Comparison of Logistic Regression and Linear Discriminant Analyses of the Determinants of Financial Sustainability of Rural Banks in Ghana

Godfred Kwame Abledu[1, *], Akuffo Buckman[1], Thomas Adade[2], Samuel Kwofie[1]

[1]Department of Applied Mathematics, Faculty of Applied Sciences and Technology, Koforidua Polytechnic, Koforidua, Ghana
[2]Department of Accountancy, Faculty of Business and Management Studies, Ho Polytechnic, Ho, Ghana

Email address:
godfredabledu@gmail.com (G. K. Abledu)

Abstract: Financial Sustainability is a primary issue for successful rural and community banks' services. Establishing a system of sustained provision of modern financial services has, however, been challenging and most controversial. Several studies have been conducted on the determinants of sustainability of institutions in various countries. However, the levels of significance of the factors that influence financial sustainability of banks vary with studies. In addition, the results are mixed and empirical evidence regarding the determinants of rural and community banks' sustainability is also missing. The objective of this study therefore was to develop a model which could be used to identify likely future rural and community banks that are non-sustainable. This study examined the determinants of financial sustainability of Rural and community banks using discriminant analysis (LDA) and logistic regression (LR) models.

Keywords: Rural and Community Banks, Linear Discriminant Analysis, Logistic Regression, Financial Sustainability

1. Introduction

The first Rural and Community Bank (RCB) was established in a farming community in the central region of Ghana in 1976 [1]. Rural communities showed tremendous interest in the community ownership and management features of RCBs, and by 1984 the number of RCBs reached 106. The introduction of a check payment system for cocoa farmers (known as the Akuafo Check operation) also spurred the establishment of local banks in many communities. In 1981 about 30 existing RCBs formed an Association of Rural Banks (ARB) to serve as a networking forum. As a network of institutions sharing a common mission, the ARB promoted and represented the RCBs and also provided training services to member RCBs.

The financial performance of many RCBs started to decline, however, for several reasons, including a drought that affected the country in 1983 (leading to high loan default rates), weak governing ability, conflicts within boards of directors, and ineffective management in many RCBs. Several reforms were undertaken to curb the deteriorating situation—exposure to risky sectors (mainly agriculture) was limited, distressed banks were closed, supervision by the Central Bank was strengthened, and RCB managers and boards of directors were offered training. Nevertheless, RCBs continued to be relevant rural finance service providers, and the Government of Ghana has consistently provided support to the RCBs by financing capacity building (in partnership with several donors), restructuring programs, and undertaking regulatory reforms.

By the end of 2008, 127 RCBs were in operation with a total 584 service outlets. RCBs are regulated by Ghana's Central Bank, the Bank of Ghana, and thereby form part of the country's regulated financial sector. RCBs are the largest providers of formal financial services in rural areas and represent about half of the total banking outlets in Ghana [2].

RCBs are relatively small financial institutions with average share capital of GHc 136,526 (US$105,263), average deposits of GHc 2.3 million (US$1.77 million), and average assets of GHc 3.8 million (US$2.4 million). Values of the three indicators, however, vary significantly. Out of the 127 RCBs, 75 percent have assets between GHc 1

million (US$771,010) and GHc 8 million (US$6.1 million), 20 percent have assets of less than GHc 1 million, and 5 percent have assets over GHc 10 million (US$7.7 million). Similarly, 44 percent of RCBs have share capital of less than GHc 100,000 (US$77,101) and only 6 percent have share capital of more than GHc 250,000 (US$192,753).

As a network, RCBs have achieved a remarkable level of service delivery and financial performance. At the end of 2008, they had deposits of GHc 343.9 million (US$265.1 million) from more than 2.8 million clients, and loans and advances of GHc 224.7 million (US$173.2 million) with about 680,000 clients. They delivered 128,875 domestic money transfers worth around GHc 63.3 million (US$48.8 million) in 2007 and 32,392 international money transfers worth GHc 9.3 million (US$7.1 million) in 2008. They also facilitated check transactions worth GHc 993.7 million (US$766.1 million) in 2008. RCBs made a consolidated profit of GHc 15.6 million (US$12.0 million) in 2008 and had a consolidated net worth of GHc 62.3 million (US$48.03 million). Several have RCBs excelled in performance, both within the financial sector and in the broader private sector. Some rural banks have figured more than once in Club 100, a group of 100 Ghanaian institutions recognized annually for business excellence.

