RESEARCH ARTICLE

SOCIOECONOMIC AND DEMOGRAPHIC INFLUENCE ON MICROFINANCE LOAN DEFAULT RATE

Henry Ajagbawa
Managing Partner, Henry O. Ajagbawa & Associates

ABSTRACT

The purpose of the study is to define a relationship between microfinance loan default rate (risk of loan repayment) and certain key socioeconomic and demographic variables. By conventional thinking, we are prone to assume that loan default rate in microfinance institution is likely to outpace the default rate in commercial banks. This assumption rests on the less stringent minimum risk acceptance criteria in lending, and the traditional in-built weaker control mechanisms and lax lending practices in microfinance institutions to accommodate the less privilege and promote financial inclusion. The socioeconomic and demographic profile of the average customer of the microfinance institution is markedly different from the clients of the commercial banks. Surprisingly, the early experience of the modern Father of microfinance, Prof. Muhammad Yunus, indicates that microfinance default rate is statistically insignificant, and repayment culture may vary from one country to the other as influenced by endogenous factors or exogenous factors like government interference. The paper explores whether default rate can be explained by the endogenous factors (independent variables) in a selected microfinance institution in Nigeria. Advanced Multivariate statistical methods of factor analysis and discriminant analysis are employed to facilitate the handling of the multiple socioeconomic and demographic factors that might influence microfinance loan default rate.

INTRODUCTION

Microfinance has become a fashionable buzz word as an alternative to conventional banking in addressing the banking needs of the poor and the huge under-banked population of the world especially in less developed countries (LDCs). Statistics shows that over 80% of the bankable populations of most LDCs are accommodated by the informal banking arrangements. In Sub-Saharan Africa only 12% of her household maintain bank accounts compared to over 91% in OECD (organization for economic cooperation and development) countries. (GCAP-World Bank, 2011). Consequently, the huge un-banked population find expression in the informal banking arrangements like the local thrifts societies. In Nigeria for example there abound a local thrifts arrangement called ESUSU which involves a trusted community person doing the rounds daily and taking collections from the community members, mostly traders and artisans, who at specified intervals and in rounds, collect the total pooled takings until everyone takes his turn. Of course there are variants to these arrangements which would appear to have served the unbanked markets well so far.

It is arrangements like this and other forms of community banking that has approximated to Microfinance institutions or are working in tandem and close cooperation with microfinance institutions. It is not uncommon to see microfinance institutions employ the services of ESUSU to help in mobilizing deposits and selling loans or down right recruit young sales agents for the same purpose.

For Example, the Noble Laureate award winning Grameen Bank, on its website, categorized the ubiquitous terminology of MFI into 10 categories ranging from the typical money lender (Esusu) to co-operatives, NGOs. Microfinance and other rural credit schemes organized through specialized banking called microfinance Bank. In all, one central theme is the provision of financial intermediation services to the poor. The by-line of Grameen Bank, typifies the very essence of Micro Finance Institutions; “Bank for the Poor”

Microfinance business is operated in well over 100 countries and over 12,000 MFB are in operation around the world (Forbes, 2011) and purported to host over 170 million active customers (Prof Yunus on CNN on Feb 16, 2011). As at 2009, gross loan portfolio of MFI’s stood at $65.2 billion with an average loan per customer at about $521.

In Nigeria, the enabling regulation, guiding the business of microfinance banks came into existence in 2006 by the issuance of a Regulatory and supervisory Framework for microfinance banks by the Central Bank of Nigeria (CBN), and with this document, the regulation and supervision of the business of microfinance became codified. The importance of this development is underscored by the fact that significant level of influence over macroeconomic variables especially through monetary policies was not within the ambit of the regulators as a huge part (80%) of banking business was done outside the banking sector. A significant part of the informal banking system is now being captured and measured in the
formal sector thereby providing more accurate monetary policy measure for macroeconomic planning and control.

