Title:

Potential of Credit Scoring in Microfinance Institution in US

(Community Venture Corp. of Kentucky Taken as Case Study)

Conference:

Selected Paper for the SAEA (Southern Agricultural Economics Association) meetings
Little Rock, Arkansas, February 5 - 9, 2005

Authors:

Mr. Jiang Hou (Primary Author)
Ph.D. Candidate and Research Assistant
Department of Agricultural Economics in University of Kentucky
337 C.E. Barnhart Bldg.
Lexington K.Y. 40506-0276
Phone: (859)797-5987
Fax: (859)257-7290
Email: jhou2@uky.edu

Mr. Jerry R. Skees
Professor
Department of Agricultural Economics in University of Kentucky
310 Charles E. Barnhart Bldg.
Lexington, KY 40546-0276
Phone: (859) 257-7262
Fax: (859) 257-7290
E-mail: jskees@uky.edu

Ms: Wei Wang
Ph.D. Student and Research Assistant
Department of Agricultural Economics in University of Kentucky
337 C.E. Barnhart Bldg.
Lexington K.Y. 40506-0276
Phone: (859)608-1852
Fax: (859)257-7290
Email: wwang5@uky.edu

Subject Areas:

1. Agribusiness and Finance
2. Research Methods-Quantitative-Econometrics

Potential of Credit Scoring in Microfinance Institution in US

(Community Venture Corp. of Kentucky Taken as Case Study)

Background of Microfinance

United States has the most advanced financial service in the world, but as many as 12 to 15 millions households have no access to the basic mainstream financial service (James H. Carr and Zhong Yi Tong). Lack of access to the financial services hampers the households to improve their circumstances and it also undermines the community they live.

One particular federal program, the SBA Microloan program, combines technical assistance and financial assistance to the often ignored underdeveloped areas including the poor rural areas. The SBA Microloan program began as the Microloan Demonstration program in 1992 under the Bush Administration. The program was designed to target underserved populations such as women, minorities, Indian tribes and the potential micro entrepreneurs in rural areas. Originally, the program was only authorized for certain states as a pilot program.

The microloan program operates through an intermediary lending process. The SBA lends money to intermediary lenders who in turn loan money to microentrepreneurs. The intermediary lender is eligible to borrow a ten year loan up to $750,000. An intermediary may borrow up to $2.5 million through the history of their participation in the program.

In FY 2003, the SBA microloan program provides over 29 million dollars in loans to over 2,400 microentrepreneurs. This is actually far less than the program has been authorized. The original authorization included funding of annually 100 million dollars. However, Congress has never appropriated that amount of funding.

There are some unique characteristics of the microloan program that distinguishes it from other SBA loan programs. Most importantly, microentrepreneurs generally have little or no collateral and little business experience. Further, microentrepreneurs generally have credit scores below 700; while most other SBA lenders have credit scores over 700.

The objective of this paper is to lay a conceptual framework for the decision making process facing an SBA microlender intermediary (CVC). The credit scoring experiences in the agricultural lending and other commercial lending will be applied in the practices of microfinance. Based on discussion about the U.S. microfinance program, there remain important issues regarding the long-term viability of this development policy. In particular, an important question is the ability of U.S. microfinance lenders to manage portfolio risks so as to balance the tradeoff between the need to payoff of the SBA loan and the desire to serve the maximum number of clients and generate significant social returns on investment. If the risk threshold is set to high, fewer clients will be served and the program will generate lower social return on investment. At the same time, a lower risk threshold will run the potential of multiple defaults or high level of arrears leading to eventual default on the SBA intermediary loan.

The Introduction of CVC's Business

In central and northern Kentucky, CVC functions as such community-based, nonprofit organization that exists to improve the quality of life for urban and rural residents. Its central mission is to provide individual and families with the skills, income, and assets, which help them achieve financial independence.

CVC shares the common characteristics that microfinance institutions (MFI) have. Microfinance refers to the provision of financial service (including credit, deposit, and insurance) to low-income clients. Contrary to other powerful financial instrument, microfinance originated and matured in developing countries like Bangladesh, Bolivia and Indonesia and then it spread into industrial economies including United States and European countries.

Since the clients served by the microfinance institutions often belong to the low-credit population and lose access to the formal financial services, there is higher probability they default on the loans. The MFI's would face great operations risk when many default cases happen. Therefore, it is necessary to establish some mechanism to screen the unqualified applicants before the transactions. In this research, the mechanism that typical MFIs (so does CVC) take to control risk will first be analyzed. Then the focus will be on the potential analysis that the credit scoring experience in the Credit card to be applied in the practice of CVC's loan practice.

The businesses of CVC contain three fields: Small business ownership; Homeownership; and Job creation through business expansion. CVC packages intensive training and technical assistance programs with flexible, affordable lending services to provide individuals with the skills and capital they need to start and grow small business and to purchase homes. It serves 31 counties in central and Northern Kentucky through its headquarters in Lexington and its two satellite offices in Covington and Campbellsville shown as Figure 1 in appendix.