Several challenges, however, remain. The Bank of Ghana (BoG) rated the performance of 17 of the 127 rural banks in operation as mediocre, based on capital adequacy, and it categorized 5 banks as distressed. Among the banks whose performance is categorized as mediocre, 6 rural banks have negative net worth. The Apex Bank of the network, which was created primarily to provide services to rural banks, is not yet fully financially self-sufficient and has inadequate resources to effectively perform its functions. The BoG, which is primarily responsible for supervising RCBs, is constrained in effectively performing its supervision role because of political and civil society pressures, resource constraints, and limited delegation of supervisory functions to the Apex Bank.

2. Financial Sustainability and Ghana's Rural Bank

Broadly, sustainability refers to the ability of administrators to maintain an organization over the long term. However, the definition of financial sustainability may vary widely for-profit organizations and nonprofits (defined as organizations that use surplus revenues to achieve their goals rather than distributing them as profit or dividends), depending on the business structure, revenue structure, and overarching goal of the organization. Sustainability is now increasingly recognized as central to the growth of emerging market economies. For the private sector, this represents both a demand for greater social and environmental risk management as well as a new landscape of business opportunities.

For both profit and nonprofit organizations, financial capacity consists of resources that give an organization the ability to seize opportunities and react to unexpected threats while maintaining general operations of the organization [3]. It reflects the degree of managerial flexibility to reallocate assets in response to opportunities and threats. Financial sustainability refers to the ability to maintain financial capacity over time [3]. Regardless of an organization's for-profit or nonprofit status, the challenges of establishing financial capacity and financial sustainability are central to organizational function [3]. However, maintaining the ability to be financially agile over the long term may be especially important for nonprofits organisations, given that many of them serve high-need communities that require consistent and continually available services. With this in mind, the goal of financial sustainability for nonprofits is to maintain or expand services within the organization while developing resilience to occasional economic shocks in the short term (e.g., short-term loss of programmme funds, monthly variability in donations).

Ghana's Rural Bank scheme was initiated in 1976, under the auspices of the Bank of Ghana (the country's central bank). The purpose of this program was to serve small borrowers and savers in rural areas, who at the time had essentially no access to institutional savings and credit facilities. RFM specialists would recognize in this program many elements of the Directed Credit Approach. For its time, however, the Rural Bank project was relatively well thought out. Many features of this program, indeed, foreshadowed the yet-to-be developed Financial Systems Approach to RFM intervention. During its first decade of operations, the Rural Bank program proved, in general, to be a success. By the late 1980s, however, many individual Rural Banks were floundering. The government attempted to reinvigorate the programmme via a macroeconomic Financial Liberalization effort initiated in 1988 and a comprehensive Rural Bank restructuring exercise begun in 1991. Despite these efforts, in the mid 1990s the 125 Rural Banks in operation were, in general, not fulfilling their promise-and struggling financially [4, 5. 6].

3. Related Works

The goal of Logistic Regression (LR) is to find the best fitting and most parsimonious model to describe the relationship between the outcome (dependent or response variable) and a set of independent (predictor or explanatory) variables. The method is relatively robust, flexible and easily used, and it lends itself to a meaningful interpretation. In LR, unlike in the case of Linear Discriminant analysis (LDA), no assumptions are made regarding the distribution of the explanatory variables. Contrary to the popular beliefs, both methods can be applied to more than two categories [7. 8]. The LR model can be expressed as the case of a dichotomous outcome variable (Y) as:

$$P(Y_i = 1 \mid X_i) = \frac{e^{\beta^T X_i}}{1 + e^{\beta^T X_i}} \qquad (1)$$

where the Y_i are independent Bernoulli random variables. The coefficients of this model are estimated using the maximum likelihood method.

Linear Discriminant Analysis (LDA) can be used to determine which variable discriminates between two or more classes, and to derive a classification model for predicting the group membership of new observations [7, 8]. For each of the groups, LDA assumes the explanatory variables to be normally distributed with equal covariance matrices. LDA is discussed further by [9]. The standard LDA model assumes that the conditional distribution of X|y is multivariate normal with mean vector μy and common covariance matrix S. With some algebra, can be shown that if x is assigned to group 1, then:

$$P(1 \mid x) = \frac{1}{1 + \left(e^{\alpha + \beta x}\right)^{-1}} \tag{2}$$

where α and β coefficients are

$$\alpha = -\log \frac{\pi_1}{\pi_0} + \frac{1}{2}(\mu_1 + \mu_0)^T \Sigma^{-1}(\mu_1 - \mu_0) \text{ and } \beta = (\mu_1 - \mu_0)^T \Sigma^{-1},$$

