Collection score and the opportunities for non-performing loans market

Eric Bacconi Gonçalves
Universidade de São Paulo, Brasil
Maria Aparecida Gouvêa e-mail: magouvea@usp.br
Universidade de São Paulo, Brasil

Abstract
The target of this study is to develop a collection score model, in a sample of 254,914 clients of a Brazilian company specialized in non-performing loan portfolio, using Logistic Regression to identify the clients who have greater propensity to pay non-performing-loans. This paper presents, additionally, a suggestion of business application.

Key words: Non-performing loans, Collection Scoring, Logistic Regression, Statistical Models

INTRODUCTION

In 1994, after financial stabilization, the Brazilian market started making use of mass credit analysis models, evaluating large volumes of proposals automatically.

Brazilian financial institutions were already massively using credit scoring models for new clients, due to the currency stability, achieved with *Plano Real*, which began in mid-1994, and resulted in high growth rates in the volume of credit to consumers.

In addition to the models used to analyze the granting of new loans, known as credit scoring, there has also been an increased use of two other models: in the first model (behavior scoring model), the purpose is to evaluate whether bank clients are able to have new loans granted; and the second model (collection scoring) evaluates the likelihood of payment to be made by clients who are already in default and require collection action (Sadatrasoul et al. 2013).

This study aims to build a collection scoring model to a portfolio of non-performing loans (NPL), seeking to, by assessing the payment profile of each type of client, define the best collection strategies. In addition, we will also propose collection strategies to be adopted according to the profile of the client identified in the analysis.

LITERATURE REVIEW

Non-performing Loans

Non-performing loan is a provision for loans that are overdue for more than 90 days. An increase in the volume of non-performing loans in a financial institution leads to the risk of bankruptcy of this company (Makri et al. 2014).

Toledo (2013) points out that since the mid-90s, after the stability achieved with *Plano Real*, the Brazilian economy has been undergoing a process of growth leveraged by the increase in lending, and consequently, according to several authors, the rapid expansion has generated a worsening in the quality of lending, causing an increase in defaults (Kauko 2012; Makri et al. 2014; Barseghyan 2010; Lu et al. 2007), resulting in loans overdue for more than 90 days. Toledo (2013) points out that, between 2002 and 2012, the credit volume increased from 25% to approximately 50% of GDP.

Collection Policies

The duty of the Collection area is to bring money to the company's cash. The purpose of this area is to accelerate collections, causing the company to minimize its need for credit facilities (Gitman 2006).

The collection policies aim to define the various possible criteria and procedures to be adopted by a company seeking to receive the amounts receivable (Assaf Neto and Lima, 2011), that is, the company's strategy to receive the amounts receivable on their maturity date. The basic procedures used are: by letter, phone, in court, visits, among others (Machado and Barreto, 2011).

According to Hoji (2014), the collection policy should be implemented in conjunction with the credit policy. The granting of credit should not be too facilitated to subsequently require the application of rigidity in the collection, or vice versa. If the difficulty in collection is already expected in the act of granting credit, the credit scoring should be even stricter.

Scoring Models

According to Crook et al. (2007), scoring models are intended to measure the risk of a portfolio, during its term. The most common is to use logistic regression as a tool for building a model; however, researchers use other techniques, such as: Decision Trees (Olson et al. 2012), Neural Networks (Olson et al. 2012), Genetic Algorithms (Gouvêa et al. 2012) and Survival Analysis (Andreeva 2003).

Gouvêa et al. (2012) propose a seven-step cycle for building a credit scoring model that can be used for making any type of scoring model:

- <u>Surveying a historical customer base:</u>
  It is necessary to assume that the clients have the same pattern of behavior over time; based on that, past information are gathered for building the model. At this stage, it is necessary to define the target audience of the model, what information will be used and what frequency of data to be collected to build the model.
- <u>Classification of clients according to their pattern of behavior and definition of the response variable:</u>
- At this stage, the groups of clients to be modeled are defined. In general, two types of classification of clients are used, known as good debtors and bad debtors. In general, in addition to good and bad debtors, there may be, also, excluded clients (individuals who have particular characteristics and should not be considered such as, for example, individuals working in the institution) and indeterminate clients (those who are in the so-

called "gray area" and cannot yet be classified as good or bad, for example, new clients). Both in the market practice and in the academic papers, the trend is to work only with good and bad debtors (Olson et al. 2012).

