Consilience: The Journal of Sustainable Development
Vol. 7, Iss. 1 (2012), Pp. 17–32

Building A Credit Scoring Model For The Savings And Credit Mutual Of The Potou Zone (MECZOP)/Senegal

Ousséni Kinda
Cheikh Anta Diop University
ousconnect@gmail.com

Audrey Achonu
Columbia University
aia2110@columbia.edu

Abstract

This article explores the management of risk credit in a savings and credit mutual in a rural community in Senegal. Using a logit model, we built a credit scoring model for this mutual. There are three main models of scoring. The principal advantage of the regression model is that it clearly shows the link between credit risk and its characteristics, and hence, has a strong predictive power. Our findings show that the variables age, age-squared, gender reimbursement history, guarantee and frequency of reimbursement are all statistically significant in regards to their relationship with the probability of repayment in the logit model. This informative credit scoring model developed for MECZOP allows for an assessment of variables that influence significantly late reimbursement; however, it needs to be tested on old clients and new borrowers.

Author's Note

Ousséni Kinda holds a master's degree in Economics and is currently pursuing a Master of Public Administration in Development Practice at Cheikh Anta Diop University. He has participated in field work at the Millennium Villages Project in Potou, Senegal addressing issues of microinsurance and microcredit. He is interested in further conducting research on economic diversification and more generally on issues of sustainable development.

Audrey Achonu is currently pursuing a Master of Public Administration in Development Practice at Columbia University. She recently interned at the Millennium Village Project in Potou, Senegal where the study was conducted. She holds a bachelor's degree in Economics with a minor in Political Science, Audrey hopes to continue working on how to foster economic growth at the macro and micro-level, ensuring that growth occurs at all levels of development.

Keywords: credit scoring model, late reimbursement, savings, credit mutual.

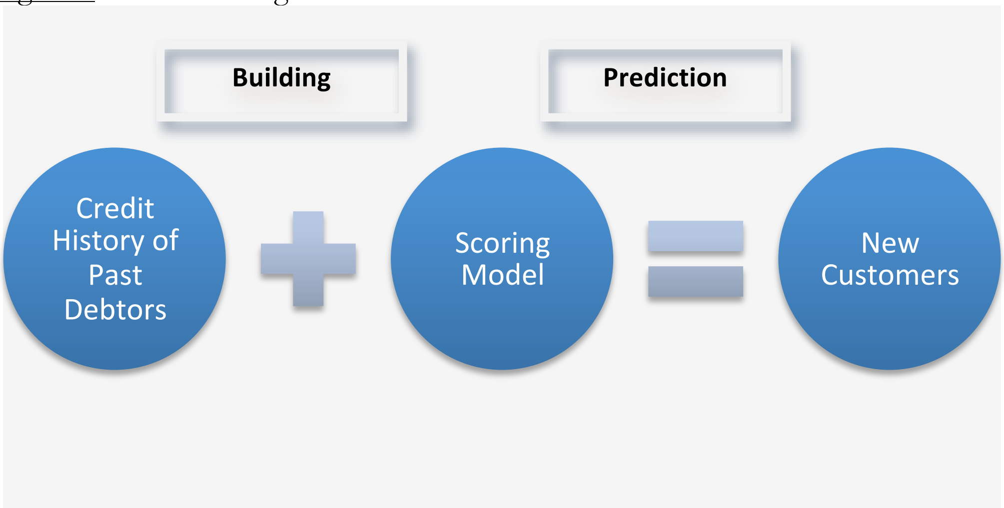
1. Introduction

Sustainable growth in Africa requires facilitating the poor's access to financial resources. Although microfinance in West Africa continues to grow, it makes up only 7-8% of the financial system comprised of loans and savings (CGAP, 2011).

There remain considerable hurdles that affect and prevent the sector from growing, namely poor governance, weak internal control, low capitalization, poor performance and insufficient management of credit risk.

There are many ways to tackle the problem of credit risk, such as through the development of a credit scoring model that will allow for forecasting of the probability that a loan is not reimbursed. In the context of this study we examine the case of the microfinance institution MECZOP[1] in Potou, Senegal and develop a credit scoring model that can be used to limit credit risk.

Figure 1: Credit Scoring Process



Source: Adapted from Liu Yang, New Issues in Credit Scoring Application, 2001

2. Credit-Scoring Model

Scoring can be defined as the use of the knowledge about the performance and characteristics of past loans to predict the performance of future loans (Schreiner, 2004, P.2). It uses a technique that awards scores to borrowers as a means of evaluating the performance of their future loans. The scoring follows the framework presented in figure 1.