Small Business Ownership Program

CVC provides training and lending services to emerging entrepreneurs in Central Kentucky. Its staff assists clients in all phases of business development with lending services and training to both new business owners and existing businesses wishing to expand. CVC has trained over 700 individuals through its business planning courses. 140 are now operating their own businesses in Central Kentucky. The business types cover almost all the industry fields**.**

Homeownership Programs

Many people never experience the benefits of home ownership because they can't get a mortgage loan. CVC offers three different programs to assist potential homeowners. Down payment & closing cost program are prepared for the families who qualify for a home mortgage but lack the down payment and up-front expenses may qualify for this program. Lease purchase program helps families who want to own a home but don't qualify for home mortgages. The family rents the home they want to own for up to two years while resolving any credit or debt issues. Once these issues are resolved, the family assumes full home ownership. Low-cost first mortgage programs offer below-market first mortgages to clients who can't afford or qualify for traditional first mortgages.

Business Expansion Program

CVC offers technical assistance and loans to existing businesses desiring to expand. Existing businesses that have been turned down by a conventional bank or credit union, lack a solid financial history, or who are creating five or more jobs are eligible through one of these programs. For example, it provides Small Business Administration (SBA) Microloan program. Such program represents entry-level solution for businesses wishing to expand. After receiving technical assistance from business consultants, clients may be able secure a loan up to $35,000 which can be used to buy machinery, equipment, furniture, inventory, and supplies. The loan may also be used to provide working capital

Practice Taken by CVC and Other MFIs to Control Risk

For the clients are usually poor people that are denied credit because of their lack of assets, MFIs use measures other than collateral to assess worthiness and offer microloans. This credit rationing process includes loan officers analyzing business plan and project's cash flow. Solidarity contracts or group lending, the use of dynamic incentives and other methods are often taken to ensure the high repayment rate. CVC also takes such measures to control risks.

Group Lending

Group lending provides a way to price discriminate first. That is forming a group between a risky and safe type would bring no mutual benefit and would lead to risk type matching. A group lending activity can also improve on matters by inducing borrowers to invest in safety operations activities. Stiglitz's model (1990) shows that if joint-liability payment is set high enough, then borrowers will always choose to do the safe activity. Group lending actually burdens the borrower with more risk and is beneficial to the lenders. From the other side, the lenders can make sure the intentions of borrowers to invest in the safe activity and can subsequently lower the interest rate to offset the burden and still can using the liability payment to cover it costs. In all, such arrangement of peer-monitoring and group enforcement can lower interest rates, derive higher welfare and raised expected repayment rates. In CVC, a peer group can be formed where loans up to $6000 are available from the Bluegrass MicroEnterprise Fund for starting or operating a small business.

Dynamic Incentives

A second mechanism of ensuring high repayment rates in microfinance entail exploiting dynamic incentives. This includes lending to individuals in small amounts, which can be termed as test lending and help establish credibility. After several times of repayment, the trust between the borrower and lenders can be attained. The borrowers' incentives can be further enhanced if they can assure of larger loan. In CVC, smaller amount of money can be lent to some borrower to help them attain some level of credit record. After ensuring the repayment of such loan, they will have chance for larger amount of loan.

Regular Repayment Schedule

Another economic advantage exploited by microfinance if the use of regular repayment schedule. Repayments begin immediately after the loan has been conferred upon the borrowers. The banks have established a way of repaying loan, which can be weekly basis. This can screen out undisciplined borrowers. The key point of such mechanism is that periodical payment requires the households or individuals to have a stream of income to reply upon. Highly seasonal occupations such agricultural cultivation can be excluded from the borrowers lists if no other guarantee. The loan officers in CVC often meet their clients monthly to make sure the loan has been used and repaid wisely. Once they find the repayment is delayed or funding was dispersed "unwisely", they can demand the borrower to go to CVC to explain or even clear the account.

Training

The training classes are provided in MFIs the practical knowledge to do the myriad of little things it takes to start and sustain an enterprise. Training in MFIs works as monitoring and screening role beyond imparting of knowledge. The business plan can also help track self-employment efforts and to indicate who is likely to be a successful entrepreneur. In the training class of CVC, the potential client must complete the screening and assessment process. It includes a free two-hour orientation workshop to learn the services that CVC provides. The important process of assessing the clients' readiness for self-employment will proceed concurrently. The training class will also teach how to write a comprehensive business plan. Once their training and business plan are completed, may they be member of a peer group.

Literature Review about Credit Scoring

To enhance the sustainability of Microfinance institution, the credit scoring has been introduced as one screening mechanism that differentiates the unqualified applicants. The credit scoring models and other experiences in the formal financial institutions (including the experience in the agricultural lending institutions) will be summed as the following parts.

The research of Galindo et al. (2000) is to find accurate predictors of individual risk in the credit portfolios of institutions. The paper made a comparative analysis of different statistical and machine learning models of classifications on a mortgage loan data. A specific modeling methodology based on the study of error curves was introduced. Finally, he discussed the possibility to use the type of accurate predictive model as ingredients of institutional and global risk models.

Many authors have examined credit scoring in the context of a traditional commercial credit environment. Mester et al. (1997) first answer the question of what the credit scoring means. Then he analyzed the methodology of credit scoring as well as the cons and pros. Finally, the contribution of credit scoring to the securitization of loan was illustrated. Joose et al. (1998) compared the performance of Logistic analysis and decision trees in a credit classification environment. Both models were used on the extensive database of Belgium's largest banks. The research result shows the Logistic models are more consistent in the credit decision process. Handzic et al. (2003) considered three Neural Network models in the light of credit loan application classification. The goal was to find the best tool for decision and the experimental results indicate that Committee Machine models were superior to other. Vasoncelos et al. (1999) investigated a solution to a credit problem in a rather peculiar environment, characterized by a stabilized economy but subject to a high interest rate. A neural network based credit scoring system has been developed and its performance was evaluated against that attained by a traditional discriminated analysis system.

