PROBABILISTIC CREDIT SCORING FOR COHORTS OF BORROWERS

Roberto Mosquera

Working Paper Nº3

Centro de Investigaciones Económicas
Universidad de las Américas
Av. Granados E12-41 y Colimes esq.
Quito – Ecuador
May 2010

Probabilistic Credit Scoring for Cohorts of Borrowers
Roberto Mosquera
CIE Working Paper Nº 3
May 2010

ABSTRACT

As one is able to score an individual based on his level of credit risk, one should be able to score a group of potential borrowers with similar characteristics. To achieve this end, one should acknowledge that the risk of a group comes from the diversity of its members. This research proposes a methodology for applying credit scoring to groups of borrowers with similar characteristics, relying on the composition of a particular group and on the distribution of the probability of default within the group. This methodology allows ranking different cohorts of the population by their risk level, and considers in the ranking the risk preferences of the decision maker.

Roberto Mosquera
Investigador
Centro de Investigaciones Económicas – UDLA
rmmosquera@udla.edu.ec

INTRODUCTION

Credit scoring has become a major facet of today's society. For instance, in the United States an individual's credit score determines his capacity to access a series of goods and services, from a house to electricity. Credit scoring's increasing significance is due to the expansion of consumption credit in the past 40 years. Nowadays, consumption credit is found in retail stores, energy companies, conventional and cellular phone companies, public utilities, and evidently in credit cards and bank loans. In this context, a precise characterization of the borrower's credit behavior is beneficial to both the lender and the borrower (Hand and Henley, 1997). Accurate estimations of consumer's credit risk lead to an effective credit policy, in the sense that neither the amount of credit granted will constrain sales and gains, nor it will allow increased losses due to uncollectible accounts (Bierman and Hausman, 1970). From the borrower/consumer's point of view, a meticulous credit scoring will prevent over commitment, but it will also grant access to valuable resources needed.

Research about credit scoring generally deals with descriptions of the methods used to estimate the probability of default, and how they improve on each other. Hand and Henley (1997), Rosenberg and Gleit (1994), and Reichert, Cho and Wagner (1983) give a tractable description of the main techniques used in credit scoring. Wiginton (1980) presents one of the first uses of the logistic regression in credit scoring. In the nonparametric field, Hand and Henley (1996) derive a k- nearest neighbor estimator for estimating the probability of default. More recently, Zhu, Beling and Overstreet (2001) analyze the conditions where combining two credit scores leads to a better model in the sense that it estimates probabilities of default more accurately than its components. One may conclude that research in this subject has focused on improving the estimation of the expected (mean) probability of default to discriminate as accurately as possible between "good" and "bad" borrowers.

Yet, researchers have overlooked the study of other applications for credit scoring models. This is a consequence of the most common use of these models: to decide automatically which applicants to accept and which applicants to reject. However, in order to determine an effective credit policy it is not enough to accurately estimate the level of credit risk of a person once he/she applies for a credit. In a context of uncertainty about the level of credit risk, it is critical that credit managers define first to whom it is convenient to offer a credit. This implies that credit scoring models should be used for strategic risk analysis of groups (clusters) of borrowers and in the definition of both credit and marketing policies. As one is able to score an individual based on his level of credit risk, one should be able to score a group of potential borrowers with similar characteristics. Very little has been written in this direction. The common practice is to estimate the score using the mean value for a cluster of the population of whatever variables considered by the model. This approach is usually expanded with a sensitivity analysis to identify best and worst case scenarios. As it will be showed, this procedure fails to capture all the sources of uncertainty that determine the level and variability of the credit risk of a cohort of borrowers.

The purpose of this research is to propose a methodology for applying credit scoring to groups of borrowers with similar characteristics. To capture a more complete picture of the likely credit behavior of a particular segment of the population, probabilistic credit scoring relies on the composition of a particular group and on the distribution of the probability of default within the group. The methodology is illustrated for a credit card issuer in Ecuador. This procedure may be employed to evaluate the credit risk of cohorts of borrowers in any industry, whenever sufficient information is available. Furthermore, it can be generalized as a methodology for cluster analysis.

RELATED LITERATURE

Even though credit scoring systems are common in several industries, one can not find much investigation on the subject in the published literature. Hand and Henley (1997) argue that this is a consequence of the confidentiality lenders maintain on their data and procedures due to security issues and the competitive advantage given by these techniques.

As mentioned before, credit scoring models are mainly used to instantly accept or reject an applicant. Accordingly, the study of this type of models has been driven only by statistical predictability of "good" and "bad" applicants. This estimation relies on techniques such as discriminant analysis, linear regression, logistic regression, decision trees, expert systems, neural networks, and dynamic programming. Durant (1941) introduced discriminant analysis to estimate the probability of default. For a succinct description of this technique the reader is directed to Rosenberg and Gleit (1994). The main limitations of this procedure are the assumption of multivariate normal distribution, especially when employing dichotomous variables; the estimation of a priori probability for weighting the groups (in general "good risk" and "bad risk") in the classification function; and assuming that the covariance matrices across groups are the same (Reichert, Cho and Wagner, 1983).