Given the level of literacy/numeracy in a LDC like Nigeria, an effective microfinance model can only be achieved if certain base characterization of microfinance bank is sold to the public to be able to attract the teeming rural and poor unbanked and under banked Nigerians into the new banking model. Experientially, the perception of most underprivileged and unbanked Nigerians about commercial banks is that banking is for the affluent and are not encouraged to initiate banking relationships given the cultural and economic distance they perceive between them and the bankers. Microfinance banks operations on the other hand are built on “simplicity” and characterized by attributes such as No collateral required for Lending (typically), Simple process for account opening and operation (most account holders are barely literate), KYC is community based (as MFI are local based and entrenched in the community they operate), and No legal instrument required for lending (at best, very basic)

The sustainability of MFB as a business model is of concern to stakeholders given the inherent weakness of the model. Concerns over high transaction costs, Risk allocation, weak governance, Information constraints, Weak capacity – financial and human and the inherent low revenue streams are issues on the front burner for debate on how best to avoid a potential implosion of the financial model. In Nigeria the Regulation and continuous dialogue on MFI is intended to assure sustainability given the importance of MFIs in the economy especially in intermediating the deficit sector as a poverty alleviation policy mechanism.

There are over 800 licensed micro finance institutions in Nigeria and by regulation, they are all licensed and regulated by the Central Bank of Nigeria. The requirement for their establishment is pretty liberal as minimum paid up capital is N20m ($125,000) compared to minimum paid up capital of the least category of commercial bank at N10b ($62.5m). There are however categories of microfinance banks, starting from the local government category with N20m capital base to the national microfinance category with a capital requirement of N2b. ($12.5m). By law, local government microfinance are expected, at this minimum paid up capital, to operate within a local government area (a borough) and must meet another level of minimum paid up capital, additional branch and other physical resource requirements if they wish to expand their operations beyond their current location.

Intrinsically therefore, MFI in Nigeria are, by design of law established to cater for the local unbanked needs of the rural areas. Operationally, they are cut out to easily meet the peculiar needs of the rural markets than the conventional banks owing to differences in socioeconomic, demographic and other distinguishing profile differences. Because of the ease of set up and very liberal start-up requirement, micro finance institutions have proliferated across the country in staggering number to over 800 within just 5 years whereas the conventional banks are only 24 and threatening to reduce further through combinations. The massive branch net-work of the regular commercial banks does not in any way undermine this point as each micro finance institution operates as a one stop shop legal and economic unitwhere all client needs are accommodated without referrals to the so called head office as in commercial banks. Compared against the bureaucratic and control centered operational framework of the commercial banks, where decisions are majorly concentrated at the headquarters, the MFB are able to better meet the peculiar needs of their local operating environment. The ease of reach, simplicity of its operations, the minimal legal documentation and generous collateral requirements make accessing loans by the under-banked relatively easier. To this extent micro finance institution could be said to be fulfilling one of the core objectives associated to justifying its very liberal standards of operations- meeting the needs of the poor who otherwise may not be able to access financial support.

Loan default is a natural consequence of banking practice which rate defers from country to country but anomalously a default rate of 5% is generally considered appropriate in the short run (Federal Reserve, 2011). In the wake of the global economic meltdown, there was an escalation of default rate in conventional commercial banks reaching an all high of 15% to 20% in some banks. Whereas, Microfinance banks continue to record the lowest default rate of below 2% (Uko, 2011) of its risk asset portfolio. One of the motivation for the research is to understand what consideration explains the distance between the default rate in conventional commercial banks and micro finance banks. Applying the socio economic and demographic characterization of the variable cases in microfinance bank, we attempt to understand the factors that best describes default and similar factors may provide some explanations on default in commercial banks should those characterization that explains default in micro finance banks are similarly prevalent in commercial banks.

Context of the Study

By intuition and to some extent, experiential knowledge borne out of an oversight responsibility over the operations of a MFI; we had listed a core of possible socioeconomic and demographic variables that are likely to feature prominently in providing explanation for the question raised in this study. We had attempted to understand what variables are likely to count in explaining the rate of default/Risk of non-payment of a microfinance loan. 10 factors were listed a priori; Size of loan, Age, Sex, number of years in current business, number of years in current business location, type of business, family size, Educational level, household income/profit per cycle, and marital status.

In essence the rigorous search for relevant literature on the study of probable relationship between Risk of repayment of a microfinance loan (default) and identified socio-economic and demographic variables earlier identified in this paper does not seem to have been addressed in existing studies. Consequently it would appear that this is an entirely green field for study. However, in isolated cases, a uni-variable study has been conducted. Refer to the study of intellectual capital to financial performance (African journal of Accounting, finance & Banking research, 2010, vol. 6, page 17-31). In the study a zero order correlation was found between intellectual capital (education) and financial performance. There were other Gender related studies (impact of women) which tend to highlight the impact of women on microfinance.

