We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 1

A Simple Fuzzy System Applied to Predict Default Rate

Fábio M. Soares, Olavo S. Rocha Neto and
Hevertton H. K. Barbosa

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/60079

1. Introduction

Fuzzy systems have been applied in a variety of problems with great success. One key factor is that the fuzzy rules database can be easily designed, in order to emulate a human rational decision making process just as experts usually do while facing hard jobs. Thus, in essence any process that requires human judgment can be translated into simple rules in a fuzzy system, provided that variables can be used in fuzzy sets or linguistic terms. One example is the prediction of a customer's default delay in payment, which seems to be a very simplistic and intuitive process and can be indeed modeled into a set of rules based on an expert's knowledge. In addition, the facility of implementing a fuzzy system can speed up the analysis of huge customer databases, since the usual manual process of analyzing each customer can be automated by a system which in theory has the same ability to infer as the human mind does.

This chapter shows the whole design of the fuzzy system to predict the customers' default rate in small and medium-sized businesses, and how this information can be used to provide a better cash flow estimate. The chapter is structured as follows: in section 2 we present the current economy scenario and why the default is a big problem; in section 3 we analyze some tools that are used to mitigate the risks; in section 4 we explain in details how the fuzzy approach can be exploited in this case; in section 5 we show the design of the fuzzy system; in section 6 we show some results from the simulations of this system; and finally in section 7 we discuss the results and conclude this chapter.

2. The default in the microeconomics

The default in the retail sector is a concerning problem in the modern world [1]. According to the formal definition, the default is a broader term. Technically it means any failure of some

INTECH
open science | open minds

© 2015 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

entity, natural or legal, to meet its legal obligations by not paying invoices of loan, services, bonds or wholesales [2]. The term default also applies to the failure of a government to repay its national debt; in that case it is national or sovereign default. In the case of customer default, the concerns are on rent, mortgage, consumer credits, utility payments or funding. While in the first case, the debt is related to the macroeconomic scenario, that means financial crisis over a whole country or continent, the latter is more related to customer's profiles and microeconomics. That led to the development of risk and credit analysis [3].

The default has a strong effect in developing companies, as well as in small and medium sized business. Mortgage and interest rates could be strongly affected by the customer default rate. Since the whole economy is tightly linked, the default represents a break in this chain, leading, in large scale, to a national level crisis.

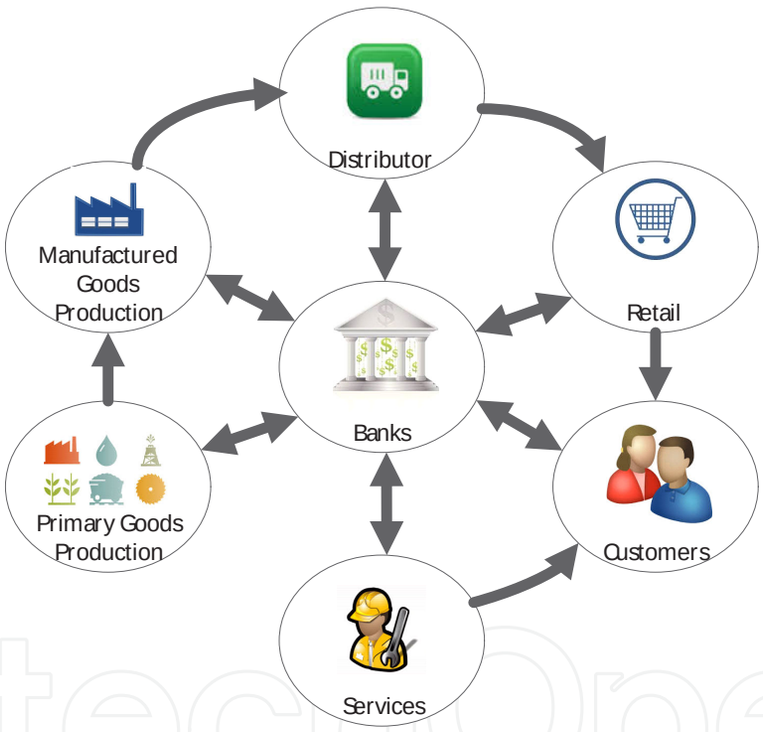


Figure 1. Economic Chain

When the default happens, it starts a shortcoming in a company's finance, and that means a loss for the provider. Some tools such as financial protection insurance and risk scores may remedy to a certain point, but in most cases they are insufficient to recover from the main problem [4]. However, if one could forecast or predict how many customers would delay payment or how much would be default, the companies would have the chance to prepare itself against a possible low cash flow.