Ordinary least squares (OLS) regression is another technique used in credit scoring models. Orgler (1970) used a linear regression to estimate a credit scoring model for commercial loans. OLS is not recommended for the estimation of probabilities of default, because the estimates will not be bounded by one, and in general this estimator is not efficient. It is better to use a logistic regression. Wiginton (1980) presented one of the first uses of the logistic regression in credit scoring. He compared the performance of a binary logit model to a linear discriminant model using in both cases the same demographic variables. His results show that the logit model predictability dominates the linear discriminant results.

Other tools employed for developing credit scoring models are decision trees, expert systems and neural networks. Rosenberg and Gleit (1994) and Hand and Henley (1997) present a review of these techniques and direct the reader to more comprehensive material. Another technique one can employ is nonparametric estimation. Hand and Henley (1996) derived a k- nearest neighbor estimator for a credit scoring model. In this investigation they argue that nonparametric estimation "enables modeling of irregularities in the risk function over the feature space" (p.78), irregularities that parametric model fail to capture due to the assumption of a specific distribution form. After empirical testing they conclude that this estimator out performs the logistic regression, the linear regression and decision trees. Another approach to improve the performance of credit score models is proposed by Zhu, Beling and Overstreet (2001); who build on the notion of second order stochastic dominance to determine the conditions where combining two credit score models leads to a better model in the sense that it estimates probabilities of default more accurately than its components.

A common problem all the techniques introduced above face is bias in the sample data. Usually, the information available consists of applicants who were approved and their posterior behavior allows the researcher or the firm to classify them as "good" or "bad". In most cases there is no information regarding the applicants that were rejected, so the sample already is biased in the sense it represents the criteria of whoever approved those solicitations. This bias affects especially those models designed to analyze the risk of new credit solicitor. This and other limitations are discussed by Capon (1982). Among other problems he mentions that even though credit score models are supposed to reflect the credit worthiness of an individual, in practice credit score designers will include any variable that statistically improves the performance of the score, for example the first letter of the last name of the person.

At this point it is convenient to clarify that there are two different types of credit scoring models. When the model is used to decide whether or not to extend credit, it is known as an application score. When the model is used to evaluate the risk of a returning or continuing (as in a credit card) customer, it is known as a behavior or maintenance score (Hand and Henley, 1997). Application scores are more common to the point that one almost automatically thinks of an application score when discussing credit scoring.

If the literature about credit scoring models in general is scarce, then published research about scoring clusters of borrowers is practically nonexistent. Because evaluating the level of credit risk for different cohorts of potential borrowers is critical to define effective credit and marketing policies, it is evident that banks and other credit granting companies are going to be particularly zealous in protecting these procedures[1]. One can only find general guidelines on how to approach the scoring of groups. The common practice is to estimate the score using the mean value for a cluster of the population of whatever variables are considered by the model. Then one should perform a sensitivity analysis to identify best and worst case scenarios. To reiterate, this procedure fails to capture all the sources of uncertainty that determine the level and variability of the credit risk of a cohort of borrowers. The next section presents the methodology for developing a probabilistic credit scoring for cohorts of borrowers.

GENERAL METHODOLOGY FOR PROBABILISTIC CREDIT SCORING

This investigation relies on stochastic simulation to construct estimates of the distribution of the credit score for a group with similar demographic characteristics. From these empirical distributions the stochastic credit score for the group is computed. One must keep in mind that the objective of a credit scoring system is to discriminate which applicants are potentially “good” clients from the general population. Thus, to develop a probabilistic credit scoring is to determine what constitutes a “good” borrower. Naturally, this classification is subjective, and in practice it should be

[1] It is logical to assume that credit granting institutions have some method to evaluate the risk of groups of borrowers with certain common characteristics.

defined by the chief officers of the credit granting institution. The main criteria to keep in mind is that it should separate “the definitely profitable [clients] from the definitely unprofitable” (Hand and Henley, 1997, p. 525). However, mainly to simplify the analysis, another approach commonly utilized is that a client is “good” if he or she is not in default or is in default for a number of days less than some cut off date (Zhu et al., 2001). In the example case presented in the following sections, a client is “good” or “bad” depending on the age of his/her debt.

Once this rule is defined, one has to select a particular model to estimate the probability of an applicant being “bad”[2]. Credit scoring extrapolates the observed credit behavior of current clients to potential applicants with similar characteristics. Therefore, the model chosen should accurately discriminate “good” from “bad” applicants, yet remain simple enough to be understood by the decision makers (Hand and Henley, 1997). Also, it is necessary that the estimated probabilities of default can be easily converted into a score, which traditionally is a number between 0 and 1000, where 1000 are the best clients. Additionally, one should be able to easily compute the marginal change in the score given by a change in each explanatory variable. These requirements are fulfilled by the logistic credit score model. Wiginton (1980) showed that given the information available,

[2] Generally, this is the probability of default, but as it depends on the definition of a “bad” client, it may have a more general sense. In this investigation, I will refer to it as the probability of default.

the logit model accurately discriminates the "good" from the "bad" applicants. Since then the logit specification has been widely used in the credit industry, thus decision makers are familiar with its characteristics. Also, the marginal effects in a logit model are directly computed by most econometric software packages.