- Selection of representative random sample from the historical base:
- To avoid any bias because of the size, it is important that the sampling is stratified equally in the pre-defined groups. The number of clients to be sampled depends on several factors, such as the size of the population and the ease of access to data, homogeneity of the population among others; however, Lewis (1992) proposes that with a sample of 1,500 clients for each type of response it is already possible to obtain robust results. Usually, studies work with two samples, the first for building the model and the second to validate and test the model.
- Descriptive analysis and data preparation:
  In this phase, each variable to be used in the model is analyzed with statistical criteria.
- Selection and application of the techniques to be used to build the model:
  In this study, we will use Logistic Regression. Gouvêa et al. (2012) conducted a literature review on the scoring models and identified the following techniques being used in these models: Linear Regression, Logistic Regression, Classification Trees, Linear Programming, Genetic Algorithms, Neural Networks, Discriminant Analysis and REAL. The results of academic studies confirm that there is no technique that proves to be always superior in relation to the others, since, depending on the data to be modeled, a technique may prevail over the others.
- Definition of criteria for the comparison of the models:
  In this step we determine the criteria for the comparison of the models; the most commonly used tools are the Gini coefficient, the ROC curve, the Kolmogorov-Smirnov (KS) test and the hit rate.
- Selection and Implementation of the best model:
  All areas involved should gather to determine the implementation plan: deadlines, stages and expected impacts should be clear to all individuals involved in order to avoid surprises along the process.

Collection Scoring Models

The collection scoring model is intended to estimate the probability of payment by clients who are already in default. This means that the target audience of the collection model consists of clients who failed to settle their obligations within the deadlines agreed with the creditors. This type of model is a tool that helps estimate the losses based on the probability of payment by clients who are already in default. Clients with different degrees of insolvencies are allocated into groups, separating those who need further collection action from those who do not need to be charged immediately (Sadatrasoul et al. 2013). Since in this case the model is built with clients who already have a relationship with the institution, the variables used in the modeling can be divided into two groups:

- Registration data: client's age, gender, marital status, address, etc. and information obtained from credit bureaus (protests, bad checks, disputes and financial constraints).
- Customer relationship with the company: late payment in previous months, length of relationship with the company, amount spent by the client with the company in previous transactions, previous contacts with the client, among others.

METHODOLOGICAL ASPECTS

Below we present some information regarding the development of this study; we used the SPSS software for Windows v.21.

Data

A company specialized in the collection of non-performing loans provided a sample of 254,914 individual clients, from a portfolio structured in May 2013, during a period of six months, and this sample only includes clients that the company has actually contacted. The clients who have not been contacted are not included in the sample due to the inability to classify them as good or bad debtors.

This type of company buys the portfolio from an institution (financial or not) at a lower price than the value of debt (in this study, the average price is 5% of the value of debt).

Definition of the Response Variable

The response variable defined will be based on the payment (or not) made by the client. Clients defined as good debtors are those who have accepted the agreement with the collection company and paid at least one installment of the amount agreed. The so-called bad debtors are defined as those who have not accepted any agreement or accepted, but breached their commitment by not paying any installment to the collection company.

Samples

Two samples were selected: one for building the model and one for validating the model. In the sample used for building the model, we selected 90,000 clients stratified by the response variable, with 45,000 clients deemed good debtors and 45,000 bad debtors; other clients remained in the sample of validation and test of the model, where we found a prevalence of good debtors.

Independent Variables

The available client registration variables, as well as the behavior variables observed were used to build the model. They are as follows:

- Client's Age
- Debt value
- Days in default
- Region of residence (North, Northeast, Midwest, Southeast and South)
- Number of residential phones in the registration
- Number of business phones in the registration
- Number of e-mails in the registration
- Number of previous contacts by telephone
- Number of previous contacts by e-mail
- Presence of restrictions on external credit bureau (protests, bad checks, Refin or Pefin)

- Score calculated by the external credit bureau
- Number of times that this client has appeared in a portfolio collected by this company.

All variables were categorized into ranges, turning into ordinal variables, in order to reduce the effect of outliers and make estimates more robust.

Logistic Regression

The Logistic Regression, as already mentioned, is the most widely used technique for this type of problem; it is based on the calculation of the probability of the client being classified in each one of the groups.

