

**ESTIMATION OF A CREDIT SCORING MODEL FOR LENDERS  
COMPANY**

**FELIPE ALONSO ARIAS ARBELAEZ**

**Proposal Report**

**Research Practice III**

**Project Advisors:**

**Juan Sebastián Bravo<sup>1</sup>**

**Francisco Iván Zuluaga<sup>2</sup>**

**EAFIT UNIVERSITY**

**DEPARTMENT OF MATHEMATICAL SCIENCES**

**MATHEMATICAL ENGINEERING**

**AUGUST 2015**

**MEDELLIN**

---

<sup>1</sup>Operation Coordinator, Linkvest Capital LLC, Miami, Florida

<sup>2</sup>Department of Mathematical Sciences, EAFIT University, Medellín, Colombia

## **Statement of the Problem**

Historically it has been seen that banks developed their own system of risk for loans to their customers, because such information is privileged, it is very difficult to know how to measure credit risk. That is why it was decided to propose a model of credit scoring to answer basic questions like: should we lend to this client? What is the loan limit? How to reduce the risk of default? Among others.

Linkvest Capital is a private equity firm that identifies, analyzes, structures, leads and supervises the businesses in which it invests. Linkvest Capital has different branches (business lines), one of them is being a mortgage loan originator and seller in Florida, United States.

Basically what is sought with the model is to decrease the probability of default using variables of individuals as their income, how long they've been a customer, among others.

## **Objectives**

### **General Objective**

Estimate a Credit Score Model using a logistic regression model to improve lending decisions on a person using historic data.

### **Specific Objectives**

- Analyze the best model to improve in our case based in the data.
- Identify the most influential variables for the company when lending money.
- Specify the model in mathematical terms.
- Estimate the probabilities of default.
- Analyze the results for decision making.

## **Literature Review**

According to Hand and Henley (1997): “Credit scoring is the term used to describe formal statistical methods used for classifying applicants for credit into ‘good’ and ‘bad’ risk classes.”

The credit scoring is assessed in terms of predictive models of payment or reimbursement by a score that measures the risk of a borrower or the operation itself. The analysis is done in terms of score is seeking to explain the financial behavior in terms of the services requested, the relationship between risk and return and the cost of the operation. (Cantón, Rubio, & Blasco, 2010).

Clearly, many models have been used to solve the credit score problem in lenders companies, we’ll mention some of the most representative and choose one to apply.

### **a) Discriminant Analysis**

It is a good role model when discriminating customers between good and bad payers, the problem that has primarily is impossible to calculate the probability of default and fails to satisfy the basic assumptions of econometrics (homoscedasticity, linearity, normality and independence) (Cantón et al., 2010).

In Altman (1968), we can see an application of this method, using the explanatory variables as ratios, he found the probability of default ratios using net income/sales, retained earnings/assets, among others.

### **b) Linear Probability Models**

In these models, least squares regression where the dependent variable takes the value of one if the client is failed and zero if the customer meets its payment obligation is used, the equation is a linear function of the explanatory variables (Hand & Henley, 1997).

Perhaps the precursor of this model is Orgler (1970) who proposed models for commercial loans, but can be successfully applied to personal loans, as in our case.

### **c) Logit Models**

Calculates the probability to belong to a group of payer or not payer. Wiginton (1980) in their study compares the logit models with discriminant analysis in the estimation of a model of default and concluded that the logit model offered a better fit in terms of probability than the discriminant analysis.

### **d) Neural Networks**

They try to imitate the nervous system, so that generate a certain degree of intelligence. Nodes that respond to certain input signals are interconnected. Rosenberg and Gleit (1994) show a summary of quantitative methods for handling credit modeling, neural networks are mentioned as a possible model but not proposed a solution on credit score but in credit fraud.

## **Justification**

For the loan to persons or entities is necessary to conduct an analysis of influential variables in the possibility that the person does not pay on time the credit. Many financial institutions use their own model appropriate to their needs, that is, to what each company believes relevant for lending money.

Linkvest Capital is a financial institution that between the services have the mortgage loan and must take into account certain variables (already categorized) to decide whether a person can repay the loan and what is the limit of the amount to pay.