π_1 and π_0 are prior probabilities of belonging to group 1 and group 0. In practice the parameters π_1, π_0, μ_1, μ_0 and Σ will be unknown, so they are replaced by their sample estimates, i. e.:

$$\hat{\pi}_1 = \frac{n_1}{n}, \ \hat{\pi}_0 = \frac{n_0}{n}$$

$$\hat{\mu}_1 = \overline{x}_1 = \frac{1}{n_1} \sum_{y_i=1} x_i \ , \ \hat{\mu}_0 = \overline{x}_0 = \frac{1}{n_0} \sum_{y_0=1} x_i \tag{3}$$

$$\hat{\Sigma} = \frac{1}{n}\left[\sum_{y_i=1}(x_i - \overline{x}_1)(x_i - \overline{x}_1)^T + \sum_{y_i=0}(x_i - \overline{x}_0)(x_i - \overline{x}_{10})^T \right] \tag{4}$$

Equation (1) is equal in form to equation (2). Hence, the two methods do not differ in functional form, but differ only in the estimation of coefficients.

Since more information is needed regarding the predictive accuracy of the methods than just a binary classification rule, [7, 8] proposed four different measures of comparing predictive accuracy of the two methods. These measures are indexes A, B, C and Q. They are better and more efficient criteria for comparisons and they indicate how well the models discriminate between the groups and/or how good the prediction is. Theoretical insight and experiences with simulations revealed that some indexes are more and some less appropriate at different assumptions. In this work, the focus is on three measures of predictive accuracy, the B, C and Q indexes. Because of its intuitive clearness the classification error (CE) is sometimes added as well. The C index is purely a measure of discrimination (discrimination refers to the ability of a model to discriminate or separate values of Y). It is written as follows;

$$C = \frac{1}{n_0 n_1} \sum_{\substack{i=0 \\ Y_i=0}}^{n} \sum_{\substack{j=1 \\ Y_j=1}}^{n} \left[I(P_j > P_i) + \frac{1}{2}I(P_j = P_i) \right] \tag{5}$$

where P_k denotes an estimate of $P(Y_k=1 \mid X_k)$ from (1) and I is an indicator function. It is seen that the value of the C index is independent of the actual group membership (Y), and as such it is only a measure of discrimination between the groups, and not a measure of accuracy of prediction. A "C" index of I indicates perfect discrimination; a "C" index of 0.5 indicates random prediction.

The B and Q indexes can be used to assess the accuracy of the outcome prediction. The B index measures an average of squared difference between an estimated and actual value:

$$B = 1 - \frac{1}{n}\sum_{i=1}^{n}\left(P_{i\_}Y_j\right)^2 \tag{6}$$

where P_i is a probability of classification into group i, Y_i is the actual group membership (1 or 0), and n is the sample size of both populations. The values of the B index are on the interval [0, 1], where 1 indicates perfect prediction. In the case of random prediction in two equally sized groups, the value of the B index is 0.75. The Q index is similar to the B index and is also a measure of predictive accuracy:

$$Q = \frac{1}{n}\sum_{i=1}^{n}\left[1 + \log_2\left(P_i^{Y_j}(1-P)^{1-Y_i} \right) \right] \tag{7}$$

A score of 1 of the Q index indicates perfect prediction. A Q index of 0 indicates random predictions, and values less than 0 indicate worse than random predictions. When predicted probabilities of 0 or 1 exist, the Q index is undefined. While the C index is purely a measure of discrimination, the B and Q indexes (besides discrimination) also consider accuracy of prediction.

Linear Discriminant Analysis (LDA) is a statistical tool that can predict the group membership of a newly sampled observation [10]. [7, 8, 10, 11] have recently proposed a new type of nonparametric LDA approach that provides a set of weights of a linear discriminant function, consequently yielding an evaluation score for the determination of group membership. The nonparametric LDA is referred to as "Data Envelopment Analysis-Discriminant Analysis (DEA-DA)," because it maintains its discriminant capabilities by incorporating the nonparametric feature of DEA into LDA. In this study, a use of two statistical tests is proposed for DEA-DA and its discriminant capability is compared with DEA from a perspective of financial analysis.

Linear Discriminant Analysis in this study by [12, 13] is used to examine empirically whether current cost accounting (CCA) information may be useful for predicting the performance of small companies. A matched sample of failed and non-failed firms is chosen and historic cost accounts are adjusted in line with the requirements of Statement of Standard Accounting Practices (SSAP). The companies are all single-plant independently owned firms in the Northeast of England; all the failed firms had ceased to trade during 1974-1980.

[14] mentioned that the LDA technique has the advantage of considering an entire profile of characteristics common to the relevant firms and another advantage of LDA in dealing with classification problems is the potential of analyzing the entire variable profile of the object simultaneously rather than sequentially examining its individual characteristics. Most recent research on the use of discriminant analysis on evaluating company performance in Ghana is by [9]. This research is using 11 ratios as independent variable to determine the performance of finance company in financial industry in Ghana.