By using a borrower's credit history, credit scoring is a tool that supports the decision making process, with a final goal of predicting the probability that a particular borrower defaults.

The implementation of a credit-scoring project follows the following process (Caire et al., 2006):

1. The definition of the segments of scoring: identifying the type of customers and products for which the scoring model will be used;
2. Selection of the type of scorecard: there are three basic methods (Schreiner, 2003; Caire, 2004; Caire et al., 2006; Sur, 2008), the statistical method which is empirically derived from data on past loans; judgmental approach which is structured from expert judgment and institutional experience; and finally the hybrid which combines both judgmental and statistical;
3. The design of the credit scorecard: The 3 D (Definition, Discovery and Development) terminology will define the framework used in this phase.

[1]Savings and Credit Mutual of the Potou Zone.

The financial institution must **define** itself what it considers a "bad" credit behavior (bad client). Bad credit behavior is when the borrower registers delinquency failing to meet his obligations with respect to the owed principle and interest. This is typically a client that if the MFI had perfect hindsight would have avoided (Caire et al., 2006). In the case of the MECZOP a bad borrower is a borrower that has been late in debt repayments by at least 5 days one or more times. The MECZOP starts to calculate penalties after 5[2] late days (Credit Policy, 2010). A borrower is considered a good borrower when he repays his loan on time. **Discovery** is the process of identifying the characteristics that are to be variables in the model. The financial institution must identify the variables that are likely to influence the risk of repayment. **Development** involves weighing the selected model factors and creating a scorecard[3];

4. <u>Testing, implementing and managing of scoring</u>: Back testing[4] is a key tool in setting scoring policy. After back testing, a pilot test is conducted on new loans, to train the users of the scoring system, notably the loan officer to ensure viability and that the model works. Moreover, the use of adequate technologies and the availability of data are tributary of the success of the credit-scoring model.

There are three credit-scoring models (Schreiner, 2003): The tree shaped scorecards (**statistical tree**)[5] is made up of leaves[6] that correspond to different segments. The segments are ordered from least to most risk. The statistical tree essentially links the past characteristics with past arrears and supposes that the future will reflect past behavior, hence allowing for the prediction of the probability of defaulting.

The **expert system**[7] also called judgmental scoring is based on the judgment and experience of the loan officers and branch managers. Loan officers and branch managers identify the client characteristics and their relevance to client behavior. The expert system differs from the trees in that the trees use quantitative experience while expert systems rely on qualitative experience. Comparing the two, expert systems have less predictive power.

The third model is the **notation by the principle of regression**. It uses mathematical formulas to establish the influence of each client characteristics on the delinquency risk (Schreiner, 2003).

---

[2] The averages depend on the general policy of each MFI.

[3] The weight for a statistical model comes from the statistical outputs, for a judgmental model weights are assigned manually based on the perceived importance of individual factors, their interactions and the implications, hybrid scorecards combine the statistical and judgmental techniques (Caire et al., 2006).

[4] Back testing is a procedure that tests the performance of a credit-scoring model on the previous loans.

[5] The functioning principle of the tree in the prediction of risk assumes the risk history associated with a segment presents the expected risk for that segment, so all loan applications that present certain characteristics identical to the risk history of past loans present the same characteristics of that segment.

[6] There are two types of trees, the four-leaf tree and the 19-leaf tree. They differ only in the number of leaves but the process is the same.

[7] There are two types of expert systems, regression and trees. The expert systems uses trees however their trees splits are based on the experience and judgment of the loan officers or branch manager, they are not based on statistical analysis. The expert system regressions use mathematical formulae like the statistical regression but the characteristics and their weights are chosen by the managers instead of derived from the data.

Of the three models, regression has the greatest predictive power and reveals the links between the risk and characteristics better than both the tree and expert system. The notation by the principal of regression assigns the weights such that they show whether the characteristics increases or decreases risk when other characteristics are held constant and by how much (Schreiner, 2003).

Taking into account its predictive nature, its efficacy, and its operational capability in assessing the probability that a potential borrower defaults contrary to the previous models presented, the regression would be the model used to predict the probability of non-repayment in the case of the MECZOP. There are several advantages to the use of credit-scoring for the MECZOP. These advantages can be regrouped into three levels: the decision making level, the administrative level and at the cost level, by reducing transaction costs.