With these general considerations, the first step is to estimate the probability of default with a logit model using demographic variables (sex, age, number of children, and city of residence, among others), and economic variables (income and indicators of wealth). The key criteria is to select variables that allow us to separate the population of interest into groups with similar characteristics, and that capture relevant differences between the members of a group. For example one may be interested in analyzing the credit behavior of single women between 25 and 35 years old who live in a particular city. In this case, the relevant differences are the variations in income and wealth levels. The estimated probability of default is converted into a score using the following formula:

$$Score = 1 - \Pr(Default) \times 1000$$

Then, it is necessary to identify the sources of risk within a particular cohort. The first source is the variability in the observable characteristics of the members of the group. For the example of single women between 25 and 35 years old who live in a particular city, one will find women with all the possible ages, with different levels of income and wealth. Given the weights for each variable computed with the logit model, each different possible combination of traits within a group will result in a different estimate of the probability of default. Thus it is necessary to consistently capture the observed variability within each group. This implies that employing only deterministic estimations for the credit risk of a group (for instance using only mean values for the different characteristics) fails to capture the real level of risk of the group. Therefore, one must use stochastic simulation techniques. The second step is to estimate the distribution for each variable that can change within a group, for every group of interest. In the case of banks or credit card operators these distributions can be estimated using their historical datasets, given the amount of information the institutions collect. In this case, it is recommended to use empirical distributions, except perhaps for the case of income, where one may consider allowing for different extreme values than the ones observed. Another source of information to estimate these distributions are census data and other surveys.

The second source of credit risk for a specific cohort is the variability of the unobserved characteristics of the members of the group. The error term captures those unobserved characteristics. Thus, to account for this unobserved variability, one estimates the residuals of the logit model. Let y be an indicator variable, such that y= 1 if the client is “bad”. The residuals are

$$R = y - \Pr(Default)$$

We are interested in the variability for each group. Thus, we require the distribution of the residuals for each particular group. From equation (2), one can see that distribution of the residuals is a mixture of the distribution of y and the distribution of the probability of default. Given that y is an indicator variable, its distribution is Bernoulli, where the probability of y= 1 (client is “bad”) is the observed proportion of bad clients in a particular cohort. The probability of default is asymptotically normally distributed (see appendix) taking the mean of the probability of default and its variance for the particular group as the distribution parameters. Thus, to generate the stochastic residuals for each group one uses

$$\begin{aligned} R_s &= Bernoulli(p) \\ &\quad - Normal\ \ \mathrm{Mean}\ \ \Pr(Default), \mathrm{Var}\ \ \Pr(Default) \end{aligned}$$

The steps to calculate a stochastic credit score are

i. Defined all relevant cohorts in the population.

ii. Estimate a logit model to calculate the probability of default for the population sample using variables that permit us to identify the groups previously defined.
iii. Estimate the distribution for each varying characteristic within a group for all the groups.
iv. Randomly draw an "individual" from the estimated distributions of a group and calculate his probability of default using the coefficients of the logit model.
v. Generate the stochastic residual using only the proportion of "bad" borrowers, the mean of the estimated probability of default and its variance for the particular group.
vi. Calculate the stochastic score:

$$\Pr(Default)_s = \Pr(Default) + R_s$$

$$Score_s = \ 1 - \Pr(Default)_s \ \times 1000$$

vii. Repeat steps iv to vi for a sufficient number of times to generate the simulated distribution of the stochastic score of the group.
viii. Repeat steps iv to vii for the other cohorts of interest.

The final issue is to rank the risk level of the specified cohorts according to their respective distributions of stochastic credit score. Once the decision makers accept a particular model of credit scoring (especially the definition of "good" and "bad"), it is reasonable to assume that they prefer score values closer to 1000 than from score values closer to 0. Thus, one can suppose that these decision makers have a utility function that ranks applicants according to the score value. In the case of scoring groups of applicants, the utility from an "applicant" becomes a lottery, due to the differences in the characteristics of the members of a cohort as discussed above. In these cases, the economic theory indicates that alternatives should be ranked by their expected utility, or by equivalent measures such as the certainty equivalent. In this case, a certainty equivalent (CE) score is defined as the score value that gives the decision makers the same utility as the distribution of stochastic credit scores of a particular cohort. Therefore, to rank the stochastic credit scores, one can compute their respective CE scores and rank them as using deterministic scores.

To apply this ranking methodology it is necessary to characterize the utility function of the decision makers. It is reasonable to assume that the degree of risk aversion does not change as the score level changes, because even a high score has a level of error, captured by the unobserved factors. This indicates that it is appropriate to use a utility function that exhibits constant absolute risk aversion. The negative exponential utility function is a natural choice. Also, decision makers in credit institutions often exhibit different degrees of risk aversion. For example, it is common that marketing managers have a lower degree of risk aversion (some even exhibit risk neutrality) than credit or risk managers. To account for these differences, we employ the stochastic efficiency with respect to a function (SERF) analysis develop by Hardaker, Richardson, Lien and Schumann (2004) in the context of agriculture economics. Basically SERF analysis ranks risky alternatives by calculating the CE score in a range of risk aversion coefficients, thus presenting the ranking for a series of decision makers. This provides a complete ranking of the risk level of particular cohorts of borrowers in a population.