Logistic regression is a form of regression which is used when the dependent is a dichotomy and the independents are of any type [15]. Continuous variables are not used as dependents in logistic regression. Unlike logit regression, there can be only one dependent variable. Logistic regression can be used to predict a dependent variable on the basis of continuous and/or categorical independents and to determine the percent of variance in the dependent variable explained by the independents; to rank the relative importance of independents; to assess interaction effects; and to understand the impact of covariate control variables.

Logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable (the natural log of the odds of the dependent occurring or not). In this way, logistic regression estimates the probability of a certain event occurring [16]. Logistic regression has many analogies to OLS regression: logit coefficients correspond to b coefficients in the logistic regression equation, the standardized logit coefficients correspond to beta weights, and a pseudo R statistic is available to summarize the strength of the relationship. Unlike OLS regression, however, logistic regression does not assume linearity of relationship between the independent variables and the dependent, does not require normally distributed variables.

Logistic regression also does not assume homoscedasticity, and in general has less stringent requirements. It does, however, require that observations are independent and that the independent variables be linearly related to the logit of the dependent. The success of the logistic regression can be assessed by looking at the classification table, showing correct and incorrect classifications of the dichotomous, ordinal, or polytomous dependent. Also, goodness-of-fit tests such as model chi-square are available as indicators of model appropriateness as is the Wald statistic to test the significance of individual independent variables.

Because both LDA and LR can be used for predicting or classifying individuals into different groups based on a set of measurements, a logical question often asked is: How do the two techniques compare with each other? In the literature, there has been considerable discussion about the relative merits of the two different techniques [17, 18]

Theoretically, LDA is considered as having more stringent data assumptions. Two prominent assumptions for LDA are multivariate normality of data and homogeneity of the covariance matrices of the groups [19, 20, 21]. However, it is not entirely clear what consequences the violation of those assumptions has on LDA analysis results. LR, on the other hand, is considered relatively free of those stringent data assumptions [22, 23, 24]. Although there is no strong logical reason to expect the superiority of one technique over the other in classification accuracy when the assumptions for LDA hold, it would be reasonable to expect that LR should have the upper hand when some of those assumptions for LDA are not tenable [23, 24].

Research findings about the relative performance of the two methods appear to be inconsistent. With regard to data normality, [25] showed that under the optimal data condition of multivariate normality and equal covariance matrices for the groups, LDA is more economical and more efficient than LR. When the data are not multivariate normal, results from some simulation studies [26, 27] indicated that LR performed better than LDA. That finding, however, has not been unequivocally supported by the studies in which researchers compared the two techniques by using extant data sets; in quite a few studies involving actual nonnormal data sets, very little practical difference has been found between the two techniques [26, 29, 30].

With regard to the condition of equal covariance matrices for LDA, there are few empirical studies comparing the relative performance of LDA and LR for unequal covariance matrices. Researchers seem to assume that LR should be the method of choice when the two groups do not have equal covariance matrices [31, 32]. Several studies that involved extant data sets did not suggest that LDA's performance would suffer appreciably because the assumption was violated [30, 33]. No one seems to have specifically manipulated that condition in simulation studies to examine its effect on the performance of LDA and LR.

The relative performance of LDA and LR under different sample-size conditions is also an issue of interest. Viewed from the perspective of statistical estimation in general, maximum likelihood estimators (as in LR) tend to require larger samples to achieve stable results than ordinary least square estimators (as in LDA). Inconsistent results have been reported about the relative performance of the two techniques with regard to sample-size conditions. For example, in a simulation study, [31] implied that LDA performed better under small sample-size conditions. [19] showed that when the techniques were applied to real data sets, the findings did not clearly confirm that conclusion.

There are several reasons for the limited internal and external validity of those studies. First, using extant data

sets gives researchers no control of data characteristics, thus making it impossible to systematically investigate the impact of each individual factor, because in extant data sets the effects of those relevant factors are often hopelessly confounded. Second, most of those studies did not provide enough information about the data characteristics, making it very difficult to synthesize the results across studies. For those reasons, simulation studies with strong experimental control are useful in assessing the effects of those relevant factors.

4. Methodology

This paper is based on a review of various published and unpublished documents, interviews with key respondents, and an analysis of data collected from the BoG, the ARB Apex Bank, and a sample of rural banks. The sampled rural banks were selected primarily to reflect the proportional representation of different categories of rural banks according to the BoG's performance classification of all 127 banks. Other factors used to select the sample of banks were location (primarily rural or periurban, and agroclimatic zone), size, and age.

The performance analysis of the rural banks is primarily presented at the network level (consolidated data for all RCBs) using secondary data available from the ARB Apex Bank, the BoG, and other secondary sources. Whenever the necessary data are not available at the network level, data from the sample banks are used, if available. The analysis includes trends, frequencies, and composition of key indicators and their comparison with data from peer-group institutions.