According to Gouvêa et al. (2012), there are three premises for the adoption of this technique, as follows:

- Absence of outliers: The outlier should be viewed from the perspective of how representative it may be in the population, and the researcher should evaluate whether it should be kept or eliminated, in case it exercises improper influence on the results.
- Low Multicollinearity: Multicollinearity means that the variables are not linearly independent. "High degrees of multicollinearity may cause the coefficients of independent variables to be erroneously estimated and even have the wrong signals." (Gouvêa et al. 2012)
- Sample size: The sample size should be adequate to allow for the generalization of the results, which can be verified with regard to the statistical significance of the tests. According to Hair et al. (2010), the minimum size recommended for the sample should be calculated in such a way that each group (Good and Bad) have at least 10 observations per predictor variable, and the total size of the sample should be above 400 observations.

For this study, we have categorized the variables seeking to reduce the effect of outliers; in order to prevent multicollinearity, the technique chosen for the selection of the variables of the Logistic Regression model was the forward stepwise; and the model was built with 90,000 cases, far above the volume proposed by Hair et al. (2010).

Performance Evaluation Criteria

The first performance evaluation criterion used was the selection of a validation sample; if the results of the validation sample are close to those of the development sample, it means that the model is appropriate to be used in other bases. Other two criteria will be used to evaluate the performance of the model: Hit rate and Kolmogorov-Smirnov Test.

Hit Rate

According to Crook et al. (2007), the hit rate is measured by dividing the total number of clients correctly classified by the number of clients who were part of the model. The same calculation must be done for each client group analyzed according to the model (Good, Bad), to understand whether the model is identifying a client type more accurately than others.

Hair et al. (2010) suggest to define the minimum acceptable hit rate the criterion of achieving a classification at least 25% better than the rate of accuracy achievable by chance alone; in this study, the probability of classifying any random client correctly by chance would be 50%; therefore, the minimum acceptable accuracy would be 62.5% (50% x 1.25); in case of different sample sizes, we should make the weighing, based on the largest group.

Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) test is a non-parametric statistical technique that aims to determine whether two samples are from the same population (Siegel 1975); in the case of this study, we seek to differentiate the good debtors from those classified as bad debtors. To apply this test, a cumulative frequency distribution is built for each sample of observations, using the same intervals for both distributions. For each interval one function is subtracted from the other. The test focuses on the largest observed deviation.

According to Crook et al. (2007), this is an important measure of separation; the higher the KS obtained in the model, the better the model is capable of distinguishing the bad debtors from good debtors.

RESULTS

Below, the results obtained in the processing of logistic regression are presented and analyzed. Finally, a proposal of an action plan for the manager of the collection company is formulated.

Logistic Regression

In this paper, initially, all variables are included to build the model; however, in the final logistic model, only some variables will be selected. The variables will be chosen using the forward stepwise method, the most widely used in logistic regression models (Norusis 2011).

The resulting model consists of 29 variables, and the most important variables for the classification of the client were the period in default, the classification of the external credit bureau and whether the client has been previously contacted by e-mail, as shown in Table 1.

Table 1: Variables in the equation

Variable	Estimated logistic coefficient (B)	Wald	Significance	Exp (B)
Overdue up to 360 days	1.833	99.210	.000	6.255
Overdue from 361-720 days	.659	267.340	.000	1.933
Overdue from 721-1080	.068	4.342	.037	1.071
Overdue from 1441-1800 days	-.213	31.447	.000	.808
Overdue above 1800 days	-2.162	2105.431	.000	.115
Client has been previously contacted by phone	.189	59.307	.000	1.208

Client has contact e-mail	1.327	3236.108	.000	3.771
Appeared more than once in the portfolio	-.271	57.587	.000	.763
Never appeared in the portfolio	.343	87.221	.000	1.409
Balance up to 500	.150	19.021	.000	1.162
Balance 1001 to 5000	-.214	68.014	.000	.807
Balance > 5000	-.952	787.678	.000	.386
Aged 18 to 40	.189	62.159	.000	1.208
Aged above 50	-.143	20.239	.000	.866
Presence of Restrictions in bureau	-.497	284.601	.000	.609
Client has 2 or more creditors	-.511	580.563	.000	.600
Client has been previously contacted by e-mail	1.363	1874.063	.000	3.909
Contact region	.209	36.770	.000	1.233
No external bureau score	2.870	6844.671	.000	17.642
External bureau score_range 1	-.411	65.871	.000	.663
External bureau score_range 3	.362	90.257	.000	1.436
External bureau score_range 4	.816	478.619	.000	2.263
External bureau score_range 5	1.558	1699.357	.000	4.749
Client has not informed telephone number	-.248	82.511	.000	.781
Client informed two or more telephone numbers	.215	95.250	.000	1.240
Client has not informed business telephone number	-.507	147.924	.000	.602
Client informed two or more business telephone numbers	.357	7.094	.008	1.429
Client has not informed home phone	-.122	25.312	.000	.885
Client informed two or more home phone numbers	.250	99.158	.000	1.284
Constant	-1.232	325.232	.000	.292

The Omnibus test measures whether the model is able to make predictions with the desired accuracy (O'Connell 2006; Menard 2002). The results of this analysis show that the significance test confirms that the model is able to properly make predictions.