The following sections of this paper demonstrate how to develop a probabilistic credit scoring model for a credit card issuer in Ecuador. The applicability of this methodology is not reduced to financial institutions. It can be applied to any firm that offers credit. Furthermore, with a proper redefinition of the categories of interest, it can be generalized as a methodology for cluster analysis.

THE DATA

This research will use the information of all principal cardholders from a major credit card issuer bank from Ecuador[3] as of February 2006. The first issue that any research about credit scoring must address is the sample bias as discussed by Capon (1982). In this case, the bank has a very lenient applicant acceptance procedure. It basically relies on the confirmation of the information presented in the application form (especially the reported income), and on requesting a guarantor for those cases where the assigned credit officer believes there is a risk. If the guarantor does not qualify, then it asks for another guarantor. Thus, for this bank the rejection rate is marginal, and any bias in the data set will be minimal.

The data set comprises information of the age of the debt (number of days in default) and the demographic characteristics of each individual. The next step is to classify each individual into "good" or "bad" credit behavior categories. For this study, "good" clients will be those who are not in default, or those who are in default for less than 90 days (92.5% of the sample). "Bad" clients will be those in default for 90 or more days (7.5% of the sample). This definition

[3] Due to confidentiality the name of the bank can not be revealed. All information that may reveal the source of the data is either omitted or replaced by a generic description.

actually combines the two criteria mentioned in the previous section because Ecuadorian law states that if a consumer credit account is in default for 90 or more days, then full provisions must be made and legal collection action must be taken. After these actions are taken, most accounts will not be profitable.

The data set was randomly partitioned into a design sample (50% of the observations) used to estimate the base logit model and the random variability in the probability of default for the different cohorts of borrowers, and a test sample (50% of the observations) employed to validate the logit model. Both samples maintain the population proportions of "good" and "bad" clients. The entire data set was used to estimate the probabilistic composition of the cohorts analyzed in this study.

The following variables are considered to estimate the base logit credit score model and to define the cohorts of borrowers:

TABLE 1
Description of the Variables

- Age of debt:	Represents the maximum number of days that an individual is in default.
- Default:	Indicator variable that marks 1 when the client is "bad".
- Sex:	Indicator variable for the sex of the client. 1 stands for women.
- Number of children:	Represents the number of persons the client is financially responsible for. It is abbreviated *num_child*.
- Vehicle:	Stands for the number of vehicles a

	client has. It is included as a measure of wealth.
- Income:	Reported monthly income of the client.
- Properties:	Stands for the number of properties (houses, apartments, or offices) the client has. It is included as a measure of wealth.
- Age	Age of the client measured in years.
- # of cell phones	Number of cell phones the client owns. It is included as a measure of wealth. It is abbreviated *numcell*.
- Cell phone	Indicator variable that marks 1 when the client has a cell phone. It is included as a measure of wealth It is abbreviated *cellp*.
- Telephone	Indicator variable that marks 1 when the client has a land line phone. It is included as a measure of wealth. It is abbreviated *tel*.
- Cell phone and Telephone	Indicator variable that marks 1 when the client has a land line phone and a cell phone. It is included as a measure of wealth. It is abbreviated *telcel*.
- City *i*	Set indicator variables that stand for the city of residence of the client. This set considers only Ecuador's nine largest cities. Each variable is abbreviated *dcityi*.
- Marital Status	Set of indicator variables that stand for marital status of the client. This set considers single, married, widower, divorced and free union.

MODEL DEVELOPMENT

In order to illustrate the methodology, this paper focuses on two main groups: single men and women residents of City 1.

Additionally, each group is further divided into seven age groups: 18 to 24 years old, 25 to 34 years old, 35 to 44 years old, 45 to 54 years old, 55 to 64 years old, 65 to 74 years old, and more the 75 years old. Consequently, there are 14 cohorts to be analyzed. In practices, management should identify all the cohorts of interest. The second step is to estimate a logit model for the entire population. This model is used to calculate the deterministic component of the stochastic score, and to generate the stochastic residuals for each cohort.