The credit scoring practice study in the agricultural lending can be summarized as following. Gallagher (2001) used the statistical model to test the financial and non-financial characteristic differences between unsuccessful and successful agribusiness loans. The fundamental principle of Gallagher's research is similar to the evaluation models of the credit worthiness of agricultural cooperatives loans taken by Rambaldi et al (1992) in which in-sample and out-of –sample prediction performance were compared to determine the best statistical model. Novak (1999) used a different algorithm recursive partitioning (which is actually decision tree as one branch of data mining) to categorize the creditworthiness of agricultural observations. Unlike the researches that took the statistical model as classification instrument, Ziari 's research (1995) applied mathematical programming in the credit scoring of agricultural loans. The prediction correctness was compared between the MP and statistical model to suggest the performances are similar to between these models.

Schreiner et al. (2003) regarded scoring as the new breakthrough in microcredit. His paper introduced the concept of credit scoring to microcredit managers. It analyzed how scoring works, its limitations and detailed technical models. The requirement for application of scoring was also advanced for a microcredit institution. The discussion in this paper drew on the long experience with scoring in developed countries. The experience with scoring in Latin America was used as the example.

Schreiner et al. (2001) argued that scoring have a place in microfinance though scoring is less powerful in poor countries than in rich countries. “The derivation of the scoring formula reveals how the characteristics of borrowers, loans, and the lender affect risk” Schreiner et al., 2001; pg. 5). He also point that scoring just complements but not replace current microfinance technology.

By the previous studies, it can be concluded the statistical model can be a reliable instrument to be used to determine the clients’ credit worthiness in the MFIs. The accurateness of the prediction model is one of the key points to apply credit scoring to MFIs’ practices.

Necessity of Scoring in MFIs

By the arrangements above, the MFIs can reduce default risk greatly. However, many drawbacks of such methods can still affect MFI’s sustainability. Since the human capital is very costly in USA, the training can be unaffordable for many micro-entrepreneurs. The regular repayment schedule can exclude the seasonal production like agriculture out of the loan portfolios and there is no evidence shown that agriculture producer tends to default the loan. For such disadvantages, scoring is another new and economical way to appraise the repayment risks. It first originated from the credit card companies. Credit card lenders often make massive numbers of small, short, unsecured microloan (similar forms also taken by Microfinance programs) at very low costs because they judge risk with statistical scoring models. Such scoring models can predict future default risk of borrowers with quantitative probability. It can reduce loan losses, enhance client loyalty, and help adjust interest rates to risk for the microfinance institutions. The time and costs spent in the collection can also be greatly reduced. Such explicit and quantitative analysis can be a big aid to decision making by the loan officers of MFIs such as CVC.

Scoring Model Introduction

Normally, scoring has two branches: subjective scoring and statistical scoring. The latter will be focus of my research. Statistical scoring forecasts risk based on quantified characteristics recorded in database. Links between risk and characteristic are expressed as sets of rules or mathematical formulae that forecast risk explicitly as a probability. It detects historical links between repayment performance and the quantified characteristics of loan applications, assumes those links will persist over time, and then forecasts future repayment risk based on the characteristics of current applications. The scoring process can be demonstrated by figure2. Many machine mining models like neural networks models can provide great explaining power but they often require the data entries more than 1000. In our case, the raw data records of all the clients in CVC are 467. Therefore, the basic statistic model like Logistic or Probit model are appropriate to be applied here.

Logistic regression is used when the outcome is a proportion (repayment probability) assumed to have a binomial distribution, with a mean that is predicted by other factors (borrowers' characteristics). Instead of predicting our Y variable (which has values 0 and 1, and expected value p that must lie between 0 and 1, the Link function called the Logistic function will be used. the Logistic function to transform our predicted response to make sure it remains between 0 and 1. The proportion p with the non-linear logistic function will be modeled as equation (1):

$$p = 1/(1+\exp(-\eta)) \quad \text{or equivalently} \quad p = \exp(\eta)/(1+\exp(\eta)) \qquad (1)$$

Where the symbol η is the linear predictor which is a linear function of the predictors

$$\eta = \alpha + \beta 1 * x1 + \beta 2 * x2 + \beta 3 * x3 + \ldots$$

Data and Variable Introduction

The data of CVC includes both demographic statistics and their business & Loan information. The demographic statistics include the following parameters: geography by county, geography by zip code, geography by urban/rural, ethnicity, gender, and income as a percentage of median, income as a percentage of poverty, income in dollars, housing - own vs. rent, citizenship, marital status, veteran status, education level, numbers in household and numbers of dependents. Business & loan information include: business status - startup vs. existing, number of Employees, training goals, total technical assistance hours, type of business, credit score, original loan amount, funding date, current balance, loan status - current/delinquent and default rate. The loans in CVC consist of two parts: one for business and the other for the housing program.