Analysing per sector, it is known that service based industries are usually more affected by defaults.

http://dx.doi.org/10.5772/60079

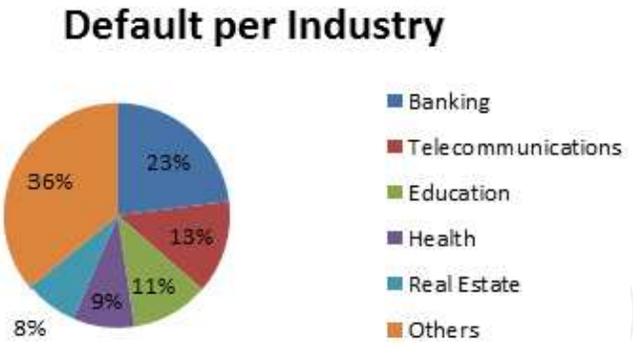


Figure 2. Defaults per Industry. Source: Serasa, 2013 [5]

2.1. Reasons for default

In order to better understand this problem, we should take into account the reasons for consumer default. The recent economic growing and integration speeded up the development of many enterprises, and much of this has been accomplished by the mechanisms of credit and financial leasing [6]. The credit offers expanded to small businesses and so have been to ordinary workers, thus increasing the economic activity [7]. In developing countries such as Brazil, many families from lower and middle classes turned out to actively participate in the economy as voracious consumers [4]. Consequently the debt of Brazilian families rose from 15% in 1992 to over 40% in 2012 [5]. Moreover, by analysing the reasons for the debt, it can be easily perceived that this index is not likely to be lowered, but limiting credit seems also not to be a good option [7]. However, the consumer default is correlated to some behaviours that can be detected in risk analysis systems [8]. Therefore by understanding the reasons for default, this problem can be more controllable.

Upon a report issued by Central Bank of Brazil [9], the causes for consumer default vary from bad financial habits (compulsivity, expenses greater than revenue) to financial problems (unemployment, little wages, crises, default from their clients).

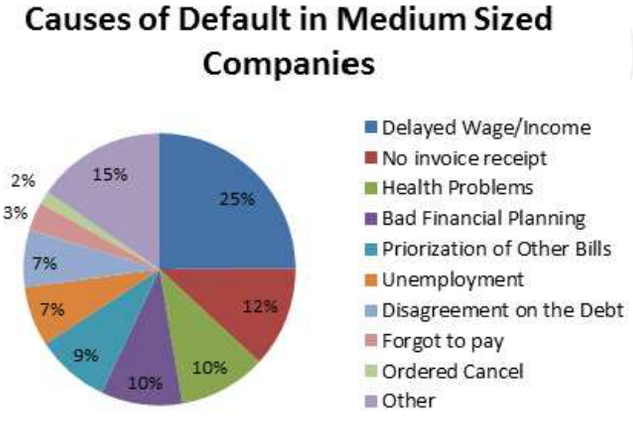


Figure 3. Causes for Defaults. Source: Annibal, 2009

Risk Analysis tools take into account as much information as possible from clients in order to evaluate the risk score of a given client. It is known that when a customer faces problems, he or she is more likely to overdue the bills or even not to pay at all. Likewise when customers always pay their bills without delay, it is a good sign they are less likely to delay.

2.2. The effects of default in the economy

The customer default prediction remains a concern for many enterprises, owners, investors and business men all over the world. Every default implies that some party is losing money, since a good or a service has been offered for free without any compensation. In large scale this leads to economy shrinking and inflation [10]. For small companies the effects may be even more drastic due to its small budget. Every small business is oriented to observe its cash flow, but in fact when a customer fails to pay its debt to the company, the cash flow accuracy is severely affected. So there is a need to estimate a percentage of default from its clients. The first and main consequence for small business is that it may not be able to meet its obligations, although it has a quite good financial planning. A second consequence is that the billing department will be overloaded since many bills remain unpaid and the company itself may fail to operate.