TABLE 2
Logit Credit Score Estimation Results

Dependent Variable: Default

Variable	Coefficient	Standard Error	Z-value	P-value
Sex	-0.4053	0.0364	-11.15	0.000
Number of Children	0.1320	0.0134	9.84	0.000
ln(Income)	-0.4730	0.0176	-26.94	0.000
Properties	0.0892	0.0239	3.73	0.000
Age	0.0530	0.0105	5.03	0.000
Age Squared	-0.0006	0.0001	-5.62	0.000
# of Cell Phones	0.2172	0.0799	2.72	0.007
Cell Phone	-1.1537	0.2244	-5.14	0.000
Cell phone and Telephone	0.6897	0.1450	4.76	0.000
City1	-0.2121	0.0560	-3.79	0.000
City2	0.8447	0.0560	15.07	0.000
City3	-0.5951	0.1046	-5.69	0.000
City4	-0.3494	0.1146	-3.05	0.002
City7	0.4441	0.1163	3.82	0.000
City9	0.4454	0.1088	4.09	0.000
Single	-0.3995	0.0676	-5.91	0.000

Married	-0.3435	0.0571	-6.01	0.000
_cons	-0.1130	0.2918	-0.39	0.699
Log Likelihood	-14488.564			

The observed variability inside each one of the groups previously defined is generated by the distributions within each group of the following random variables: age, number of children, income, properties, number of cell phones, has cell phone, and cell phone and telephone. Because of the discrete nature of these variables (with the exception of income), their distributions are estimated using discrete empirical distributions for each one of the 14 cohorts. Regarding income, one requires a parametrical distribution form allows considering values of income over the range of those observed in the data. Therefore, income's distribution is taken as log normal, where the mean and variance of the distribution were estimated for each cohort using maximum likelihood (MLE).

The unobserved variability within each cohort is captures by the stochastic residuals. To use equation (3), one requires the proportion of default and the mean and variance of the estimated probability of default for each group.

TABLE 3
Data for Generating Stochastic Residuals

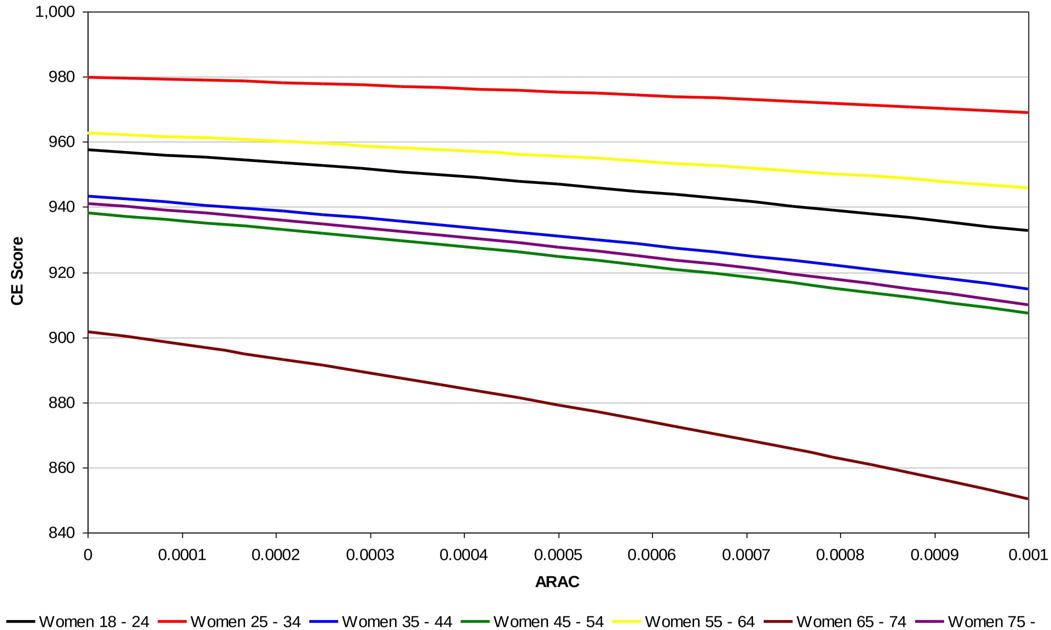
Cohort		Proportion of default	Estimated Probability of default	
			Mean	Standard Error
18 - 24	Women	3.85%	2.79%	0.0069
	Men	0.00%	4.59%	0.0155
25 - 34	Women	1.73%	3.13%	0.0105
	Men	3.73%	4.33%	0.0156
35 - 44	Women	4.59%	4.11%	0.0204
	Men	6.08%	5.61%	0.0306
45 - 54	Women	5.11%	4.41%	0.0237
	Men	7.75%	6.53%	0.0370
55 - 64	Women	2.66%	4.36%	0.0218
	Men	8.05%	5.46%	0.0266
65 - 74	Women	8.97%	3.97%	0.0159
	Men	0.00%	4.10%	0.0194
75 -	Women	5.00%	3.93%	0.0181
	Men	0.00%	4.41%	0.0215

With this information, the stochastic credit score of each cohort was simulated for 500 iterations to generate the empirical distribution of the score for each group. The next section presents the resulting risk rankings.