Our literature review neither established the existence of a study of the influence of socioeconomic and demographic factors on microfinance default rate nor did it reveal the application of Multivariate statistical analysis tool to evaluate simultaneous impact of these variables on the default of microfinance loans. Therefore, our study is intended to add to the body of knowledge in the area of the application of multivariate statistical tools on the influence of socioeconomic and demographic factors on microfinance loan default rate.

**Conceptual Framework**

We have postulated a probable hypothesis that the rate of default of a microfinance loan is affected by a list of socioeconomic and demographic variables including; size of loan, Gender of obligor, Educational level, Number of years in current business, industry type, Household income, family size, Number of years in current location, Marital status, and Age. These 10 items constitute the predictor variables in our model while the loan default rate is the dependent variable. Our hypothesis is;

**Null Hypothesis**

\[ H_0: \ D = F (X_n) \]; where D is default and \( X_n \) is the identified socioeconomic and demographic variables

**Alternative Hypothesis**

\[ H_1: \ D = F (Y_n) \]; where D is default and \( Y_n \) is other variables outside the observed variables

Because of the large size of the predictor variables and the variable cases. There is a necessity to employ more advanced statistical tools to handle the data set. But first we need to make a brief note on the data source and treatment.

Data Source: Data was spooled out of a customer data base of a microfinance Institution. The MFI has a customer base of over 10,000 customers but with only 2080 loan customers active in its books. We obtained at random a sample of 500 clients from the population of 2080 loan customers of the MFI. The data base was not well updated. We uploaded the information gaps to the data base for update. Subsequently, we spooled out secondary data from the updated database of the MFI

Description of the Variables: Each case (client) is described by the 10 predictor variables. Each of the variables or attributes (characteristics) has a unique measurement scale. Some are numeric while others are not. We have used dummy scales to measure the nonmetric variables. Table 1 depicts the measurement scale of each of the variables and the ascribed dummy variables to each.

The dependent variable, loan default rate, is a nonnumeric nominal dichotomous variable. The state either is default or non-default and so we can simply represent the variable on the binary scale. Therefore we have 10 independent variables and a dependent dichotomous variable which can only be handled by multivariate statistical methods. The default state is represented as 1; while non-default is at 0.

**The analysis**

The objective of the paper is to define a relationship between certain socioeconomic and demographic variables, and the loan default rate in a microfinance bank. The data set we have obtained for the study presents us with a dependent variable that is categorical and dichotomous and so amenable to the statistical analytical tool of discriminant analysis. This would assist in the explanation of the behavior of the group membership and how each of the extracted factor best explains the membership of each variable to the two groups; default and non-default customers. In this manner, we would establish a relationship between the default and the extracted variables. The choice of Discriminant analysis is best suited to the extent that the dependent variable is a categorical variable amongst other reasons including dealing with multiple independent variables.

However, the sheer number of the independent variables is likely to crowd out the relative importance of some of the more potent variables in explaining the states of default or non-default and so we will employ the analytical tool of factor analysis to reduce the variables to a more manageable size. Factor analysis would facilitate the weeding out of variables that are too linearly correlated and provide poor explanation for the behavior of the dependent variable. It helps in achieving data reduction by extracting variables with more explanatory relevance from the array of variables in the model.

A set of descriptive statistics would be employed to provide some sense making in understanding the character of the predictor variables. This would establish frequency statistics, mean, median, mode, percentiles, range standard deviation, variance, skewness, and kurtosis. In some manner it would show us if the possible existence of outliers can be tolerated or not and whether it would materially affect the use of the data as is or some level of de-noising is required before the data can be used.

We employed the IBM SPSS 19 package in executing the analysis for the research

**Factor Analysis**

We have identified ten potentially and equally probable variables that may best describe behavior of default or non-default.
Factor analysis is a tool that helps in data reduction and structures detection. In this section our aim is to utilize factor analysis to reduce the variables to a manageable size that best describes the phenomenon under study in a manner that only statistical insignificant data is sacrificed in the process of the data reduction. Essentially we are employing explanatory factor analysis in this session to the extent that we are exploring the strength of variables in explaining the concept under review. And the R-type factor analysis is used here as a clear variable type-client is under review.

In our research design, as captured under the section Description of the Variables, we listed 10 independent variables and stated the measurement type and scale. The population size is 2080 and a sample size of 500 customers, spooled randomly from the population. In our variable selection, we could only obtain variables for which information exist in the database or for which we could upload and update the database. There are some other relevant and perhaps impactful variables like the debt/income ratio which was not captured by the organization as rigorous risk asset criteria were not applied in granting loans to clients in the MFB. However, the absence of this variable in our opinion, and perhaps a few that might exist, does not affect the research effort in any material manner and is consequently unlikely to affect the results to be obtained. Experiences in local financial markets may also support this position. Indeed, it is quite paradoxical that in most LDC, it is more common to find the more affluent defaulting in loan repayment than the less affluent.