The independent variables that were associated at a significance level a = 0.05 with the dependent variable "financial sustainability" were entered in a principal components analysis (PCA). Sixteen variables satisfied the above criterion, so 16 principal components were extracted from the analysis. Applying Kaiser's criterion (eigenvalue >1), we retained 6 mutually independent factors. The assumptions for the two models were all fulfilled and the components due to their extraction methods followed the multivariate normal distribution and were mutually independent. The variance - covariance matrices of the groups were equivalent. We used the standardized canonical discriminant function coefficients and the unstandardized function coefficients for discriminant analysis and Z statistic (squared Wald statistic) for logistic regression, to evaluate how much each one of the variables contributes to the discrimination between two groups.

To identify the factors that influence the financial sustainability of RCBs, we utilized panel data on RCBs in Ghana for the years 2005 through 2012. This yielded unbalanced panel data for 190 RCBs. The RCBs' data is collected from individual institutions as reported to MIX market. In this study financial self-sufficiency is used as

dependent variable since the study seeks to identify determinants of financial sustainability of RCBs. The independent variables are: Inflation rate(x_1), Interest rate (x_2), Portfolio at risk (x_3), Operating expense /asset ratio (x_4), Debt/equity ratio (x_5), and Deposits to total assets (x_6).

The contribution of the respective variables to the discrimination depends on how large the coefficients are. We also compared the sign and magnitude of coefficients. Box's M test was used to check the equality of the covariance matrices, and it was revealed that they were equal (P>.05), thus this assumption for discriminant analysis was met. For each model, we plotted the corresponding response operating characteristics (ROC) curve.

An ROC curve graphically displays sensitivity and 100% minus specificity (false positive rate) at several cutoff points. By plotting the ROC curves for two models on the same axes, one is able to determine which test is better for classification, namely, that test whose curve encloses the larger area beneath it. All analyses were performed using the SPSS version 17.0 software.

5. Empirical Results

In this section, a discussion of the determinants of financial sustainability of RCBs which are measured by using LDA and LR is presented. Using PCA and applying Kaiser's criterion, 6 variables of our original data were extracted. These variables were used in both discriminant and logistic regression analyses, and both techniques revealed the same results. We observe that the direction of the relationships was the same, and there were not extreme differences in the magnitude of the coefficients. The overall correct classification rate was 81.3% for discriminant analysis and 83.1% for logistic regression analysis.

Approximately 68.4 %(130) of the sampled RCBs was used as training set to create the model. The remaining RCBs, 31.6% (60) was used to validate the model results. The classification function was used to assign cases to groups. The binary grouping variable was defined to be 0 if the RCD is not sustainable and 1 if MFI is sustainable. After the appropriate functions were calculated, the individual RCBs in both the training and validation sets were classified from the estimate functions, that is, the functions estimated from the training sets).

The logistic regression classified 88 (79 + 9) of the 130 RCBs in the training set correctly for a 67.7% classification rate (see Table 1). In the validation set, 38 (25 + 13) of the 60 RCBs were correctly classified, for a 63.3% correct classification rate.

The discriminant analysis correctly classified 86(80 + 6) of the 130 RCBs in the training set, for a 66.2 percent correct classification rate. In the validation set 36(33+3) of the 60 RCBs were correctly classified, for a 60 percent correct classification rate. The prior probabilities used were 0.66 of non-sustainability, and 0.34 of sustainability.

Table 1. *Summary of Classifications of RCBs by Logistic Regression and Discriminant Function Methods.*

Cases	Discriminant Analysis			Logistic Regression Analysis		
	Actual Group		Classification Rate (%)	Actual Group		Classification Rate (%)
	0	1		0	1	
Training Set						
Default Group(0)	80	7		79	7	
Non-Default Group(1)	37	6		35	9	
Total	117	13	67	114	16	71
Validation Set						
Default Group(0)	33	0		25	6	
Non-Default Group(1)	24	3		16	13	
Total	57	3	59	41	19	62

Of particular interest is the pattern of errors. When the cases that were misclassified in the validation set by each procedure are examined critically, it is found that some overlap. Sixteen cases were misclassified the same by both procedures. All the 16 were positive that were classified as negative. In addition, logistic regression misclassified six negatives as positive that the discriminant analysis classified properly. Thus there is a clear difference in the types of cases misclassified by the two procedures. The discriminant functions consistently misclassify many more RCBs into the 0 group than the logistic function. These associations may be studied by inspection of the equations estimated. Tables 2 and 3 present the results of the Discriminant Analysis and the Logistic Regression models of the Financial Sustainability of RCBs. The estimated functions are for logistic regression:

$$LR(X) = 0.547 - 8.61X_1 - 4.32X_2 - 1.98X_3 - 5.23X_4 - 8.82X_5 + 6.13X_6 \qquad (8)$$