PROBABILISTIC CREDIT SCORING RESULTS

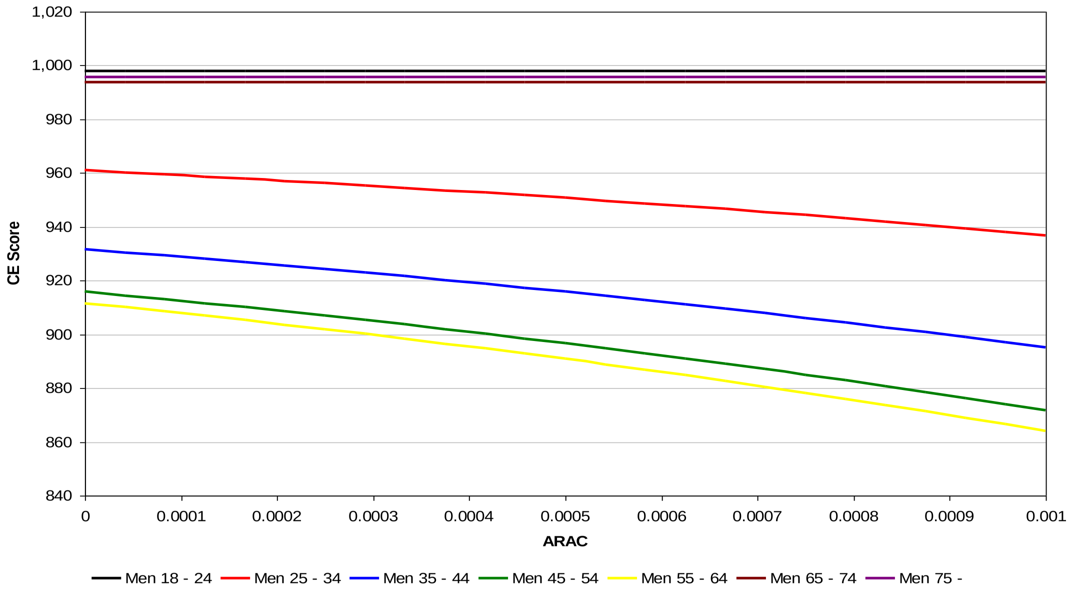
SERF analysis is used to calculate the risk rankings for the different cohorts. This paper evaluates a series of hypothetical decision makers with absolute risk aversion coefficients (ARAC) that range from risk neutrality (0) to extreme risk aversion (0.001). Since the utility functions are defined as functions of the score, there is no clear guideline in the theory to choose the upper limit for the ARAC. In such situations, using SERF to rank the stochastic credit scores has an additional advantage, since SERF's risk ranking is robust to whatever upper boundary ARAC one could choose. However, one should take care to not use an upper ARAC that is completely unrealistic.

FIGURE 1 - Risk Ranking for Single Women Living in City 1



Recall that a decision maker prefers the cohort with the highest CE score. Then for single women that live in city 1, the group between 25 to 34 years old is the least risky (most preferred), followed by the group between 55 to 64 years old, 18 to 24 years old, 35 to 44 years old, 75 years old or more, 45 to 64 years old, and 65 to 74 years old (least preferred). This ranking is consistent for all the ARAC between 0 and 0.001. Also, one should remember that this ranking reflects the information available. If one includes more variables, for example some savings indicator, the results will be different.

FIGURE 2 - Risk Ranking for Single Men Living in City 1



For single men that live in city 1, the group between 18 to 24 years old is the least risky (most preferred), followed by the group with 75 years old or more, 65 to 74 years old, 25 to 34 years old, 35 to 44 years old, 45 to 64 years old, and 55 to 64 years old (least preferred). This ranking is consistent for all the ARAC between 0 and 0.001. One should note that this is not the same ranking as for women. Each group has a different composition that implies different risk sources. This is reflected in distinct risk rankings.

Now, it is natural to wonder if a deterministic risk ranking of the cohorts differs from the stochastic ranking presented. The deterministic ranking is calculated by estimating the probability of default for the "average individual" of each cohort, and then converting this probability in a score using equation (1).

TABLE 4

Probabilistic Credit Scoring for Single Women Living in City 1 by Cohorts of Age

Cohort (years)	Deterministic Credit Score		Probabilistic Credit Scoring		
	Average Score (Risk Neutral)	Risk Ranking	CE Score Risk Neutral	CE Score Most Risk Averse	Risk Ranking
18 - 24	973.3	2nd Most Preferred	957.4	932.6	3rd Most Preferred
25 - 34	972.7	3rd Most Preferred	979.9	968.9	Most Preferred
35 - 44	968.6	4th Most Preferred	943.2	914.8	4th Most Preferred
45 - 54	966.2	Least Preferred	938	907.2	6th Most Preferred
55 - 64	966.3	6th Most Preferred	962.7	945.8	2nd Most Preferred
65 - 74	967.6	5th Most Preferred	901.7	850.3	Least Preferred
75 -	978.3	Most Preferred	940.9	909.9	5th Most Preferred

TABLE 5

Probabilistic Credit Scoring for Single Men Living in City 1 by Cohorts of Age

Cohort (years)	Deterministic Credit Score		Probabilistic Credit Scoring		
	Average Score (Risk Neutral)	Risk Ranking	CE Score Risk Neutral	CE Score Most Risk Averse	Risk Ranking
18 - 24	963.1	3rd Most Preferred	997.9	997.8	Most Preferred
25 - 34	963	4th Most Preferred	960.9	936.8	4th Most Preferred
35 - 44	958.2	5th Most Preferred	931.6	895.1	5th Most Preferred
45 - 54	953.2	Least Preferred	915.9	871.7	6th Most Preferred
55 - 64	956.2	6th Most Preferred	911.6	863.9	Least Preferred
65 - 74	966.3	2nd Most Preferred	993.7	993.6	3rd Most Preferred
75 -	974.2	Most Preferred	995.7	995.7	2nd Most Preferred