### Interpretation of the Results of Factor Analysis

Communalities is defined as the total amount of variance an original variable shares with all other variables included in the analysis. That is the amount of variance in each variable that is accounted for. Through an iteration process, it finds a clear combination of variables that explains the variation in the original variable and repeats the process over and over without affecting the previous component explaining the initial variation. By this method we are able to ascertain the relative variances of each variable included in the analysis.

The higher the variance explained in the variable, the more relevant the variable would be in explaining the dependent variable. The size of the communalities in the factor matrices represents an index for measuring or assessing how much of the variance in the particular variable is accounted for in the factor solution. The higher the communalities, the higher the variance extracted from the factor solution and hence the higher the explanation of the variables by the factor components. Usually we select extracted communalities associated with a variable with high value for further consideration.

Typically the communalities table as in Table 3, list out the initial and extracted communalities for all the variables and first hand you can see which variables have the highest communalities and therefore variance. Low communalities also explains some level of co-linearity amongst the

<table>
<thead>
<tr>
<th>S/N</th>
<th>Name of Variable</th>
<th>Type of Data</th>
<th>Measurement Method</th>
<th>Measurement scale</th>
<th>Description of the Basis of Measurement of the Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Size Of Loan</td>
<td>Numeric</td>
<td>Scale</td>
<td>Actual</td>
<td>The actual loan size is used. N500, 000.00 is the highest loan volume obtainable and obtained. There are however, a few outliers as would be shown in our descriptive statistics</td>
</tr>
<tr>
<td>2</td>
<td>Type Of Business</td>
<td>Non-numeric</td>
<td>Ordinal</td>
<td>0-11</td>
<td>We narrowed down the possible businesses under this business scale to 12 types and denote each by a number ranging from 0-11. For example, artisans may be represented by 1 and Trading by 2……….</td>
</tr>
<tr>
<td>3</td>
<td>No of years in current business</td>
<td>Numeric</td>
<td>Scale</td>
<td>Actual</td>
<td>This variable also describes some level of stability and success of the business. The more the number of years spent in that business the more the clients is perceived to be successful in that business. We have used the actual number of years spent in the business as our measurement basis.</td>
</tr>
<tr>
<td>4</td>
<td>Number of years in current Location</td>
<td>Numeric</td>
<td>Scale</td>
<td>Actual</td>
<td>This variable tends to describe some level of stability or success of the business. The longer the years in the current business location, the more stable the business is perceived. We have adopted the scale of 0-2 as a basis of measurement of the years spent in a business location with each number denoting incrementally longer number of years spent by the business in the location.</td>
</tr>
<tr>
<td>5</td>
<td>Family size</td>
<td>Numeric</td>
<td>Scale</td>
<td>0-2</td>
<td>This scale denotes the number of persons in the family counted as the couple and their children. The scale 0-2 denotes incrementally, the number of the family members in each client family.</td>
</tr>
<tr>
<td>6</td>
<td>Age</td>
<td>Numeric</td>
<td>Scale</td>
<td>Actual</td>
<td>The actual chronological age of the client has been used as the measurement basis.</td>
</tr>
<tr>
<td>7</td>
<td>Gender</td>
<td>Non-numeric</td>
<td>Nominal</td>
<td>Binary</td>
<td>The binary scale; 0 = male and 1 = female; represents the gender of the owner of the business.</td>
</tr>
<tr>
<td>8</td>
<td>Educational Level Household Income</td>
<td>Non-numeric</td>
<td>Ordinal</td>
<td>0-3</td>
<td>Educational Level can at best be descriptive. The scale 0-3 denotes incrementally the level of educational attainment of the owner of the business.</td>
</tr>
<tr>
<td>9</td>
<td>Marital Status</td>
<td>Non-numeric</td>
<td>Nominal</td>
<td>Binary</td>
<td>The binary scale =Married and =unmarried. Other states such as divorce, widowed which may exist is statistically insignificant and so not material to our analysis. In any event, the state of unmarried should adequately summate the other un captured states</td>
</tr>
</tbody>
</table>

These variables were established a priori from our simple understanding of probable influencers of default in a micro finance institution. It is however most improbable that very meaningful results can be obtained from the data in their present state.