At the administrative level, credit scoring offers privileges such as improved efficient analysis of credit applications, homogenizing the decision making process, increasing the sense of security for loan officers, a better allocation of funds and a flexible credit pricing policy.

The quantitative model allows for the adoption of a common evaluation criterion that increases the consistency of lending policies (Vigano, 1993). Since credit scoring analyzes each customer using a similar set of rules, there is consistency in the evaluation process by the entire staff, increasing the efficient analysis of credit application. The loan officers as a result are secure because they know that the decision to allocate credit was objective.

Another important advantage of scoring is how it affects the allocation of funds by allowing the MECZOP to optimize its classification and selection of members to whom credit would be allocated. Credit-scoring model facilitates the treatment of high and low risk borrowers, vis-à-vis the denial of or disbursement of credit. This allows the MFI to focus its energy on further analyzing the applications of average risk borrowers, which easily translates into increased efficient treatment of credit demand.

A flexible pricing policy can also be developed based on the envisaged level of risk. Low risk borrowers[8] can be compensated and encouraged to continue to borrow by being offered lower prices (interest rates or lower guarantee). Consequently higher risk borrowers can face greater interest rates or be subject to a higher required guarantee.

At the decision making level, the credit scoring has the advantage of reducing human error taking into account many determinants of reimbursement. As for the reduction in transaction costs, the use of credit scoring will allow for the reduction of unpaid loans as well as the time invested in recovering.

Even if credit scoring presents many advantages, there are nonetheless some inconveniences. Credit scoring can reduce access for new borrowers, as those without credit history are likely to score poorly and highly likely to be declined access to credit[9]. It can affect the pricing of credit by making it more expensive for borrowers to borrow – increasing the stringency of credit disbursement. Having a low credit score makes it difficult to get a loan at a reasonable interest rate, while those who have a good score benefit from better interest rates. This could result in the exclusion of certain groups.

---

[8] Borrowers below the tolerated risk limit.

[9] The statistical model of credit scoring can become "uncompromising" because it is based uniquely on mathematical formulas that segregate certain clients who may have a low score but whose project ideas may have potential.

In addition to addressing both the advantages and disadvantages of this approach, it is important to elaborate on the course of action that will give rise to a credit scoring model that is adapted to the specific needs of MECZOP.

3. Credit scoring: Theoretical framework

The goal of credit scoring for microfinance and other financial purposes is to discriminate between bad and good loans (Gool et al., 2009). As such, it can allow a given MFI to manage the risks of non-repayment of loans. Adverse selection and moral hazard, the dual problems of asymmetric information, justify the management of risk for MFIs. Adverse selection designates a situation in which clients know more about their risk level than their lender. In the microfinance context, borrowers hiding information to benefit from loans is an example of adverse selection. The outcome of this phenomenon is an increase in the interest rate by the MFI to protect itself against defaults. This could result in the selection of high-risk clients, because good clients will fear being able to reimburse the credit. Moral hazard takes place after the credit application and approval. It is the inability of the MFI to control the behavior of the borrower as it pertains to the purpose and objective of the loan without incurring additional costs. Borrowers can at any time adopt opportunistic behavior that could compromise their capacity to reimburse the loans. Credit scoring alleviates these problems by increasing the ability of lenders to predict the risk of different borrowers and pricing accordingly.

4. Description of the Selected Variables

The selection of the variables is from evaluation of the socio- economic and demographic environment of Potou. It was also based on an assessment of the information available. A key determinant of variable selection was the information provided by the loan officer as well as a thorough review of MECZOP's loan application.

Dependent variable

The dependent variable in this model is late repayment, labeled in the data as "empr-retard". It is a dummy variable coded one if a repayment was late and zero if the loan payments were on time.

The independent variables

After a review of the literature (Schreiner[10], 2004; Gool and al., 2009), the independent variables are arranged in three categories according to the determinants of repayment: socio–economic characteristics of the borrower, characteristics of the loan and those related to the experience of the loan officer. The characteristics of the borrower provide insight into the likelihood of repayment.

[10]Schreiner classified the explanatory determinants of reimbursement into three categories: the socio-economic characteristics of the borrower, the characteristics of the borrower, and the loan characteristics.