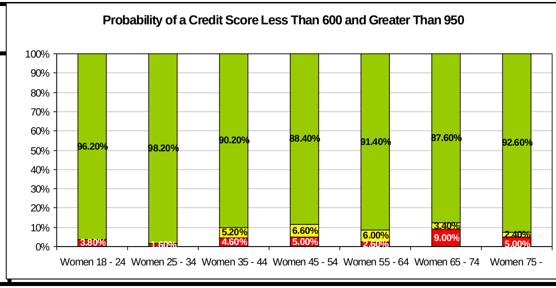
These results conclusively show that ignoring the sources of credit risk leads to incorrect risk ranking of the cohorts. For women the most preferred cohort as ranked by the deterministic approach is women older than 75 years old. When one accounts for all the sources of risk this cohorts becomes the fifth most preferred. On the other hand women between 25 and 34 years old are ranked fourth by the deterministic score. Yet, when one considers all the sources of risk it becomes the most preferred. For men the risk ranking produced by the deterministic approach is more similar to the probabilistic risk ranking, but still one can find important differences. Again, the most preferred cohort as ranked by the deterministic approach is men older than 75 years old. However, probabilistic credit scoring indicates that the most preferred group by its risk level is men between 18 to 24 years old. These results imply that a credit policy based on the deterministic results will extend credit to inappropriate (in a credit risk sense) cohorts of the population. Also, it is usual to belief that middle age people have a lower level of credit risk because they have more stable sources of income. Yet the probabilistic risk rankings for both men and women indicate the young people, starting their careers have a lower risk level. Thus, the issuer should focus to capture clients from that segment of the population, with the added benefit that they will probably stay as clients for the rest of their lives.

Probabilistic credit scoring also allows calculating the probability of observing individual score values within a particular range for each cohort. These probabilities can be considered as an additional measure of the credit risk level of a cohort. The CE scores and the probabilities of observing score values within a particular range can be summarized in score tables that permit a rapid evaluation of the cohorts for decision making. Table 6 presents an example of these score tables. For instance, a woman whose age is between 25 and 34 years old (the most preferred cohort for women) has a 98.2% chance of having a credit score of 950 or higher, and a 1.6% probability of having a credit score lower than 600. On the other hand, a man whose age is between 18 and 24 years old (the most preferred cohort for men) has a 99.6% probability of having a credit score of 950 or higher, and a 0.4% chance of having a credit score lower than 600. Similar comparison can be made for the other cohorts.

TABLE 6 Score Table for Single Men and Women Living in City 1

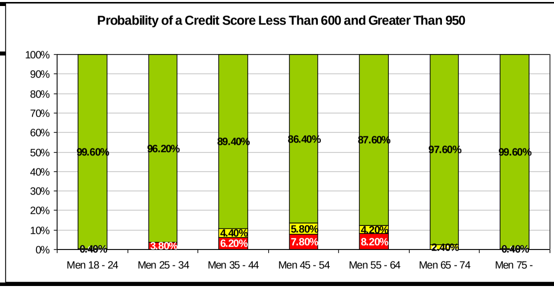
Women

Cohort (years)	CE Score Risk Neutral	CE Score Most Risk Averse	Risk Ranking
18 - 24	957.4	932.6	3rd Most Preferred
25 - 34	979.9	968.9	Most Preferred
35 - 44	943.2	914.8	4th Most Preferred
45 - 54	938	907.2	6th Most Preferred
55 - 64	962.7	945.8	2nd Most Preferred
65 - 74	901.7	850.3	Least Preferred
75 -	940.9	909.9	5th Most Preferred



Men

Cohort (years)	CE Score Risk Neutral	CE Score Most Risk Averse	Risk Ranking
18 - 24	997.9	997.8	Most Preferred
25 - 34	960.9	936.8	4th Most Preferred
35 - 44	931.6	895.1	5th Most Preferred
45 - 54	915.9	871.7	6th Most Preferred
55 - 64	911.6	863.9	Least Preferred
65 - 74	993.7	993.6	3rd Most Preferred
75 -	995.7	995.7	2nd Most Preferred



CONCLUSIONS

This paper presents a novel methodology for developing a probabilistic credit score for different cohorts of population. This procedure may be employed to evaluate the credit risk of cohorts of borrowers in any industry, whenever sufficient information is available. Probabilistic credit scores accounts for all the sources of credit risk within a particular group with similar characteristics. The resulting credit scores provide useful information for strategic risk analysis of groups (clusters) of borrowers that can be directly applied to define both credit and marketing policies. Also, probabilistic risk ranking includes the degree of risk aversion of the decision maker. Therefore, these rankings lead to define to whom it is convenient to offer a credit.