In our research design, as captured under the section Description of the Variables, we listed 10 independent variables and stated the measurement type and scale. The population size is 2080 and a sample size of 500 customers, spooled randomly from the population. In our variable selection, we could only obtain variables for which information exist in the database or for which we could upload and update the database. There are some other relevant and perhaps impactful variables like the debt/income ratio which was not captured by the organization as rigorous risk asset criteria were not applied in granting loans to clients in the MFB. However, the absence of this variable in our opinion, and perhaps a few that might exist, does not affect the research effort in any material manner and is consequently unlikely to affect the results to be obtained. Experiences in local financial markets may also support this position. Indeed, it is quite paradoxical that in most LDC, it is more common to find the more affluent defaulting in loan repayment than the less affluent.
Table 2 Interpretation of the Descriptive Statistics

<table>
<thead>
<tr>
<th>S/N</th>
<th>Name of variable</th>
<th>Elements of the measurement Variable</th>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Size of Loan</td>
<td>Actual</td>
<td>The mean and median at N274, 695 and N250, 000, range at N484, 738, and mode at N500, 000. Given a few outlier above the mode, a large number of loans smaller than the mean exist thus giving a huge standard deviation of N154, 029. Less than 5% of the sample is below N50, 000 while the world average loan size is N80, 000. The mean does not do a good job at describing the loan size of our sample population. Given the even spread of the data around the mean and existence of extremely few outliers, we have not done any repairs on the data</td>
</tr>
<tr>
<td>2</td>
<td>Type of Business</td>
<td>Householdmaterial=0;artisan=1;transport =2;telecomms=3;agric=4;generalservices =5;education=6;food&amp;edible=7;clothing=8;prod/building mat=9;Electronics/Autoparts=10;general trading=11</td>
<td>4 out of the 12 defined businesses describes our sample population up to 78%. Household is 9.8%, Artisans is 21.2%, food is 13.8%, and general trading is 20.2% making up cumulative frequency distribution of 78% of the total sample businesses. This descriptive statistics should be compared against the results of the other statistics to draw relevant conclusions as to whether the typical business funded by MFB has bearing to the default rates in the business of MFB.</td>
</tr>
<tr>
<td>3</td>
<td>No of years in current business</td>
<td>Actual</td>
<td>The mean, median and mode are at 9.0640 years, 8years and 10 years respectively. The range of 31 years corroborates the high standard deviation at 5.60023years, indicating the central measures don’t do a good explanatory work of the years spent in the business. At 95% percentile, the sample population indicates a range at 31years, which shows that less than 5% of the customers are over and above 31 years’ experience in business.</td>
</tr>
<tr>
<td>4</td>
<td>Number of years in current Location</td>
<td>Less than 2 years = 0</td>
<td>The variable is almost uni-modal with median and mode at 1year and the mean at 1.2485 years. And with a standard deviation at 0.83245years and range at 2years respectively, the variability from the central tendency is pretty high and so the descriptive also do not explain well enough the sample population</td>
</tr>
<tr>
<td>5</td>
<td>Family size</td>
<td>3-4 = 0; 5-6 = 1; 7&gt; = 2</td>
<td>The sample population indicates a mean of 0.5853 on a measurement scale of 0-2 which translates to an average family size of 5 persons. However 50% of the sample population has family size at level 4 persons using the upper limits of the interval classification.</td>
</tr>
<tr>
<td>6</td>
<td>Age</td>
<td>Actual</td>
<td>The mean, median and mode age is 38.43 yrs. 37yrs. and 35yrs respectively. The standard deviation is 8.88yrs. The distribution is fairly normal and a good representation of the sample population. However the range at 69 years shows from the composite statistics indicates the presence of outliers, however not significant enough to distort the distribution</td>
</tr>
<tr>
<td>7</td>
<td>Gender</td>
<td>Female = 0; Male = 1</td>
<td>The descriptive for this variable may be somewhat meaningless. The average of a man and woman cannot tell us anything. Here in lies the dilemma of statistics. While computation can be aided by statistics the judgment call is about the most important aspect in sense making.</td>
</tr>
<tr>
<td>8</td>
<td>Educational Level</td>
<td>Pry Education &amp; below = 0</td>
<td>At a mean of 2.3414 on a measurement scale of 0-3, it shows that there is a pretty high level of education amongst the micro finance customers. The mode and median at 3 respectively shows a there are more people with higher education than the mean value. The percentile report shows that only 10% of the sample population have education less than level 2.</td>
</tr>
<tr>
<td>9</td>
<td>Household Income</td>
<td>Actual Profit per cycle for the relevant period</td>
<td>The statistics on household income shows that there are definite outliers which would render the statistics almost meaningless if used in its pure state. The mean for example at N10,359n while the median and mode are at N150,000 and N100,000 respectively couple with a range at N5b could mean the presence of unjustifiable outliers or significant transcription error that must be corrected. The minimum would be set at N40, 000 at which point from the percentile report only 10% data loss would result and the maximum set at N1.5m at which point only a maximum of 95% data loss may result</td>
</tr>
<tr>
<td>10</td>
<td>Marital Status</td>
<td></td>
<td>In this context the descriptive statistics on marital status is meaningless as for example there is no meaning to the average of being married and being unmarried</td>
</tr>
</tbody>
</table>