Based on the above, we have the need to estimate an appropriate model to calculate the probability of default on a loan to a person in Linkvest Capital, thus being able to avoid having flows of money due to a bad loan analysis.

## Scope

The main goal is to estimate a credit score model to estimate the probability of default on personal loans using macroeconomic and personal data.

It is important to evaluate the model with real data and take into account the basic assumptions of the model chosen.

## Methodology

The methodology adopted will be based on the realization of a model by linear regression and logit model using a variation of the model called model of binary logistic regression, this is because we need as output the probability of a dependent variable (Cantón et al., 2010).

This model in most cases does not produce inefficient estimators, easily supports categorical variables, estimates the probability of loan default to the values of the independent variables and determine the influence of each independent variable on the dependent variable depending on the Odd Ratio, if this value is close to one then there is an increased probability of default and one close to zero indicates a better chance of fulfillment (Hand & Henley, 1997).

The logistic regression model can be formulated as:

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \mu \quad (1)$$

where  $\mu$  is the disturbance and  $p$  is the probability of default and can be estimated as follows:

$$p = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}} \quad (2)$$

Then, our response is going to be

$$Default = \log \left( \frac{p}{1 - p} \right) \quad (3)$$

The problem is initially having organized data, specify the model in terms of variables used to explain the dependent variable as in Equation 3, estimate the model and calculate the probability of default with Equations 1 and 2.

After the estimation of the model we have to check for the basic econometric assumptions as linearity, homoscedasticity, normality and independence. If the model have good estimations based on the significance levels and the basic assumptions so we are going to have a model that estimate the default probability on a loan.

## Schedule

<b>Week</b>	<b>Task</b>	<b>Description</b>
<b>4-5</b>	Proposal report	Oral and written proposal report
<b>6-7</b>	Specification of the model	Specify the independent variables that explain the probability of default
<b>8-9</b>	Estimation of the model	Run the model and check the econometric assumptions
<b>10</b>	Progress presentation	First presentation of the project showing the actual scope
<b>11-14</b>	More estimations	Run the model with different combination of dependent variables
<b>15-16</b>	Final report	Write the final report with all the results obtained with the automated report
<b>17</b>	Extra improvements	If all the schedule is running well, this week we are going to look for some new features we can do with the model
<b>18</b>	Final presentation	Final presentation of the project

## Resources

The resources are basically the time each tutor is going to give for the project. The student is going to work on his personal computer and only need information for the company which it's already given.

## Intellectual Property

According to the internal regulations on intellectual property (EAFIT, 2015), the results of this project are product of the co-autorship between Francisco Zuluaga, Juan Bravo and Felipe Arias. According to the internal regulations of Linkvest Capital, the information is confidential and only can be used by people in the company, so the University and the people involved in the project only can use this project according to Linkvest Capital permission.

## References

- Altman, E. I. (1968, September). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, XXIII(4), 589-609.
- Cantón, S. R., Rubio, J. L., & Blasco, D. C. (2010, June). A Credit Scoring Model for Institutions of Microfinance under the Basel II Normative. *Journal of Economics, Finance and Administrative Science*, 15(28).
- EAFIT, U. (n.d.). *Reglamento de propiedad intelectual*. Retrieved 22 Agosto 2015, from [http://www.eafit.edu.co/institucional/reglamentos/Documents/Reglamento\\_Propiedad\\_Intelectual.pdf](http://www.eafit.edu.co/institucional/reglamentos/Documents/Reglamento_Propiedad_Intelectual.pdf)
- Hand, D. J., & Henley, W. E. (1997). Statistical Classification Methods in Customer Credit Scoring: A review. *Journal of the Royal Statistical Association*, 160(Part 3), 523-541.
- Orgler, Y. E. (1970, November). A Credit Scoring Model for Commercial Loans. *Journal of Money, Credit and Banking*, 2(4), 435-445.
- Rosenberg, E., & Gleit, A. (1994, August). Quantitative Methods in Credit Management: A Survey. *Journal of Operations Research*, 42(4), 589-613.
- Wiginton, J. C. (1980, September). A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior. *Journal of Financial and Quantitative Analysis*, 15(3), 757-770.