In addition, the analysis can be extended by changing the definition of “good” and “bad” clients so that it does not only consider credit risk, but also includes some explicit indicator of profitability. Then the scores would be a joint indicator of the risk level and the profitability of a particular cohort. This allows developing an integral risk management policy, where one takes risks according to the potential benefits. Furthermore, by changing the definitions of “good” and “bad”, this methodology can be generalized as a methodology for cluster analysis in any context.

APPENDIX

Asymptotic Distribution of the Probability of Default in a Logit Model

Define the logistic cumulative distribution function (CDF) as

$$F(X_i \beta) = \frac{e^{X_i \beta}}{1+e^{X_i \beta}}$$

Then the estimated CDF is

$$F(X_i \hat{\beta}) = \frac{e^{X_i \hat{\beta}}}{1+e^{X_i \hat{\beta}}}$$

And the partial derivative of the CDF with respect to β is

$$\frac{\partial F(X_i \hat{\beta})}{\partial \beta} = \mathrm{f}(X_i \hat{\beta}) X_i$$

where $f(X, \hat{\beta})$ is the logistic probability density function (pdf).

Let β be the true population value of the coefficients. Take a first order Taylor expansion of the CDF around β .

$$F(X_i \hat{\beta}) = F(X_i \beta) + \mathrm{f}(X_i \beta) X_i (\hat{\beta} - \beta)$$

$$\sqrt{n} F(X_i \hat{\beta}) = F(X_i \beta) + \mathrm{f}(X_i \beta) X_i \sqrt{n} (\hat{\beta} - \beta)$$

The estimated coefficients have an asymptotically normal distribution:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[d]{} N(0, \mathrm{var}(\hat{\beta}))$$

Then, apply the delta method to the Taylor expansion of the CDF to get its asymptotic distribution:

$$F(X_i \hat{\beta}) \overset{a}{\sim} N(F(X_i \beta), \mathrm{f}(X_i \beta)^2 X_i \,\mathrm{var}(\hat{\beta}) X_i^{'})$$

Equation A6 is empirically approximated by

$$F(X_i \hat{\beta}) \overset{a}{\sim} N\left[\text{mean } F(X_i \hat{\beta}), \mathrm{var}(F(X_i \hat{\beta})) \right]$$

References

Bierman, H. and Hausman, W. H., 1970, The Credit Granting Decision, *Management Science*, Vol. 16, No. 8, Application Series, pp. B519-B532.

Capon, N., 1982, Credit Scoring Systems: A Critical Analysis, *Journal of Marketing*, Vol. 46, No. 2, pp. 82-91.

Edelstein, R. H., 1975, Improving the Selection of Credit Risks: An Analysis of a Commercial Bank Minority Lending Program, *The Journal of Finance*, Vol. 30, No. 1, pp. 37-55.

Frydman, H., Kallberg, J. G. and Kao D., 1985, Testing the Adequacy of Markov Chain and Mover-Stayer Models as Representations of Credit Behavior, *Operations Research*, Vol. 33, No. 6, pp. 1203-1214.

Hand, D. J. and Henley, W. E., 1996, A k-Nearest-Neighbour Classifier for Assessing Consumer Credit Risk, *The Statistician*, Vol. 45, No. 1, pp. 77-95.

Hand, D. J. and Henley, W. E., 1997, Statistical Classification Methods in Consumer Credit Scoring: A Review, *Journal of the Royal Statistical Society*. Series A (Statistics in Society), Vol. 160, No. 3, pp. 523-541.

Hardaker, J. B., Richardson, J. W., Lien, G., and Schumann, K. D., 2004, Stochastic efficiency analysis with risk aversion bounds: a simplified approach, *The Australian Journal of Agricultural and Resource Economics*, 48:2, pp. 253-270.

Lawrence, E. C. and Arshadi, N., 1995, A Multinomial Logit Analysis of Problem Loan Resolution Choices in Banking, *Journal of Money, Credit and Banking*, Vol. 27, No. 1, pp. 202-216.

Orgler, Y. E., 1970, A Credit Scoring Model for Commercial Loans. *Journal of Money, Credit and Banking*, Vol. 2, No. 4, pp. 435-445.

Reichert, A. K., Cho, C., and Wagner G. M., 1983, An Examination of the Conceptual Issues Involved in Developing Credit-Scoring Models, *Journal of Business & Economic Statistics*, Vol. 1, No. 2, pp. 101-114.

Rosenberg, E. and Gleit, A., 1994, Quantitative Methods in Credit Management: A Survey, *Operations Research*, Vol. 42, No. 4, pp. 589-613.

Sexton, D. E., 1977, Determining Good and Bad Credit Risks Among High- and Low-Income Families, *The Journal of Business*, Vol. 50, No. 2, pp. 236-239.

Wiginton, J. C., 1980, A Note on the Comparison of Logit and Discriminant Models of Consumer Credit

Behavior, *The Journal of Financial and Quantitative Analysis*, Vol. 15, No. 3, pp. 757-770.

Zhu, H., Beling, P.A., and Overstreet, G.A., 2001, A Study in the Combination of Two Consumer Credit Scores, *The Journal of the Operational Research Society*, Vol. 52, No. 9, Special Issue: Credit Scoring and Data Mining, pp. 974-980.