Thus in Table 3, the initial communalities are estimates of the variance in each variable and always 1 for principal component extraction.

Extracted communalities explain estimates of the variance in each variable accounted for by the component. Out of the ten variables, we have chosen only seven with extracted components greater variables than 0.5 as the others are not statistically significant for further consideration. Consequently seven variables extracted would be further considered in the course of the analysis; See the extraction column in Table 3 as the criteria index. The scree plot can be further used to weed out the components that are not good indicators of the variables.

The results shows three sections of the variance explained. The first shows the Eigen values which is the amount of variance in

independent variables which renders them inefficient explanatory variables. The component matrix helps in assessing the relative variance performance of all the variables with respect to the four components extracted as in the scree plot. At the point where the rate of change in the scree plot is low we tend to cut off our components as becoming less efficient.

The un-rotated and rotated factor matrices represent degree of correlation or association of each variable with respect to each factor. The objective of producing each of these matrices is to maximize the association of each variable to a single factor. If our results of the un-rotated component matrix does not appear satisfactory enough a further confirmation may be done to test the adequacy of the results of the un-rotated matrix. In our analysis here the matrix shows how well each of the variables is explained by the four components derived.
the original variables accounted for by each component while the second section shows the percentage of variance column which indicates the proportion of the variance accounted for by each component to the total variance in all the variables. Lastly, the cumulative percentage gives the proportion of variance explained up to the 10th component. In this case we have cumulative variance up to the 4th component. We ordered an Eigen value of 1 in our principal component analysis and so all the Eigen values with value greater than 1 represent principal component from the extracted solution. Consequently the first six components with Eigen value greater than 1 and collectively representing a cumulative explanation of 54.5% of the variance of all the components will constitute the principal component in our analysis.

This implies that 54.5% of the variability of the original 10 variables can be explained by the extracted variables. By reduction to 4 variables from 10, we have at best an average explanation of the variability, with about 45.5% possible loss of information on all the other 6 variables. A reasonable trade off: The Rotated matrix component also helps to explain same. The scree plot provides a schematic approach to efficiently choosing the number of components; from the scree plot, the components 1 through 4 have very sharp slopes while it basically flattens subsequently. We therefore extract component 1 through 4 while the other may contribute very little to the solution individually. The component matrix provides the basis for the selection of the relevant factors based on the highest component scores which corresponds with the variables with the highest variances and so represents better explanation of the mode.

Figure 1

Table 3 Communalities

<table>
<thead>
<tr>
<th>Variables</th>
<th>Initial</th>
<th>Extraction</th>
<th>Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Loan Balance actual</td>
<td>1</td>
<td>0.529</td>
<td>Yes</td>
</tr>
<tr>
<td>2 Gender</td>
<td>1</td>
<td>0.578</td>
<td>Yes</td>
</tr>
<tr>
<td>3 Years in current Location</td>
<td>1</td>
<td>0.329</td>
<td>No</td>
</tr>
<tr>
<td>4 Level of Education</td>
<td>1</td>
<td>0.672</td>
<td>Yes</td>
</tr>
<tr>
<td>5 Family size</td>
<td>1</td>
<td>0.489</td>
<td>No</td>
</tr>
<tr>
<td>6 Age</td>
<td>1</td>
<td>0.599</td>
<td>Yes</td>
</tr>
<tr>
<td>7 Years of experience in Business</td>
<td>1</td>
<td>0.571</td>
<td>Yes</td>
</tr>
<tr>
<td>8 Household Income</td>
<td>1</td>
<td>0.571</td>
<td>Yes</td>
</tr>
<tr>
<td>9 Marital status</td>
<td>1</td>
<td>0.480</td>
<td>No</td>
</tr>
<tr>
<td>10 Type of Business</td>
<td>1</td>
<td>0.579</td>
<td>Yes</td>
</tr>
</tbody>
</table>