3. Existing tools for mitigating the risks

The defaults cannot be prevented, but can be forecasted. Whenever a lender wants to issue credits to a borrower, he may perform an analysis on the financial statements of the borrower in order to assess its capability to comply with its debts [11]. However many aspects may be not considered in traditional risk and credit analysis, since many risk analysis are conducted by means of likelihood and probabilities [2]. On the other hand, there is a call for simpler and quicker risk analysis [12, 13].

Although most of these tools address the credit analysis in the form of a loan, the same procedure applies to any customer that is buying goods/services from a supplier [14], especially if it is in the form of leasing or even contracting. Since we are dealing here with the problem of predicting default, our goal is to forecast when a customer will not pay his/debt causing a default. To that end, the methodologies fall particularly on risk and credit analysis, bankruptcy prediction and probability of default.

3.1. Risk and credit analysis

In recent decades, a number of objective, quantitative systems for scoring credits have been developed. The risk of credit is assessed by comparison of accounting ratios of potential borrowers with industry or trends in the financial variables. The banks are provided with many of these ratios, since they are the main credit providers, but that information is not always available to enterprises. Traditional credit risk analyses are implemented in expensive expert systems whose development is very time-consuming. On the other hand simpler forms to grant

http://dx.doi.org/10.5772/60079

credit may be achieved by the use of reduced models, such as Balanced Scorecard, Jarrow-Turnbull, among others [15].

Balanced Scorecard, also known as BSC, is actually a management technique aimed at assessing an enterprise's performance from four perspectives: financial, customer, internal processes, and learning. A balanced score of these indicators makes a system that helps the enterprise to select and focus strategies to achieve goals in the near future. The customer and financial perspective of this analysis composes a good index to evaluate the risk of servicing a given client [12]. But unfortunately this is not always enough to define strategy, and there should be also other performance indicators to determine the risk in a more accurate way.

A reduced form of risk model was published by [16] which is an extension of the Merton model [17] to a random interest rates framework. In this model, risk is modeled as a statistical process. The value of risk is evaluated using a continuous probability of default, estimated in two approaches: Deriving Point in Time (PIT) or Through the Cycle (TTC). The main difference between these approaches regards to internal and external factors. The term PIT applies to probabilities of default that are dependent of general credit conditions or external factors, while TTC applies to probabilities of default that are not subjected to external factors [18].

PIT Factors	TTC Factors
GDP Growth Rates	Revenue Growth
House Price Indices	Number of Default Cases
Unemployment	Load to Value Ratio

Table 1. Factors Analyzed in PIT and TTC Probabilities

3.2. Bankruptcy prediction

One benefit of risk analysis is that it allows the prediction of bankruptcy for a given entity. One of the oldest methods for bankruptcy prediction was published in 1968 by Altman. His formula is used to predict the probability of bankruptcy within 2 years by using Z-scores. The Z-score is a linear combination of four or five coefficient-weighted common business ratios.

$$Z = 1.2T_1 + 1.4T_2 + 3.3T_3 + 0.6T_4 + 0.99T_5 \quad (1)$$

where:

T_1 is the Working Capital / Total Assets

T_2 is the Retained Earnings / Total Assets

T_3 is the Earnings Before Interest and Taxes / Total Assets

T_4 is the Market Value of Equity / Book Value of Liabilities

T_5 is the Sales or Revenue / Total Assets

Z is the score which denotes where an entity will face bankruptcy or not. The Bankruptcy threshold varies on the entity's activity, but in general it is defined as follows

$$\begin{gathered} Z > 2.99 - \text{Non Bankrupt} \\ 1.81 < Z < 2.99 - \text{Gray Zone} \\ Z < 1.81 - \text{Bankrupt} \end{gathered} \tag{2}$$

Altman Z-Score [19] model was found to be 72% accurate in predicting bankruptcy two years before the event with only 6% of false negatives. It is still well accepted by auditors, management accountants and financial directors for load evaluation. However, this model is not recommended for use with financial companies such as banks or factoring, because the balance sheets of companies are usually opaque and the model does not address off-balance sheet items. For prediction of default for financial companies, the Merton Model is used.

Although additional methods for bankruptcy prediction have been developed by taking into account more data, their practicability turned out to be expensive, since it depends on a lot of data to be collected [20].

3.3. Probability of default

Given that many methods of risk and credit analysis, and bankruptcy prediction are based on stochastic models, we are now focusing on the measures for evaluating the probability of default. Most of methods exploit logistic regression functions as well as inversed probability distribution formulas.