Next, we tested the hit rate of the model. Table 2 shows that the hit rate of this model is 83.9% in the development sample, and 83.4% in the validation sample. The percentages of accuracy for good and bad debtors are close to each other and there is no change when changing from the development sample to the validation sample, which indicates a good result for the model.

Table 2: Hit rates

Sample			**Predicted**		**% hit**
			Bad	**Good**	
Development	Observed	Bad	38.495	6.505	85.5
		Good	7.968	37.032	82.3
	Total		46.463	43.537	83.9
Validation	Observed	Bad	51.317	8.721	85.5
		Good	18.712	86.164	82.2
	Total		70.029	94.885	83.4

According to Sicsú (2010), models with KS above 0.70 are deemed to have excellent discrimination, while models with KS between 0.60 and 0.70 have very good discrimination. For these data, the result of the KS test achieved for the development sample was 0.680, while in the validation sample it reached 0.679, indicating, just as the hit rate, that the results of the development samples are good and very close.

Proposed action

To propose an action, we will use the entire portfolio (covering the development and validation samples), where each client receives a score determined by the logistic model. The clients are divided into twenty equally sized ranges (each one with approximately 5% of the population); in each one of these ranges, the clients are highlighted as good or bad. If the model is well adjusted, the highest concentration of bad debtors will be in the lower ranges, while the so-called good debtors should be located more frequently in the higher ranges (Lewis 1992; Mays 2001). Table 3 below shows the distribution in the twenty ranges.

Table 3: Distribution of Good and Bad Debtors according to the score range

Score Range	**Good Debtors**	**Bad Debtors**	**Total in the Range**	**% of Good debtors within the Range**
Range 1	1,439	11,307	12,746	11.3%
Range 2	1,483	11,231	12,714	11.7%
Range 3	1,724	11,136	12,860	13.4%
Range 4	1,957	10,710	12,667	15.4%
Range 5	2,234	10,507	12,741	17.5%
Range 6	2,853	9,902	12,755	22.4%
Range 7	3,458	9,348	12,806	27.0%
Range 8	4,599	8,076	12,675	36.3%

Range 9	5,957	6,788	12,745	46.7%
Range 10	7,697	5,055	12,752	60.4%
Range 11	9,092	3,650	12,742	71.4%
Range 12	10,059	2,688	12,747	78.9%
Range 13	10,993	1,772	12,765	86.1%
Range 14	11,590	1,137	12,727	91.1%
Range 15	11,988	756	12,744	94.1%
Range 16	12,309	441	12,750	96.5%
Range 17	12,452	289	12,741	97.7%
Range 18	12,542	154	12,696	98.8%
Range 19	12,658	65	12,723	99.5%
Range 20	12,792	26	12,818	99.8%

The collection scoring model developed achieved a good division, since the percentage of good debtors increases at each range. Clients in the ranges between 14 and 20 can be approached with more flexible collection policies, such as discounts of lower value; on the other hand, clients between ranges 1 to 5, could be the focus of more aggressive collection policies (e.g. higher discounts).

FINAL CONSIDERATIONS

The purpose of this study was to adapt a collection scoring model, using logistic regression, to a portfolio of non-performing loans, and the results were appropriate.

This study presented a proposal on how to adapt the existing offers to the customer profiles identified by the model developed, since in a market with customized products, a model that allows for the differentiation of customer profiles is able to help managers to determine offers and targeted strategies according to their audience.

This study poses some limitations. The first was the use of secondary data provided by a company; so it is not possible to assure that all variables for the development of the model were made available; likewise, the clients had already been previously classified as good or bad debtors by the collection company. A second limitation was the low number of academic studies on collection scoring available in the literature; according to Sadatrasoul et al. (2013), the difficulty in obtaining data bases for this type of study inhibits the publication of further studies.

