.

Lee, T. S., Chiu, C. C., Lu, C. J. and Chen, I. F. (1982). Credit scoring using the hybrid neural discriminant technique. Expert Systems with Applications, Vol. 23, Issue 3, 245-254.

Leech, N.L., Barrett, K.C. and Morgan, G.A.(2005). SPSS for intermediate statistics: Use and Interpretation, New Jersey, Lawrence Erlbaum Associates, Inc.

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

Khandker, S.(2003). Microfinance and Poverty: Evidence using panel data from Bangladesh; World Bank Policy Research Working Paper 2945.

Mario, J.M. and Claudio, G.(2010). Systemic Risk, Index Insurance and optimal management of agricultural loan portfolios in Developing Countries, *American Journal of Agricultural Economics*,93,2,399-406

Martell, T.F. and Fitts, R.L. (1981). A quadratic discriminant analysis of bank credit card user characteristics, *J. of Economics and Business* 33, 153-159.

Mester, L. (1997). What's the point of credit scoring? Federal Reserve Bank of Philadelphia Business Review, September/October, 3-16.

Merton, R.C. (1974). On the pricing of corporate debt: the risk structure of interest rates, *Journal of Finance 29, 449-470.*

Mortensen, T., Watt, D.L. and Leistritz, F.L.(1988). Predicting Probability of Loan Default. *Agricultural Finance Review* 48 : 8-14

Novak, M.P. and LaDue E.L.(1999). Application of Recursive Partitioning to Agricultural Credit Scoring. *J. Agric and Appl. Econ*. 31,1:109-122.

Osei-Boateng, C. and Ampratwum, E.(2011). The Informal Sector in Ghana. Friedrich Ebert Stiftung.Ghana office.

Pyndick, R.S. and Rubinfeld, D.L.(1998). *Econometric Models and Economic Forecast*. 4th ed, McGraw Hill.

Reichert, A.K.,Cho, C.C. and Wagner, G.M.(1983). An examination of the conceptual issues involved in developing credit scoring models. *Journal of Business and Economic Statistics*, 1(10),101-114.

Robinson, M. (2001). *The microfinance revolution*. Washington DC, World Bank Publications.

Schlesselman, J.J.(1982). Case control studies: Design, control, analysis. New York, Oxford University Press.

Steenackers, A. and Goovaerts, M. J. (1989). A Credit Scoring Model for Personal Loans. Insurance: Mathematics and Economics, Vol. 8, Issue 1, 31-34.

Turvey, C.G.(1991). Credit Scoring for Agricultural Loans: A Review with Applications. *Agricultural Finance Review* 51 : 43-54.

Updegrave, W.L.(1987). How Lender Size You Up. *Money*. 23-40.

Wiginton, J. C. (1980). A note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour, *Journal of Financial and Quantitative Analysis* 15, 757-770.