We are able to derive from the component matrix the relevant variables with corresponding components scores. By working through each of the components and mapping to all the variables associated to that column we can derive the variable with the highest component score as the variable to be chosen for further analysis and as having greater predictor capacity than the others. The rotated component matrix should be obtained as confirmation of the reliability or adequacy of the results obtained from the un-rotated component matrix. Please see the set of variables thrown up by either of the un-rotated and the rotated component matrix as summarized in Table 4;

Table 4 Summary of (UN) rotated matrix

<table>
<thead>
<tr>
<th>Components</th>
<th>Un-rotated matrix</th>
<th>Rotated matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Age</td>
<td></td>
<td>Years of experience</td>
</tr>
<tr>
<td>2 Loan balance</td>
<td>Household income</td>
<td></td>
</tr>
<tr>
<td>3 Level of Education</td>
<td>Level of Education</td>
<td></td>
</tr>
<tr>
<td>4 Gender</td>
<td></td>
<td>Marital status</td>
</tr>
</tbody>
</table>

Table 4 shows the four variables with explanatory power over 54% of the phenomenon under review are; Years of experience in business, House hold income, Level of Education and finally marital status

Discriminant Analysis

The foregoing section on factor analysis has thrown up four variables as having possible impact in our analysis as defined in table 4. Our next step is to find the differences between the two groups- default and non-default customer and to see which of the identified variables best separate them and the extent of the influence of any of the variable to the different group members. In so doing we will be able to postulate a relationship that best explains the behavior of all variables to the dependent variable and how this behavior is escalated in the different groups.

Discriminant Analysis is used to model the value of a dependent categorical variable based on its relationship to one or more predictor variables (independent variables). In discriminant analysis the dependent variable is dichotomous; two states of either being a default or non-default customer. The two states of the dependent variable is mutually exclusive and exhaustive. The two grouping of the dependent variable would have to be explained by the independent variable so we can understand the differences between the two groups and perhaps the relationship. Understanding the difference of the two groups helps to determine the variables that best describes each group better and the extent to which such variables describes the different groups in comparison to each other.

A discriminant function is formed;

\[ Z = a + W_1X_1 + W_2X_2 + \ldots \ldots + W_nX_n \]

Where \( Z \) is the discriminant Z score of the discriminant function
\( a= \) intercept; \( W= \) discriminant weight for independent variable and \( X= \) independent variable

The discriminant is a variate which gives a summated position of all the weighted independent variables to explain the behavior of the discriminant function and takes the form of equation as above. The discriminant analysis results processed on the IBM SPSS 19 is summarized in table 7;

Explanation Of The Results On Discriminant Analysis

Table 5 shows the summary of group means of each of the state of the dependent variable and their corresponding group
differences. Discriminant analysis is used as a technique in testing the hypothesis that the group means (centroid) of a set of independent variables for two or more groups are equal. By comparing the group means of the two groups for the various variables we settled for, we can see whether the groups have marked differences across each group and across each variable:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Years of Experience in Business</th>
<th>Household Income</th>
<th>Education</th>
<th>Marital Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Group 1</td>
<td></td>
<td></td>
<td>1157</td>
<td>0.0424</td>
</tr>
<tr>
<td>1 Group 2</td>
<td></td>
<td></td>
<td>0.157</td>
<td>2.775</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
<td>0.8095</td>
<td>0.0345</td>
</tr>
</tbody>
</table>

Household income is more pronounced than in Years of experience in Business and that more pronounced than education with the smallest group mean difference occurring in marital status. This implies that Household income is the best explanatory variable of the difference in the group behavior, followed by years of experience and then educational attainment. In that order with marital status having almost the same mean across the group.

A more reliable way to test the relevance of the group difference is to check the level of significance. Typically, any variable with a significance level greater than 0.1 may be statistically insignificant for use. From the Tests of Equality of Group Means we can see that only the variable, Educational attainment has significance level lower than prescribed. The nearest variable that may be considered is Years of experience in Business at 0.179. The Wilks’ Lambda statistics similarly supports our choice of a more potent discriminatory variable in Education and Years of experience. It can be seen that both variables have the smallest coefficient. Consequently we may actually restrict our analysis to the two variables as the others add very little to the analysis.