Knowing about the borrowers' external conditions (social and economic) is an important way of knowing what factors could reduce their willingness to repay. The characteristic of the loan is an important factor because of how the demanding contractual conditions of the MFI can induce or discourage borrowers to respect their obligations. The loan officer plays an important role in identifying a bad borrower, by evaluating their credibility prior to credit disbursement.

For socio-economic characteristics of the borrower, the variables are as follows:

- The age of the borrower
- Sex (gender)
- The number of dependents
- The number of loans already obtained
- The reimbursement of past loans
- The number of years spent practicing an economic activity

The variables for the characteristics of the loan are as follows:

- The purpose of the loan
- The guarantee provided
- The frequency of repayment
- The loan amount
- The duration of the loan
- The time between the demand and the loan disbursement

The final category is the experience of the loan officer.

5. Description of the Sample

The sample size is made up of 30 borrowers for whom the loans were disbursed and or reimbursed during the period from January 1st 2007 to the 31st December 2010. In the sample we have 15 good borrowers and 15[11] bad borrowers. The sampling was based on the total number of customers who applied for a credit in order to avoid the reject inference bias[12].

A sample was randomly drawn that only constituted retail loans. However, the MECZOP's database only gathers data on clients to whom a credit has been given. So the problem of reject inference bias does not occur.

[11] The bad borrowers were selected using MECZOP's criteria.

[12] It is the process of deducing how a rejected applicant case would have behaved had it been granted the credit: the performance classification will be assigned to rejected cases. The data is then included in the scoring model development process. The accounts with known classification are to be augmented, in order to obtain a complete picture of the population applying for credit. The scoring model that inferred the information of the rejected applicants should theoretically be better than one built only on those accepted credit (Liu, 2001, P.27). However the effectiveness of this approach is still being disputed.

6. Model Specification

The variable that the regression seeks to explain is coded Y = 1 or Y = 0. The independent variables that can affect the dependent variable are noted with X. The relationship between these explanatory variables and the dependent variable is explained by the discrete probability model displayed below:

$$\begin{cases} \Pr(Y=1) = F(X\beta) \\ \Pr(Y=0) = 1 - F(X\beta) \end{cases}$$

Where F is a function defined by the interval [0,1]; and denotes the coefficients that are associated to the vector X.

Empr-retard, the variable defining reimbursement, is equal to zero if the borrower repays without delay and equal to one if the borrower is late in repaying the loan. The empirical econometric model has the following form:

$$\text{empr\_retard} = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i$$

P(.) is the probability that the event "late reimbursement" occurs. This is the probability that the event is observed and not the likelihood that the event itself takes place. Consequently, it is possible to create a model for the probability that an individual reimburses the loan late (EMPR_RETARD =1). This model using the logistic function is presented as follows.

$$\log\left[\frac{\Pr(\text{empr\_retard}=1)}{\Pr(\text{empr\_retard}=0)}\right] = \alpha + \sum_{k=1}^{K} \beta_k X_k$$

Equivalently, Pi $=\frac{e^{Y_i}}{1+e^{Y_i}}$ defines the probability that a potential borrower will be late in reimbursing, where Y_i represents the score function defined by the model.

Given the assumptions and the conditions of estimation, some of the variables defined above where not used in the final model. The whole equation is:

$$P(\text{empr\_retard}=1) = \beta_0 + \beta_1 \text{AGE}_i + \beta_2 \text{AGE}^2{}_i + \beta_3 \text{SEXE}_i + \beta_4 \text{ANTCD\_CRD}_i + \beta_5 \text{NB\_CREDIT}_i + \beta_6 \text{GARANTIE}_i + \beta_7 \text{EXP\_AGT}_i + \beta_8 \text{MONTANT\_CREDIT}_i + \beta_9 \text{FREQ\_REMB}_i + \beta_{10} \text{ANNEE\_EXP}_i + \varepsilon_i$$

where empr_retard is the dependent variable. The explanatory variables are the following:

- age (AGE)
- age^2(AGE^2)
- sexe (SEX)

- reimbursement history (ANTCD_CRD)
- the number of credit (NB_CREDIT)
- guarantee (GARANTIE)
- the experience of loan officer (EXP_AGT)
- the loan amount (MONTANT_CREDIT)
- the frequency of reimbursement / repayment (FREQ_REMB)
- years practicing an economic activity (ANNEE_EXP)

The table 1 below is a summary of the different variables used in the model with their anticipated signs.