Future studies could focus on other techniques to develop models for this type of portfolio, such as, for example, neural networks or genetic algorithms; another opportunity to deepen the study is to understand more extensively the range of offers of the company and build a profitability projection in line with its existing policies.

Bibliography

Andreeva, G. 2003. *European generic scoring models using logistic regression and survival analysis,* Young OR Conference, April 2003, Bath.

Assaf Neto, A., F. G. Lima. 2011. *Curso de Administração Financeira*. Atlas, São Paulo.

Barseghyan, L. 2010. Non-performing loans, prospective bailouts , and Japan ' s slowdown. *Journal of Monetary Economics* **57**(7), 873–890.

Crook, J. N., D. B. Edelman, L. C. Thomas. 2007. Recent developments in consumer credit risk assessment. *European Journal of Operational Research* **183**(3), 1447–1465.

Gitman, L. J. 2006. *Principios de Administracao Financeira*. Bookman, Porto Alegre.

Gouvêa, M. A., E. B. Gonçalves, D. M. N. Mantovani. 2012. Aplicação de Regressão Logística e Algoritmos Genéticos na Análise de Risco de Crédito. *Revista Universo Contábil*, 84–102.

Gouvêa, M. A., L. C. Prearo, M. D. C. Romeiro. 2012. Avaliação da adequação de aplicação de técnicas multivariadas em estudos do comportamento do consumidor em teses e dissertações de duas instituições de ensino superior. *Revista de Administração* **47**(2), 338–355.

Hair, J. F., W. C. Black, B. J. Babin, R. E. Anderson. 2010. *Multivariate Data Analysis*. Pearson, São Paulo.

Hoji, M. 2014. *Administração Financeira e Orçamentária*. São Paulo.

Kauko, K. 2012. External deficits and non-performing loans in the recent financial crisis. *Economics Letters* **115**(2), 196–199.

Lewis, E. M. 1992. *An Introduction to Credit Scoring*. San Rafael. Available at https://books.google.com.br/books/about/An_Introduction_to_Credit_Scoring.html?id=XJKXGgAACAAJ&pgis=1

Lu, D., S. M. Thangavelu ,Q. Hu. 2007. *The Journal of Development Biased Lending and Non- performing Loans in China ' s Banking Sector*.

Machado, M. A. V., K. N. B. Barreto. 2011. Decisões Financeiras de Curto Prazo das Pequenas e Médias Empresas Industriais: um estudo exploratório. *SOCIEDADE, CONTABILIDADE E GESTÃO*. Available at http://www.atena.org.br/revista/ojs-2.2.3-08/index.php/ufrj/article/view/920

Makri, V., A.Tsagkanos, A. Bellas, A. Tsaganos, A. Bellas. 2014. Determinants of Non-Performing Loans : The Case of Eurozone. *PANOECONOMICUS* **77**(April 2013), 193–206.

Mays, E. 2001. *Handbook of Credit Scoring*. Chicago: Global Professional Publishi. Available at https://books.google.com/books?id=FTWntrkW77EC&pgis=1

Menard, S. 2002. *Applied Logistic Regression Analysis, Volume 106; Volume 2002*. Thousand Oaks: SAGE Publications. Available at https://books.google.com/books?id=EAI1QmUUsbUC&pgis=1

Norusis, M. 2011. *IBM SPSS Statistics 19 Guide to Data Analysis*. Pearson, São Paulo.

O'Connell, A. A. 2006. *Logistic Regression Models for Ordinal Response Variables, Issue 146*. Thousand Oaks: SAGE Publications.

Olson, D. L., D. Delen, Y. Meng. 2012. Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems* **52**(2), 464–473.

Sadatrasoul, S. M., M. R. Gholamian, M. Siami, Z. Hajimohammadi. 2013. Credit scoring in banks and financial institutions via data mining techniques : A literature review **1**(2), 119–129.

Siegel, S. 1975. *Estatística Não-Paramétrica Para as Ciências do Comportamento*. Mc Graw-Hill, São Paulo.

Sicsú, A. L. 2010. *Credit Scoring: desenvolvimento, implantação, acompanhamento*. Edgar Blucher, São Paulo.

Toledo, R. P. P. 2013. *Mercado brasileiro de non-performing loans (NPL): uma abordagem teórica e prática na precificação de ativos*. São Paulo. Dissertação (Mestrado) - Escola de Economia de São Paulo, Fundação Getulio Vargas, São Paulo, 2013.