And for discriminant analysis:

$$DA(X) = 0.683 - 6.92X_1 - 3.02X_2 - 2.07X_3 - 4.87X_4 - 8.09X_5 + 6.91X_6 \qquad (9)$$

where X_1 is Inflation rate, X_2 is Interest rate, X_3 is Portfolio at risk, X_4 is Operating expense /asset ratio, X_5 is Debt/equity ratio and X_6 is Deposits to total assets.

As we might have expected from the earlier analyses, the two functions are quite similar.

Tables 2. *Discriminant Analysis of the determinants of Financial Sustainability of RCBs.*

Variable	Parameter Estimate	Standard Error	t- Value	Pr > |t|
Intercept	0.547	0.03811	13.77	<.0001
Inflation rate(x_1)	-8.61	0.00073668	-9.96	<.0001
Interest rate (x_2)	-4.32	0.00083207	-3.87	<,0001
Portfolio at risk(x_3)	-1.98	0.03314	-6.89	<.0001
Operating expense /asset ratio (x_4)	-5.23	0.02389	8.11	<.0001
Debt/equity ratio (x_5),	-8.82	0.02474	18.76	<.0001
Deposits to total assets(x_6)	6.13	0.08431	-6.6	<.0001

Tables 3. *Logistic Regression Analysis of the determinants of Financial Sustainability of RCBs Logistic Regression*

Variable	Parameter Estimate	Standard Error	t- Value	Pr > |t|
Intercept	0.683	0.03811	13.77	<.0001
Inflation rate(x_1)	-6.92	0.00073668	-9.96	<.0001
Interest rate (x_2)	-3.02	0.00083207	-3.87	<,0001
Portfolio at risk (x_3)	-2.07	0.03314	-6.89	<.0001
Operating expense /asset ratio (x_4)	-4.87	0.02389	8.11	<.0001
Debt/equity ratio (x_5),	-8.09	0.02474	18.76	<.0001
Deposits to total assets (x_6)	6.91	0.08431	-6.6	<.0001

Inflation rate(x_1) is included in the macroeconomic variables to show how the economic situation might affect the sustainability of RCBs. In this study inflation rate has a value of -8. 61 from LR model and -6.92 from LDA model which are all statistically significant at a 5% level. This is highly expected because the inflation rate is used to calculate the RCBs cost of capital that lowers the financial sustainability. Therefore it is clear that RCBs operating in a low inflation country are more successful in becoming self-sustainable whilst RCBs in high inflated countries find it more difficult. The main reason is the erosion of RCB's equities due to inflation as higher rates of inflation results in large part of equities being lost.

The coefficient of interest rate (x_2), (-4.32 and -3.02 from

LR and LDA models respectively), is highly significant at 5% level. The relationship between the interest rate (x_2) and repayment rate is of particular interest. Interest rates are related to the price of capital. The lower the interest rate, the higher the repayment rates of borrowers. The implication is that a higher interest rate increases cost function, which affects their level of markup and thereby reduces the ability to repay borrowed fund. It also implies that that the high interest rate in a country negatively affects the sustainability of the local RCBs.

The coefficient of Portfolio at risk(x_3) is -1.98 and -2.07 from LR and DA models respectively. This variable is significant at a 5% level. As the portfolio at risk (PAR) indicates the portion of portfolio which is at risk of defaulting is clear that a low value will enhance the possibility for a RCB to be sustainable in the long run. The high debt ratio or leverage allows the RCB to be more profitable, thus sustainable and to reach a greater clientele base [32, 34].

The coefficient of operating expense /asset ratio (x_4) is -5.23 and -4.87 from LR and DA models respectively. The result is statistically significant at the 5% level and implies that a decrease (an increase) in this variable increases (decreases) the financial sustainability of RCBs. This result collaborates the finding of [32, 34] that poor expenses management to be among the main contributors to poor financial institutions' profitability.

The coefficient of debt/equity ratio (x_5), (-8.82 and -8.09 from LR and DA models respectively), is statistically significant at 5% level. This may be due to the fact that RCDs in Ghana do not pay dividends and this makes equity a relatively cheap source of finance compared to debt financing. A number of studies provide empirical evidence supporting this negative relationship between debt level and firm's performance or profitability [34].