Table 1: Variables, Anticipated Sign and Description

Variable	Type	Anticipated sign	Description
Dependent Variable *Late Repayment/ Reimbursement*	Binary		Whether the borrower repaid his loan late
Independent Variables			
Age	Quantitative	(-)	Age of the borrower
Sex	Binary	(-)	Sex of the borrower
Reimbursement history	Binary	(+)	Reimbursement of previous loans
Number of credit	Binary	(-)	Number of loans obtained
Guarantee	Binary	(-)	Nature of the guarantee
Experience of the loan officer	Quantitative	(-)	Number of years the loan officer has been working in the field
Loan amount	Quantitative	(-)	Loan amount
Frequency of Reimbursement	Binary	(-)	Frequency of loan reimbursement
Years practicing an economic activity	Quantitative	(-)	The number of years spent practicing an economic activity

Source: from the study, 2011

The significance of the individual coefficients as well as the percentage of variation in the dependent variable explained by the independent variables will be the basis upon which the strength of the model will be determined.

7. Results of the Regression

The estimation of the model has given the following results.

Table 2: Estimation Results

Logistic Regression				**Number of observations = 30** **Wald chi2(10) = 16.57** **Prob> chi(2) = 0.0845** **Pseudo R^2=0.6057**
Tardiness (EMPR_RETARD)	Coefficients	Robust Standard Error	Z	P>|z|
AGE	2.443829	0.9512724	2.57	0.010***
AGE^2	-0.0255646	0.009814	-2.60	0.009***
SEXE	-2.443102	1.209673	-2.02	0.043**
ANTCD_CRD	9.251004	3.879988	2.38	0.017**
NB_CREDIT	-0.3984008	1.853168	-0.21	0.830
GARANTIE	-7.183442	3.178908	-2.26	0.024**
EXP_AGT	-0.3516837	0.3161463	-1.11	0.266
MONTANT_CREDIT	3.79E-06	2.47E-06	1.53	0.126
FREQ_REMB	9.149246	3.380873	2.71	0.007***
ANNEE_EXP	-0.1810599	0.1125361	-1.61	0.108
$\square_0$	-53.52892	21.45759	-2.49	0.013
Variables significance : ***=1%., **=5% ,*=10%				

Source: From the study, Stata, 2011

Using Y to refer to the dependent variable and the score function (Tardiness (EMPR_RETARD)), the equation is as presented below:

Y =-53.52+2.44AGE-0.025AGE^2-2.44SEXE+9.25ANTCD_CRD-0.39NB_CREDIT-7.18GARANTIE-0.35EXP_AGT+3.79e^{-6}MONTANT_CREDIT+9.14FREQ_REMB-0.18ANNEE_EXP

The probability that a borrower is late with reimbursing is given by:

$$P = \frac{e^y}{1+e^y}$$

This probability corresponds to the credit scoring model of the MECZOP. As an example, assume there is a borrower with the following characteristics: **Age** = 45, **gender (SEX)** = masculine, **reimbursement history** = not late, **number of credit obtained** = less than three, **guarantee** = yes, **experience of the loan officer** = 10 years, **amount of loan** = 275,000 FCFA, **frequency of reimbursement** = many, **Years practicing an economic activity** = 30 years. The score function of this borrower is then calculated to equal -0.24, and the probability that the borrower is late in repaying is 44.02%.

8. Validity of the model and the significance of the individual coefficients

The probability (> chi (2)) tests null hypothesis (H_0) that all the estimated parameters are null, against the alternative hypothesis (H_1) that at least one of the parameters is not null. This test allows us to measure the contribution of the variables in the model to explaining the variation in the dependent variable. The value of the probability (>chi (2)) = (0.0845) indicates that all the retained variables in the model jointly contribute to explaining the probability of repayment. Robust standard errors were estimated correcting for heteroscedasticity.

From an econometric point of view, an analysis of the table indicated that six (06) out of the (10) explanatory variables are significant.

AGE, AGE^2 and FREQUENCY OF REIMBURSEMENT have a significance of 1%. The analysis of the estimated coefficients of Age and Age^2 indicates that age has a concave shape and impact on late reimbursement, Age has a positive influence on late reimbursement, whilst Age^2 has a negative impact on it, setting then the turning point (vertex of the parabola) at 47.8 years[13]. This means that younger borrowers have a tendency to be late with reimbursing while older don't. The significance of frequency of reimbursement (FREQ_REMB) shows that multiple repayments option has a positive effect on late reimbursement. This result does not match the predictions that were made on this variable. This could be because the loans disbursed are mostly for a short duration and thus applying a high frequency of due repayments can affect the capacity of the borrower to repay his/her loan on time.