The choices of the variables are determined by each loan's characteristics and limited by the data completeness. The macroeconomic factors that also affect the repayment rate of the loan are beyond our study. The important characteristics of one loan often include the interest rate the lender charged, the amount of principal, the length of terms, and the characteristics of the borrowers that include the race, gender, income level and credit score. These important variables will be used to explain the difference of the default rate of loans. Therefore, the dependent variable Writtenoff of our study is one dummy variable that the loan has been the written off or not. If the some loan has been repaid, it equal to one, otherwise 0. Agri is a dummy variable that determines whether the loan is for agriculture related business or not. Orig_prin is the principle of the loan that have been lent out, by which to make sure the amount of the loan have some effect on the repayment rate. The distribution of the principles of loans is shown in Figure 3. More than 90 percent of loans are less than 30,000 dollars, which also reflects one important facets of Microfinance loan: small amount of each loan. Month_due is the terms of the loan. Long term loan often means higher default risk. So the sign of Month_due is expected to be negative. The distribution of the terms of loans is shown in figure4. Normally, the term for the business program is less than 5 years and that for the housing program is often concentrated in 10, 15 20 or 20 years. Curr-int is the interest rate of the loan. In classical financial analysis, the higher the interest rate, the higher credit risks. Therefore, the sign of interest rate is expected to be positive. Shown by figure 5, the interests of business loans often concentrate on 5 percent per year while those for housing loans are often more than 10 percent. Gend is the Gender of the client. Lex is also a dummy variable that determines the borrower lives in Lexington or not. For more that sixty percent clients live in Lexington, it is important to point the location's effect on the repayment rate. Incomeamount is the amount of the borrowers. Higher income normally means larger capacity to repay the loan so the sign of Incomeamout is expected to be positive. As shown by figure 6, the income level of CVC's borrower has an approximate normal distribution with mean about $30,000. White/Black are both dummy variables that clients' race are black or white persons. Housing means the use of loan are for housing or business. It is also a dummy variable. Creditscore is the credit score under the original valuation system of CVC. If the scoring system is effective, the sign should be significantly positive. Therefore the meanings of all variables are shown in Table 1. Table 2 shows the descriptive statistics for the data of CVC.

Table1 Variables Description Table

Written_off	Dummy Variable, whether the borrower repays loan or not
Agri	Dummy Variable, whether the loans is for agricultural business or not
Orig_prin	The original amount of money that was lend out
Month_due	Terms of loan
Curr-int	Interest level of loans
Incomeamount	Yearly income Level of loans
Housing	Dummy variables, 1 meaning the loans for housing program; 0 means the loans for enterprise program
Gend	Gender of Borrowers
White	Dummy variables, whether the borrowers are white people
Black	Dummy variables, whether the borrowers are black people
Lex	Dummy variable, whether the borrowers live in Lexington or not
CreidtScore	Credit score provided by the credit agency

Table 2 Descriptive statistics for the data of CVC

Variable	Label	N	Mean	Std Dev	minimum	Maximum
Written Off	WrittenOff	646	0.8869969	0.3168419	0	1.0000000
agri		646	0.3715170	0.4835847	0	1.0000000
ORIG_PRIN	ORIG_PRIN	646	14755.85	22176.47	0	211000.00
MONTHS_DUE	MONTHS_DUE	646	103.6640867	118.0916088	1.0000000	372.0000000
CURR_INT	CURR_INT	646	8.8440430	4.4258070	0	13.7500000
gend		646	0.0650155	0.2467442	0	1.0000000
Lex	Lex	646	0.5743034	0.4948313	0	1.0000000
IncomeAMOUNT	IncomeAMOUNT	574	30689.11	16332.11	27.0000000	100000.00
white		646	0.0789474	0.2698656	0	1.0000000
black		646	0.0619195	0.2411961	0	1.0000000
Housing	Housing	646	0.3746130	0.4843979	0	1.0000000
CreditScore	CreditScore	525	606.7485714	71.7781890	433.0000000	804.0000000

In accordance with the suggested research method, the following model is specified as:

$$\Pr ob(Writtenoff = 1) = f(\text{orig\_prin Months\_due curr\_int gend Lex incomeamount white black housing creditscore})$$

Statistical Test of Variables

Before the regression analysis, it is necessary to do the multicollinearity diagnostic test. Multicollinearity is a result of strong correlation between independent variables. The existence of multicollinearity inflates the variance of the parameters estimates. It may also result in wrong signs and magnitudes of regression coefficient estimates, and consequently in incorrect conclusion about relationships between independent and dependent variables. The collinearity diagnostic statistics are based on the Inflation Factor for each variable. Since for each independent variable, Tolerance = 1 – Rsq, where Rsq is the coefficient of determination for the regression of that variable on all remaining independent variables, low values indicate high multivariate correlation. The Variance Inflation Factor is 1/Tolerance, which is the number of times the variance of the corresponding parameter estimate is increased due to multicollinearity as compared to as it would be if there were no multicollinearity. There is no formal cutoff value to use with VIF for determining presence of multicollinearity. Values of VIF exceeding 10 are often regarded as indicating multicollinearity. The test result for multicollinearity is shown in Table 3. All the VIF values are less than 10, which tell there are no serious multicollinearity relationships between independent variables.