From the pooled within-group matrices with a diagonal of unity, we can affirm that the only two variables with reasonable level of correlation are level of education and marital status. Since also we see corroborating evidences in the very small group mean difference for marital status as well as the low level of statistical significance it is easy to conclude what variable should be dropped from further analysis. The remaining three variables, education, House hold Income and years of experience in business all have very low level of correlation and so have both reasonable significance level and good group differences worthy of further review. However, only two of the three variables is consistent with both the significance test and the Lambda test and so we may also drop Household income and settle for only the remaining two variables.

The standardized canonical discriminant function coefficient measures the overall model fit. The squaring of the canonical correlation gives the proportion of the dependent variable explained by the independent variable. Thus, the squaring, which gives a coefficient of 0.567, 0.168, 0.157 and 0.0424 (Table 8) corresponds to the variables; education, Years of experience, household income and marital status respectively. What this shows is that the dependent variable is partly measured by education to the power of 56.7%. That is the variability of the dependent variable is measured by this proportion. The level of education therefore, appears very dominant in explaining the behavior of default.

The classification Function coefficient Matrix provides weights that can be assigned to the variable to form the multivariate equation describing the relationship under study. Table 6 summarizes the coefficient for the separate grouping and the discriminant function can thus be formed.

We actually can form a discriminant function from this table in a manner where we take the coefficient that is maximized corresponding to a particular variable. For example the coefficient of level of education is largest under the default category and so it helps to form a discriminant function in the same manner as a simple regression equation is assembled. The summary of the discriminant function is that;

People with higher levels of Education are more likely to default than those with relatively lower level of education. Does this have implication for the behavior of default in commercial banks. Anecdotally, from a practitioner perspective, the more sophisticated, wealthy and enlightened clients are prone to exhibit tendencies to default than the simple average everyday client. Where possible, they seek all legal and commercial loopholes to avoid loan repayment.

Customers with more years of experience in running their business are less likely to default than those with fewer years of experience.

Default is more likely to occur with single (unmarried) customers than with their married opposites and finally.

The variable of household income completely failed to discriminate between both groups and so does not function in the discriminant equation.

CONCLUSION AND RECOMMENDATION

The results are certainly interesting and resonates with common place pedestrian logic. Education for example throws up issues that are worth looking at in the future. One is likely to agree that considering the loan amounts involved, being very small, it is possible that the more educated people may not take the business as serious as the less educated people, who might feel that the opportunity cost is bleaker than their educated counterpart. This implies that, because the more educated ones have more options and opportunities at other jobs, the possibility of abandoning the less lucrative microfinance business once the opportunity comes up is pretty high. Some have only taken the option as a get-away strategy in preparation for better opportunities and are on the way once that opportunity comes knocking.

Surprisingly, the level of household income appeared totally a non-discriminating variable as one would have assumed while the level of significance of the variable; years of experience in
the business cast doubt over the potency of the variable as a discriminating factor amongst the group.

In all, I am of the opinion that more research would produce new insights on the multiple levels of relationships and behavior of the various independent variables in response to the dichotomous dependent variables. In particular, more predictor variables might throw new light on possible pattern of relationship amongst the various variables. The research was in respect of a single MF institution with a national license and customer base across the different geographies of Nigeria. Therefore, we cannot fathom what pattern of relationships are likely to emerge if the analysis were to focus strictly within region in the country or if cross sectional data across various MF institutions were to be used. Further, given that multiple and different behavioral traits of MFB is noticed across regions of the continent, it might be interesting to apply the techniques to cross sectional data across the African sub region to view the possible patterns emerging across the different regions.

References
1. CBN (2013). Micro finance institutions. (2013) regulation and supervisory framework for MFBs in
7. GDRC (2013) Global Development Research Centre. Testing the tools of development: credit
11. http://www.webcrawler.com/info.wbcrlw305.03/search/ web?q=microfinance+loan&cid=126729304&ad.network =s&ad.keyword=microfinance%20loan&ad.creative=236 11594744&ad.position=1t1&ad.placement=&ad.matchty pe=b&ad.aceid=&ad.ismobile=&ad.device=c&ad.device model=&ad.segment=info.wbcrlw305.03www.gdrc.org/i cm/references/microfinance.html
15. Oceanic bank micro finance
19. This day Newspaper (2011) Publication of Sunday, February 13th 2011, page 30. Features in

*******