Gender (SEX), reimbursement history (ANTCD_CRD) and guarantee (GARANTIE) are significant at 5% level. The relationship between SEX and late repayment identified in the regression results is similar to conclusions drawn in numerous studies (Bert et al., 2011; AusAID, 2008; Armendariz et al., 2008). It indicates that a woman is more likely to repay her loan on time. The significance of the reimbursement history confirms that the credit history of a borrower is a good indicator of the likelihood of timely reimbursement.

The variable GUARANTEE is significant such that it enables the risk to be shared between the MECZOP and the borrower. The borrower that guarantees his/her loan is expected to repay loan on time. As for the remaining variables, they are not significant but however nonetheless contribute to explaining late reimbursement.

The problem of normalizing the variance makes it such that the numeric values (estimated coefficients) of the variables cannot be interpreted directly (Hurlin, 2002). The sign of the parameters only indicates the direction that the explanatory variables associated with the different parameters influences the probability of late reimbursement. Consequently, computing the marginal effects will measure the sensibility of the probability of reimbursement with respect to the variations in the exogenous variables.

[13] This is calculated using the formula of the vertex point. In the estimated equation, the variables other than age and age^2 are held constant. The result is obtained by differentiating the equation relatively to age. 47.8 years is therefore the turning point from which the effect of age on the variable reimbursement changes sign. Hence, the probability that a borrower is late with reimbursing increases with age until 47.8 years after which each additional year reduces the probability of tardiness.

9. Analyzing the Marginal Effects/Discrete Changes

Only the marginal effects of the significant variables will be analyzed. These effects are presented in the table 3 below.

Table 3: Marginal Effects

Variables	Marginal effects (dy/dx)
Age	0.2050036
Age²	-0.0021445
Sex	-0.2049425
Reimbursement History	0.7760315
Guarantee	-0.6025916
Frequency of Reimbursement	0.7674955

Source: From the study, Stata, 2011

The marginal effect indicates that when a borrower is below 47.8 years, the probability that he is late with reimbursing increases by 0.2, while when a borrower is above 47.8 years, the probability of lateness decreases by 0.002, ceteris paribus. For example, taking the example presented above and holding all else equal, the probability of late reimbursement changes from 44.02% to 30.15% when the age of the borrower increases from 45 to 55. This corroborates the fact that older borrowers have better credit risk than younger borrowers.

The marginal effect of gender (SEX) indicates that if the borrower is a woman, the probability that she is late giving the reimbursement decreases by 0.20. Where reimbursement history indicates tardiness, the probability that the borrower is late giving the reimbursement increases approximately by 0.77. In addition, when the loan includes a guarantee[14], the probability that the borrower is late decreases by 0.60. When the frequency of reimbursement is high, requiring multiple payments, the probability of default increases by 0.76.

To evaluate the quality of the prediction, it was necessary to generate a prediction table from the model to evaluate its ability to predict the realization of the events associated with the dependent variable.

[14] The guarantee in this case is all other forms of guarantee other than the mandatory 20% of the loan amount deposited when the loan amount is below 250,000 FCFA. As a result the guarantee variable is coded 0 when 20% of the loan amount is required and 1 when it is not but another form of guarantee (material or other is required).

10. Quality of the Prediction: Table Of Classification And The Receiver Operating Characteristic (ROC) Curve

The quality of the prediction is presented on the table 4 below.
Table 4: Classification Table

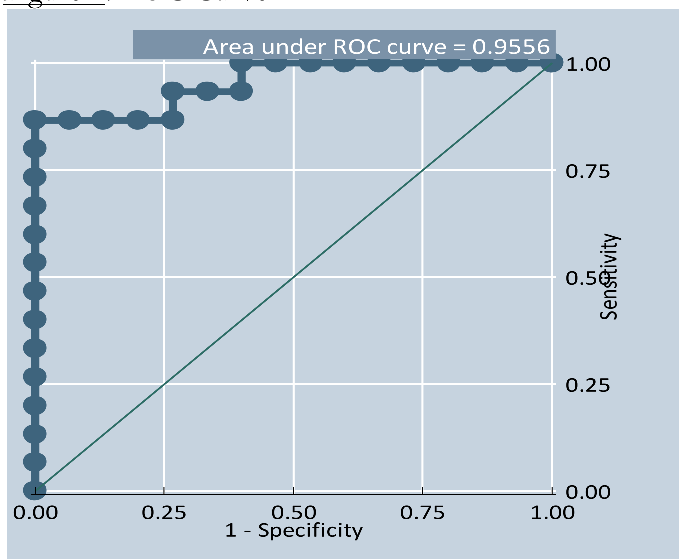
Threshold	0.5
Sensitivity	86.67%
Specificity	93.33%
Positive Predicted Value (late reimbursement)	92.86%
Negative Predicted Value(reimbursing loan on time)	87.50%
Correctly classified	90%