Table 3 Test Result for Multicollinearity of Independent Variables

Variable	Tolerance	Variance Inflation
Agri	0.5467	1.8292
ORIG_PRIN	0.76354	1.30969
MONTHS_DUE	0.34151	2.92819
CURR_INT	0.22709	4.40353
gend	0.99531	1.00472
Lex	0.97295	1.0278
IncomeAMOUNT	0.88143	1.13452
Housing	0.1537	6.50627
white	0.95692	1.04501
black	0.95656	1.04541

Regression Results

The Logitistic Model is specified and the results of the model is presented in table4

Table4 Regression Result of Logistic Model

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.6005	1.3204	1.4693	0.2255
agri	1	0.8667	1.0569	0.6725	0.4122
ORIG_PRIN	1	0.000038	0.000015	6.0974	0.0135
MONTHS_DUE	1	-0.00545	0.00252	4.6751	0.0306
CURR_INT	1	-0.2181	0.0788	7.6703	0.0056
gend	1	-5.8713	250.3	0.0006	0.9813
Lex	1	0.6612	0.1771	13.9320	0.0002
IncomeAMOUNT	1	7.834E-6	5.832E-6	1.8040	0.1792
Housing	1	-0.3677	1.1909	0.0953	0.7575
white	1	-0.0584	0.2953	0.0391	0.8432

black	1	-0.2312	0.2758	0.7030	0.4018
CreditScore	1	0.00240	0.00143	2.8262	0.0927

Examining the regression output, it can be found that the original principle, interest rate, term (months due) and location (Lexington)'s effect on the repayment are significant. However not all of them are consistent with the expected sign. The positive sign of original principle mean that larger loan tend to be repaid with higher probability. The Lexington's clients have a significantly higher repayment rate than other areas. The negative sign of Months_due means longer term of loan brings higher default risk. The loan with higher interest rate tends to default shown by the sign of current interest. The effect of income amount and credit scoring are marginally significant. That is the borrowers with higher income level tend to repay loan but the tendency not very apparent. The borrowers with higher credit score are more inclined to repay loan. The inclination is also not statistically convincing. For the other variables, it can be found that race and gender of borrowers have no difference on the repayment of the loan. Whether the loan is used for business or housing also has no effect on the repayment loan. The Logistic model can be specified as:

$$\Pr ob(Writtenoff = 1) = Logit(4.107 + 0.000073\text{orig\_prin} - 0.0107\text{Months\_due} - 0.45\text{curr\_int} - 15.9\text{gend} + 1.137\text{Lex} + 0.000014\text{incomeamount} + 0.0246\text{white} - 0.336\text{black} + 0.868\text{housing} + 0.0034\text{creditscore})$$

By the same procedure, the Probit model is specified and the result is presented in table5

Table5 Regression Result of Probit Model

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	4.0742	2.5990	2.4573	0.1170
agri	1	1.7076	2.3260	0.5389	0.4629
ORIG_PRIN	1	0.000073	0.000029	6.3406	0.0118
MONTHS_DUE	1	-0.0110	0.00527	4.3723	0.0365
CURR_INT	1	-0.4529	0.1584	8.1782	0.0042
gend	1	-15.8725	1182.8	0.0002	0.9893
Lex	1	1.1487	0.3184	13.0194	0.0003
IncomeAMOUNT	1	0.000013	0.000011	1.5548	0.2124
Housing	1	-0.7458	2.5910	0.0828	0.7735
white	1	0.0275	0.5572	0.0024	0.9606
black	1	-0.3291	0.4768	0.4766	0.4900
CreditScore	1	0.00350	0.00260	1.8094	0.1786

By analyzing the result above, we can find they are amazingly consistent with that of Logit model. The hypothesis test used on the predictors' form each model was based on the null hypothesis that the coefficient is equal to zero. At confidence level of 90%, the variable failed to reject the null hypothesis are gend, incomeamount, white, black, housing and creditscore. The estimation for the Probit model is

$$\Pr ob(Writtenoff = 1) = \Phi(1.632 + 0.000037\text{orig\_prin} - 0.00523\text{Months\_due} - 0.2175\text{curr\_int} - 5.8716\text{gend} + 0.6548\text{Lex} + 0.000008052\text{incomeamount} - 0.0566\text{white} - 0.2344\text{black} + 0.4388\text{housing} + 0.00233\text{creditscore})$$

Estimated Repayment Rate Comparison

It is necessary to calculate how each variable will affect the repayment rate quantitatively. Illustrated as an example, the interest rate will be allowed to vary. Other continuous variables such as original principal, terms of loans, income amount and credit score will be fixed at the mean level of CVC's clients. The typical clients are chosen as a white, male living in Lexington whose loan is not for housing. In other words, the dummy variables gend, lex, white will set equal to 1 and black, housing to 0. The results are presented in table4. It can be found as the interest increase the typical client tends to default more.

Table 6 Estimated Repayment Probability Change as Interest Rates Vary

	Probit Model		Logistic Model	
Interest	X*Beta	Φ (X*beta)	X*Beta	1/(1+exp(-x*beta)
1	-1.699	0.04466	-7.7701	0.00042198
3	-2.134	0.01642	-8.6705	0.00017154
5	-2.569	0.00510	-9.5709	0.00006972
7	-3.004	0.00133	-10.4713	0.00002834
9	-3.439	0.00029	-11.3717	0.00001152
11	-3.874	0.00005	-12.2721	0.00000468
13	-4.309	0.00001	-13.1725	0.00000190
15	-4.744	0.00000	-14.0729	0.00000077

Marginal Effect Analysis of Variables

In the following, the marginal impact of all variables at their mean levels would be examined. That is how the change of the variable at their mean level will affect their probability to repay the loans. The dummy variables will be fixed as the typical client described above. For the Logit model, the marginal impact is expressed as $\partial P / \partial X_i = L(X_i^{'}\beta)\beta, where, L(X_i^{'}\beta) = e^{(X_i^{'}\beta)} / (1 + e^{(X_i^{'}\beta)})^2$

For the Probit Model, such effect can be expressed as:

$\partial P / \partial X_i = \phi(X_i^{'}\beta)\beta,$

$\phi(X_i^{'}\beta)$ is the density of standard normal distribution.