The variable, Deposits to total assets(x_6) has a positive coefficient of 6.13 from the LR model and 6.91 from the LDA model. This value is significant at a 5% level. This positive value is associated with the benefit that institutions and banks have from having lots of deposits from the public. The financial benefit comes from the fact that the interest rate paid on deposits is always cheaper than borrowing from other institutions and at the same time deposits mobilization can release RCBs from their dependence on donor funds, government subsidies and external credit.

Table 4 presents sensitivity and specificity of both approaches at various cutoffs of the probability of having any record of financially sustainable. Table 4 displays the accuracy measures from the models with adjustment for endogeneity, standardized to the overall distribution of the other covariates in the model among default cases for sensitivity and the distribution among non-default cases for specificity. Sensitivity is estimated to be slightly lower and specificity is slightly higher from the LR model compared to the LDA model. The estimated sensitivity of default is 86.0% (95% CI = 83.4% to 88.1%) from the LR model and 86.5% (95% CI = 84.3% to 89.1%) from the LDA model. The estimated specificity is 88.1% (95% CI = 87.1% to 89.1%) from the LR model and 87.3% (95% CI = 86.1% to 88.5%) from the LDA model.

Table 4. Sensitivity and specificity of logistic regression and discriminant analysis models, at various cutoff points for the probability of sustainability.

*Cutoff Values	Discriminant Analysis		Logistic Regression Analysis	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
0.05	94.9	8.3	100	25.1
0.10	92.3	23.3	100	25.8
0.25	69	69.2	92.3	37.1
0.50	28.2	95.8	71.8	70.4
0.75	5.1	100	25.6	78
0.90	0	100	5.1	76.8

P*(default)=values less than or equal to the cut-off value indicate that the RCB is not sustainable; those greater than the cut-off value indicate that the RCB is sustainable.

6. Conclusion

The objective of this paper was to use linear discriminant analysis (LDA) and logistic regression (LR) models to examine the determinants of rural and community banks' financial sustainability. These are two of the most widely used statistical methods for analyzing categorical outcome variables. The models were used to classify RCBs as sustainable and not sustainable. While both are appropriate for the development of linear classification models, linear discriminant analysis makes more assumptions about the underlying data. Hence, it is assumed that logistic regression is the more flexible and more robust method in case of violations of these assumptions.

In general, both models produced similar results. Both methods estimated the same statistical significant coefficients, with similar effect size and direction, although logistic regression estimated larger coefficients.

The overall classification rate for both models was good, and either can be helpful in predicting the possibility of the sustainability of RCBs. Logistic regression slightly exceeds discriminant function in the correct classification rate but the differences in the AUC were negligible, thus indicating no discriminating difference between the models. Ultimately, the choice of analysis method will depend on the particular characteristics of the application, including the plausibility of required assumptions and computational convenience.

References

[1]   Agresti, Alan (1996). An Introduction to Categorical Data Analysis. NY: John Wiley.

[2]   Ajai, N and Azeb, F. (2010). Rural Banking: The Case of Rural and Community Banks in Ghana. Agriculture and Rural Development Discussion Paper 48.

[3]   Altman, E. I.; R. Eisenbeis, (1987) "Financial Applications of Discriminant Analysis: A Clarification," Journal of Financial and Quantitative Analysis

[4]   Aryeetey, Ernest. 1996. "Rural Finance in Africa: Institutional Developments and Access for the Poor," in The Annual World Bank Conference on Development Economics 1996, edited by M. Bruno and B. Pleskovic. The World Bank, Washington, DC: 149-188.

[5]   Baron, A. E. (1991). Misclassification among methods used for multiple group discrimination: The effects of distributional properties. Statistics in Medicine, 10, 757-766.

[6]   Bayne, C. K., Beauchamp, J. J., Kane, V. E., and McCabe, G. P. (1983). Assessment of Fisher and logistic linear and quadratic discrimination models. Computational Statistics and Data Analysis, 1, 257-273.

[7]   Bowman, Woods, "Financial Capacity and Sustainability of Ordinary Nonprofits," Nonprofit Management and Leadership, Vol. 22, No. 1, Fall 2011, pp. 37–51.

[8]   Cleary, P. D., and Angel, R. (1984). The analysis of relationships involving dichotomous dependent variables. Journal of Health and Social Behavior, 25, 334-348.

[9]   Cox, D. R., and Snell, E. J. (1989). The analysis of binary data (2nd Ed.). London: Chapman and Hall Dattalo, P. (1994). A comparison of discriminant analysis and logistic regression. Journal of Social Service Research, 19, 121-144.

[10]   Dey, E. L., and Astin, A. W. (1993). Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. Research in Higher Education, 34, 569-581.

[11]   Dowling, L, (2011): The Indian Microfinance Institutions (Development and Regulation) Bill of 2011: Microfinance Beginnings and Crisis and How the Indian Government is Trying to Protect Its People... International Lawyer, Winter2011, Vol. 45 Issue 4, p1083-1091, 9p.