Source: From the study, Stata, 2011

The table shows that at a threshold of 0.5, the overall accuracy of the model to predict a delay in reimbursement (with a probability of 0.5 or greater) is 90 %[15]. A sensitivity of 86.67 shows that among the 15 events "delay in reimbursement", 13 will test with delay and 2 will test without delay. A specificity of 93.33% means that of the 15 events "reimbursement on time" 14 will test without delay and 1 will test with a delay. The probability (92.86%) that a borrower is in arrears for reimbursement when the "delay in reimbursement" is observed represents the *positive predictive value*. The negative predicted value (87.50) is the probability that a borrower is not in arrears for repayment when the event "reimbursement on time" is observed.

Another good way to assess the fitness of the model is to plot the proportion of observations classified as "tardiness in repayment" (sensitivity) against the proportion of observations misclassified as "tardiness in repayment" (1-specificity) to give a ROC curve as shown in figure 2.

Figure 2: ROC Curve



Source: From the study, Stata, 2011

[15]This rate is equal to the sum of the cases correctly predicted divided by the total number of observations.

Our model has an area[16] of 0.9556. As a result, this shows that in almost 96% of cases, if a positive event (delay in reimbursement) and a negative event (non-delay in reimbursement) are randomly selected, the model will assign a higher probability to the event "delay in reimbursement" which denotes its accuracy prediction. As a result, this shows that in almost 96% of cases, if a positive event (delay in reimbursement) and a negative event (non-delay in reimbursement) are randomly selected, the model will assign a higher probability to the event "delay in reimbursement" which denotes its accuracy prediction.

11. Conclusion

In the context of Senegal, credit risk has been identified as a threat to the sustainability of credit access in the rural context. The implications of poor credit risk management are severe. The MECZOP's operation is funded primarily by external borrowing and its own capital. Repayment constitutes a serious risk for the MECZOP's source of funds because it may not be in a position to meet its own financial obligation. Additionally, the economic development opportunities of the rural community of Leona, where the MECZOP operates is linked to the MFI's capacity to provide loans and secure savings.

The MECZOP's credit scoring model allows for a greater grasp of relevant variables that influence significantly late reimbursement. Age explains the late reimbursement and the MECZOP would have to keep an eye on borrowers that are under 47.8 years of age. Given that women are good borrowers, the credit portfolio allotted to women should be increase to allow them to better develop their income generating activities. Additionally, such a policy acknowledges the important role women play in the economy. MECZOP Also has to focus on loans that can be repaid with multiple payments. Collateral minimizes the risk of non-reimbursement, it is important that the MECZOP in addition to the forced–savings demands collateral for loans that are high. The MECZOP would benefit from being reluctant to issue credit to borrowers that have previously being late in repaying their loans. In addition to the addressed variables above, the MECZOP should take into consideration the other determinants of the model to better address credit risk.

Once the predictive power of the model has been confirmed, the MECZOP will be able to use it to select its borrowers more judiciously. To do so MECZOP would have to define scoring thresholds that correspond to the acceptance, rejection or re-evaluation of an application based on the borrower's credit score.

[16] An appraisal of the predictive power of the model is the area under the ROC line which varies between 0.5 and 1. Each point on the ROC plot represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions of results) has an ROC plot that passes through the upper left corner, where the true-positive fraction is 1 (perfect sensitivity), and the false-positive fraction is 0 (perfect specificity). The theoretical plot for a test with no discrimination is a 45° diagonal line from the lower left corner to the upper right corner. Qualitatively, the closer the plot is to the upper left corner, the higher the overall accuracy of the test (Zweig&Campbell, 1993, P.565).

References

Armendariz et al., (2008). Gender Empowerment in Microfinance. June 2008.

AusAID, (2008).Microfinance, gender and aid effectiveness. AusAID Office of Development. October 2008.