The marginal effect of both models can be shown by table5. For a white male with mean living level who live in Lexington and borrow money for non-housing use, for each unit of interest increase, the repayment probability will decrease 0.02%. As for the other variable like original principle, it can be explained when the loan amount increase 10,000 dollars, the repayment probability will increase for 0.45% for an average client.

Table 7 Marginal Effect Analysis of Variables

	Logistic	Probit
ORIG_PRIN	4.482E-08	9.02E-10
MONTHS_DUE	-6.33E-06	-1.3E-07
CURR_INT	-0.000263	-5.6E-06
IncomeAMOUNT	9.753E-09	1.73E-10
CreditScore	2.822E-06	4.16E-08

Statistically Effectiveness Test

From the both logistic and probit model, we have got the similar analysis result, in which the sign and significance of coefficient are same; only the magnitude is some different. Now, we will determine the effectiveness of the models. The percentage of correct predictions will first be computed. A threshold of 0.5 was used to calculate the percent of correct predictions for the Logit and Probit model. Both concordant percent of prediction is 85.3%. It tells the prediction power of both models can be reliable based on the CVC data.

On the other side, the likelihood ration index (LRI) can be calculated using L_u, the value of the likelihood function when all parameters are present, and L_c, the value of the likelihood function when all the slope coefficients are restricted to zero. LRI=1- L_u/L_c . By the SAS output, we can find the null hypothesis that coefficients are equal to zero can be rejected significantly. This tells the explaining power of the variables in both models is satisfactory.

Conclusion:

Based on the data of CVC, two statistical models Logit and Probit were established. First, it is need to select the appropriate variables to estimate the repayment rate of loan. Several characteristics of loan clients were chosen. By the regression results, we can find the important variable that affects a borrower's repayment of loan significantly. Such variables can be the focus when the loan officer determine whether a loan be lend to some applicants. Using the coefficient of the variables, it can be determined any applicants' repayment probability. To further quantify such rate, marginal effect of variables was analyzed. The effectiveness of such models was tested to show it is a statistically reliable. Therefore, these models can be applied in the CVC's loan applicant evaluation practice, which can screen out the unqualified applicant at a satisfactory level.

This research could enhance CVC's self-sufficiency, which is the precondition of large-scale outreach to the economically active poor. It assumes only being profitable, can an CVC grow up and meet the widespread and long run client demand for convenient, appropriate financial service.

Finally, it must be pointed out the limitation of CVC's data (just 467 clients) prevent the use of more advanced models. When more data become available, the effectiveness of the refined model such as data mining models can be substantiated.

Appendix:

Figure1 CVC Service Distribution In Kentucky

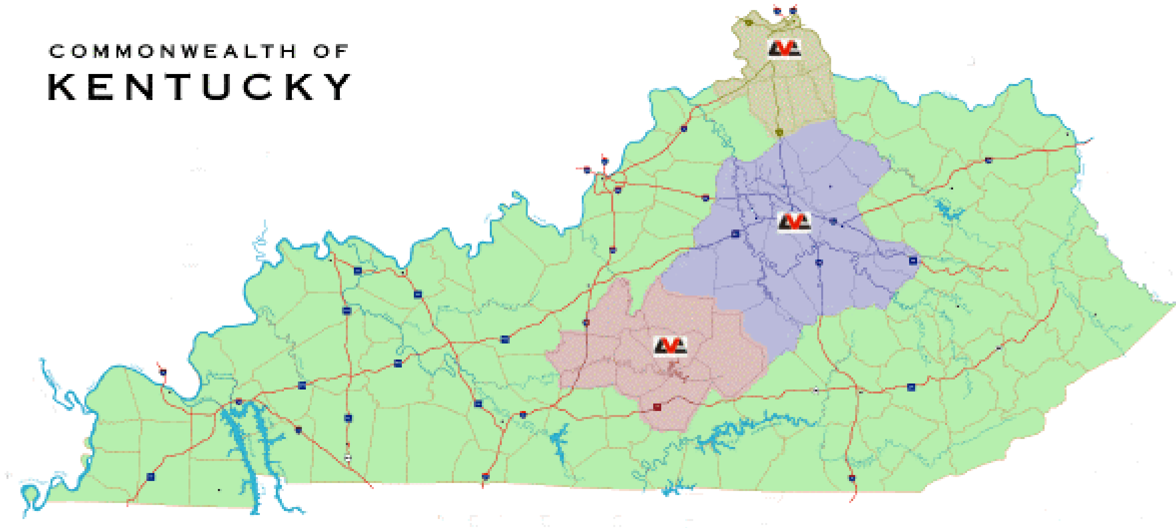


Figure2 Diagram for Credit Scoring Process

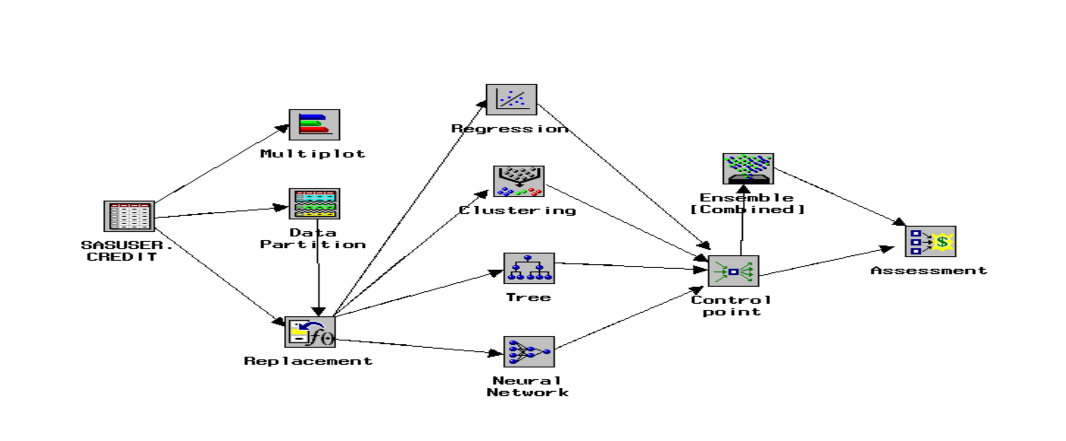


Figure 3 Distribution of the Loans' Amount

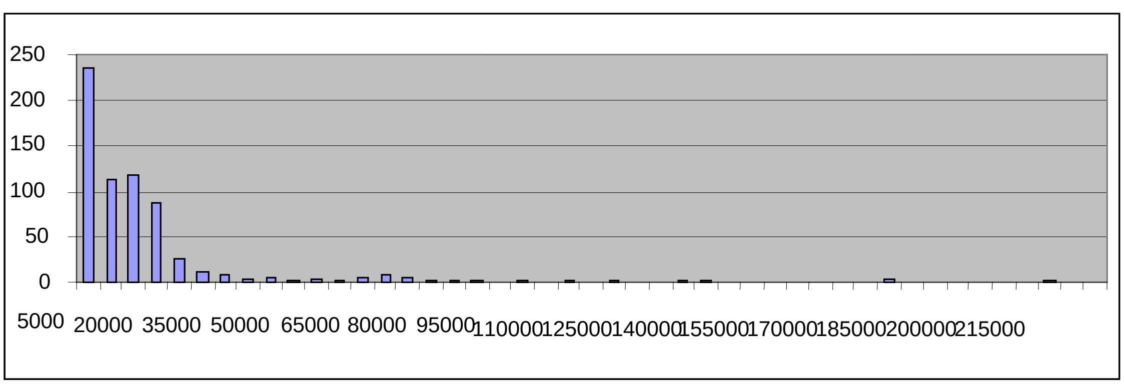


Figure 4 Distribution of the Terms of Loans

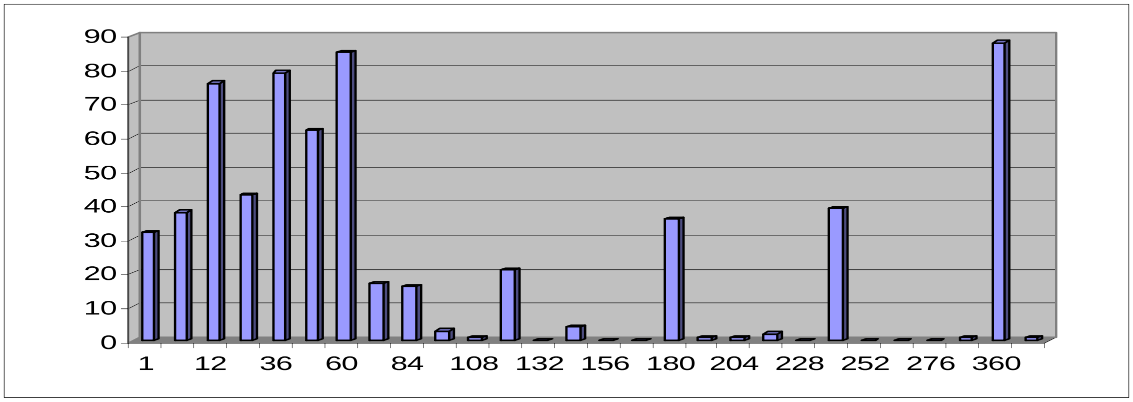


Figure 5 Distribution of Interest Rates of Loans

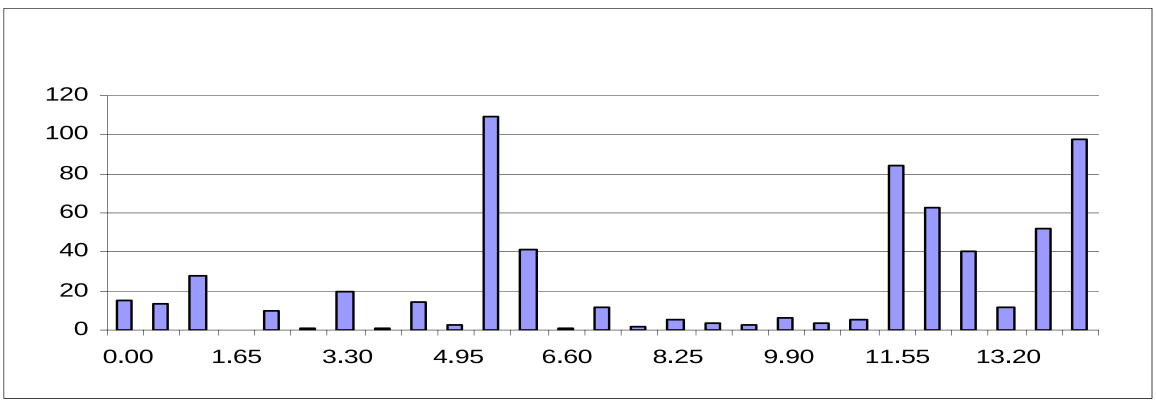
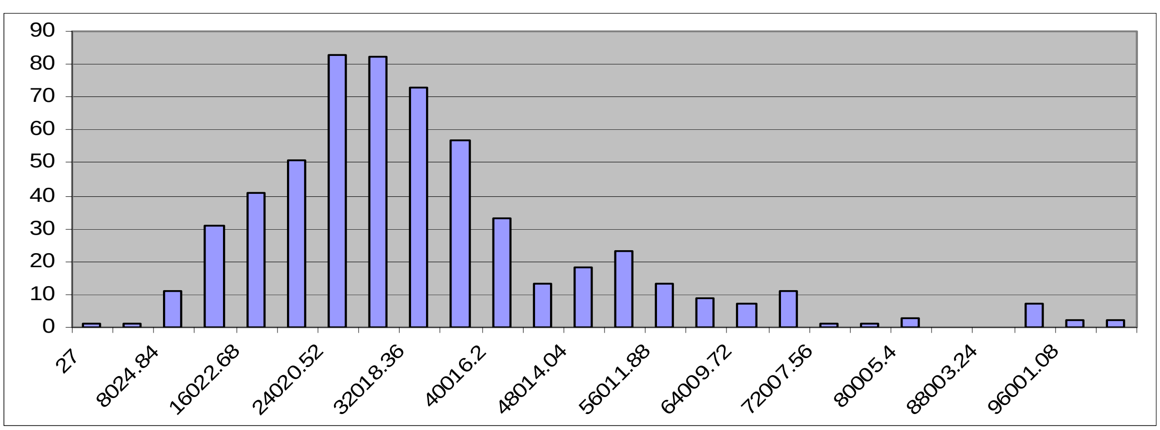


Figure 6 Distribution of the Income Level of Borrowers



Reference:

Carr J. H. and Zhong Y. T. “Replicating microfinance in the United States” Woodrow Wilson Center Press, 2002

Diagne A. “Determinants of Household Access to and Participation in Formal and Informal Credit Markets in Malawi” FCND Discussion Paper No. 67, May, 1999

Gallagher “Distinguishing Characteristics of Unsuccessful versus Successful Agribusiness Loans” Agricultural Finance Review, Spring 2001

Greene W. H, Econometric Analysis, Prentice Hall, 2002

Handzic M., Tjandrawibawa and Yeo J. “How Neural Networks can help loan officers to make better informed applications decisions” Informing Science” June, 2003

Mester L. J. “What is the Point of Credit Scoring?” Business Review, September, 1997

Novak P. Michael and LaDue L. Eddy “Stabilizing and Extending Qualitative and Quantitative Indicators of Creditworthiness in Agricultural Credit Scoring Models” Agricultural Finance Review Vol. 57, 1997

Novak P. Michael and LaDue L. Eddy “Application of Recursive Partitioning to Agricultural Credit Scoring” Journal of Agricultural and Applied Economics, 31,1, April, 1999

P. Joos, K. Vanhoof, N. Sierens and H. Ooghe (1998), "Credit Classification: A Comparison of Logit Models and Decision Trees", in Applications of Machine Learning and Data Mining in Finance, ISBN 0947-5125, pp. 59-73