[12]   Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. Journal of the American Statistical Association, 70, 892-898.

[13]   Essel, Thomas T. 1996. "The Credit Lending Operations of Rural Banks and Their Impact on Rural Development: a Case Study of Kakum Rural Bank" (Unpublished M. Phil. thesis on file at the Department of Development Studies, University of Cape Coast, Cape Coast, Ghana).

[14]   Fraser, M. W., Jensen, J. M., Kiefer, D., and Popuang, C. (1994). Statistical methods for the analysis of critical events. Social Work Research, 18(3), 163-177.

[15]   Hackler, Darrene, and Gregory D. Saxton, "The Strategic Use of Information Technology by Nonprofit Organizations: Increasing Capacity and Untapped Potential," Public Administration Review, Vol. 67, No. 3, 2007, pp. 474–487.

[16]   Harrell, F. E., Jr., and Lee, K. L. (1985). A comparison of the discrimination of discriminant analysis and logistic regression under multivariate normality. In P. K.

[17]   Sen (Ed.), Biostatistics: Statistics in biomedical, public health and environmental sciences (pp. 333-343). Amsterdam: North Holland.

[18]   Johnson, B., and Seshia, S. S. (1992). Discriminant analysis when all variables are ordered. Statistics in Medicine, 11, 1023-1032.

[19]   Knoke, J. D. (1982). Discriminant analysis with discrete and continuous variables. Biometrics. Vol. 38, pp. 191-200.

[20]   Meshbane, A., and Morris, J. D. (1996, April). Predictive discriminant analysis versus logistic regression in two-group classification problems. Paper presented at the annual meeting of the American Educational Research Association, New York. (ERIC

[21]   Documentation Reproduction Services No. ED 400 280) Press, S. J. and Wilson, S. (1978). Choosing between logistic and discriminant analysis. Journal of American Statistical Association. Vol. 73, pp. 699-705.

[22]   Imam Ghozali (2006). Aplikasi Analisa Multivariate Dengan Program SPSS. Badan Penerbit Universitas Diponegoro, Semarang Indonesia.

[23]   Johnson R. A., and Wichern, D. W. (1988). Applied multivariate statistical analysis (2nd Ed.).

[24]   Englewood Cliffs, NJ: Prentice Hall.

[25]   Kevin Keasey and Robert Watson (1986), "Current Cost Accounting and the Prediction of Small Company Performance" Journal of Business Finance & Accounting. Oxford: Vol. 13, Issue. 1; pg. 51.

[26]   Muhammad Rubini Kertapati, Nuradli Ridzwan Shah and Abdul Hadi, (2004) "Evaluating Company's Performances Using Multiple Discriminant Analysis" UIBMC Hyatt Hotel Kuantan.

[27]   Neter, J., Wasserman, W., and Kutner, M. H. (1989). Applied linear regression models (2nd Ed.). Boston: Irwin.

[28]   Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd Ed.). Mahwah, NJ: Erlbaum.

[29]   Tabachnick, B. G., and Fidell, L. S. (1996). Using multivariate statistics (3rd Ed.). New York: HarperCollins.

[30]   Toshiyuki Sueyoshi and Shiuh-San Hwang (2004) "A Use of Nonparametric Tests for DEA-Discriminant Analysis: A Methodological Comparison" Asia-Pacific Journal of Operational Research, Vol. 21, No. 2 179-195.

[31]   Wilson, R. L., and Hargrave, B. C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. Educational and Psychological Measurement, 55, 186–195.

[32]   Yeboah, Abor. 1994. "Can Ghana's Rural Banks Be Used to Stimulate Peasant Agriculture" The Case of the Kaaseman Rural Bank and its Cocoa-farming Customers' (Unpublished B. Com. thesis on file at Department of Economics & Business Studies, University of Cape Coast, Cape Coast, Ghana).

[33] Cowart, T. W., Lirely, R. and Avery, S. (2014) Two Methodologies for Predicting Patent Litigation Outcomes: Logistic Regression Versus Classification Trees. American Business Law Journal. 51, 843-877.

[34] Ben Bouallègue, Z., 2013: Calibrated short-range ensemble pre-cipitation forecasts using extended logistic regression with interaction terms. Weather Forecasting, 28, 515–524.

[35] Jakob W. Messner, Georg J. Mayr, Achim Zeileis, and Daniel S. Wilks, 2014: Heteroscedastic Extended Logistic Regression for Postprocessing of Ensemble Guidance. *Mon. Wea. Rev.*, 142, 448–456. doi: http://dx.doi.org/10.1175/MWR-D-13-00271.1.