Bert et al., (2011). Women and Repayment in Microfinance: A Global Analysis. World Development Journal. Elsevier May 2011.

Caire et al., (2006). A handbook for developing credit scoring systems in microfinance context. USAID, February 2006.

Caire, D., (2004). Building Credit Scorecards for Small Business Lending in Developing Markets, *Bannock Consulting*, November 2004.

Gool, J. V. et al., (2009). An Analysis of the Applicability of Credit Scoring for Microfinance, University of Southampton, May 2009.

Hurlin, C., (2002). Econométrie des Variables qualitatives. Cours De Maîtrise D'économétrie. Université d'Orléans.

MECZOP, (2010). Credit policy paper

Schreiner, M., (2003). Scoring: The Next Breakthrough in Microcredit? CGAP Occasional Paper N°7.

Schreiner, M., (2004). Benefits and Pitfalls of Statistical Credit Scoring for Microfinance, *Center for Social Development*, Washington University in St. Louis, USA.

Sur, K., (2008). Credit Risk Analytics: A Cornerstone for Effective Risk Management, *an Oracle White Paper, Oracle financial services,* October 2008.

Vigano, L., (1993). A Credit Scoring Model for Development Banks: An African Case Study. Savings and Development. Quarterly Review. Vol 17.N°4.

Yang Liu, (2001). New Issues in Credit Scoring Application, *Work report No 16*, Institut für Wirtschaftsinformatik, 2001.

www.cgap.org

Zweig, M. H., Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry 39 (review), *561-577*.

Annexes

Logit regression results

Logistic regression	Number of obs	=	30
	Wald chi2(10)	=	16.57
	Prob > chi2	=	0.0845
Log pseudolikelihood = -8.1992591	Pseudo R2	=	0.6057

empr_retard	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	2.443829	.9512772	2.57	0.010	.5793603	4.308298
age2	-.0255646	.009814	-2.60	0.009	-.0447997	-.0063295
sexe	-2.443102	1.209675	-2.02	0.043	-4.81402	-.072183
antcd_crd	9.251004	3.880015	2.38	0.017	1.646315	16.85569
nb_credit	-.3984008	1.853171	-0.21	0.830	-4.030548	3.233747
garantie	-7.183442	3.178927	-2.26	0.024	-13.41402	-.9528599
exp_agt	-.3516837	.3161474	-1.11	0.266	-.9713213	.2679538
montant_cr~t	3.79e-06	2.47e-06	1.53	0.126	-1.06e-06	8.64e-06
fréq_remb	9.149246	3.380901	2.71	0.007	2.522803	15.77569
année_exp	-.1810599	.1125368	-1.61	0.108	-.4016279	.0395081
_cons	-53.52892	21.45769	-2.49	0.013	-95.58521	-11.47263

Marginal Effects/Discrete Changes

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.2050036	.0259957	7.89	0.000	.1540529	.2559542
age2	-.0021445	.0002753	-7.79	0.000	-.0026842	-.0016048
sexe	-.2049425	.1061814	-1.93	0.054	-.4130542	.0031692
antcd_crd	.7760315	.1436279	5.40	0.000	.494526	1.057537
nb_credit	-.0334203	.1610299	-0.21	0.836	-.3490332	.2821926
garantie	-.6025916	.1997236	-3.02	0.003	-.9940426	-.2111406
exp_agt	-.0295014	.0219555	-1.34	0.179	-.0725334	.0135306
montant_cr~t	3.18e-07	2.02e-07	1.57	0.116	-7.88e-08	7.15e-07
fréq_remb	.7674955	.1424011	5.39	0.000	.4883945	1.046596
année_exp	-.0151884	.0064859	-2.34	0.019	-.0279005	-.0024763

Classification table for late in reimbursement

Classified	True D	True ~D	Total
+	13	1	14
-	2	14	16
Total	15	15	30

Classified + if predicted Pr(D) >= .5
True D defined as empr_retard != 0

Sensitivity	Pr(+| D)	86.67%
Specificity	Pr(-|~D)	93.33%
Positive predictive value	Pr(D| +)	92.86%
Negative predictive value	Pr(~D| -)	87.50%
False + rate for true ~D	Pr(+|~D)	6.67%
False - rate for true D	Pr(-| D)	13.33%
False + rate for classified +	Pr(~D| +)	7.14%
False - rate for classified -	Pr(D| -)	12.50%
Correctly classified		90.00%