Rambaldi N. Alicia, Zapata O. Hecter, and Cristy D. Ralph “Selecting the “Best” Prediction Model: an Application to Agricultural Cooperatives” Southern Journal of Agricultural Economics, July, 1992

Robinson M. S., The Microfinance Revolution: Sustainable Finance for the Poor, Open Society Institute, World Bank, 2002

Schreiner M. “ Scoring: The Next Breakthrough in Microcredit” CGAP Occasional Paper No. 7, 2003

Schreiner M. “Credit Scoring for Microfinance: can it work?” Journal of Microfiance, Vol. 2 No. 2, 2001

Schreiner M. “Evaluation and Microenterprise Programs in the United States” Journal of Microfinance Vol. 4 No. 2 2002

Schreiner M. “A Scoring Model of the Risk of Costly Arrears at a Microfinance Lender in Bolivia” Http://www.microfinance.com, 1999

Scorsone E. A. “The Role of Socioeconomic Factors in Entrepreneurship Development: A County Level Analysis” To be Published

Stiglitz, J. E., Weiss, A. “Credit Rationing in Market with Imperfect Information” American Economic Review 71 (3)

Swain B. R. “Credit Rationing in Rural India”, Journal of Economic Development, Vol. 27 No. 2, December, 2002.

Vasconcelos G.C. Adeodato P. J. L. and Monteiro D.S.M.P “A Neural Networks Based Solution for the Credit Risk Assessment Problem” Proceedings of the Brazilian Conference on Neural Networks PP 269-274, July, 1999

Zeller M., Ahmed A., Babu S. “Rural Financial Policies for Food Security of the Poor: Methodologies for a Multicountry Research Project” FCND Discussion Paper No. 11, April, 1996

Ziari A. Houshmand, David J. Leatham, and Calum G. Turvey “Application of Mathematical Programming Techiniques in Credit Scoring of Agricultural Loans” Agricultural Finance Review, Vol. 55, 1995

WWW.cvcky.org