The Probability of Default may be used in two ways: to address the causes of default; to predict and prevent new cases of default. Camargos et al [7] performed a survey to find conditioning factors that lead small business to default, as depicted in figure 4.

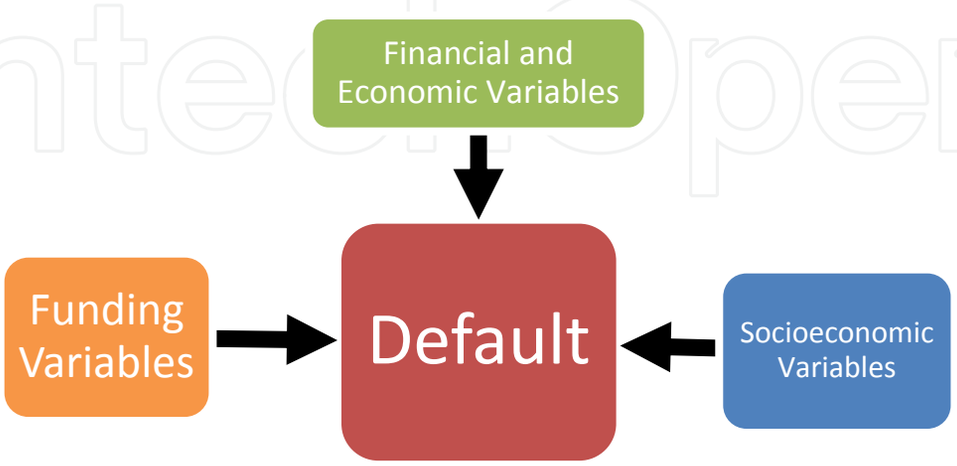


Figure 4. Variables that influence the default according to [7]

http://dx.doi.org/10.5772/60079

This survey has been conducted in an important Brazilian Program for encouragement of entrepreneurship among small-sized businesses. The method used to assess the risk of default was the logistic regression:

$$P(X_1) = \frac{1}{1 + e^{-X_1}} \tag{3}$$

The equation has only one dependent variable X_1, as the variable influencing on defaut. A threshold value of 0.5 is selected to determine whether a case is to be classified as compliant or default. Considering the probability as an input for the binary logistic regression variable Y, and then rearranging the coefficients, we obtain a linear logarithmic model:

$$\begin{aligned} Y(X_1) &= \frac{P(X_1)}{1 - P(X_1)} \\ Y(X_1) &= e^{\beta_0 + \beta_1 X_1} \\ \ln Y(X_1) &= \beta_0 + \beta_1 X_1 \\ \ln\left(\frac{P(X)}{1 - P(X)}\right) &= \beta_0 + \sum \beta_i X_i \end{aligned} \tag{4}$$

where:

P(X) is the probability of default according to the set of variables X

β_0 is a model bias constant

β_i is coefficient for the variable Xi

X_i variable taken into account in the model

Other models include bivariate probit model by Jacobson and Roszbach [21], to estimate default probabilities and the effects of default-risk-based acceptance rule changes on a bank's portfolio. Katchova and Barry [6] used the distance-to-default approach to determine the Value at Risk (VaR). All these models use logistic regression functions on multiple variables. By investigating these models amongst others, Odeh et al [22] applied a conceptual model for predicting default in agricultural loans, assuming the expected loss is expressed as a result of three components.

$$EL = PD*LGD*EAD \tag{5}$$

where

EL is the expected loss in monetary units

PD is the probability of default in percentages

LGD is the percentage of loss from the loan volume suffered by the granting institution

EAD is the loan amount plus accrued fees

Usually the Probability of Default is expressed in terms of N customers, so the equation 5 can be rearranged in the form:

$$EL^P = \frac{\sum_i^N PD_i}{N} * LGD * EAD \tag{6}$$

where now

EL^P is the expected loss on a specific portfolio

PD_i is the probability of default for a specific loan

N is the total of granted loans

Combining the logistic regression (eq. 4) with the conceptual model (eq. 6), we can express the maximum likelihood estimation as in the equation:

$$\ln\left(\frac{PD_i}{1-PD_i}\right) = B'X + \varepsilon \tag{7}$$

where

PD_i is the probability of default as stated in the equation 6

B is a vector of coefficients

X is a vector of explanatory variables and ε is a stochastic error

The coefficients may be determined empirically and vary from many aspects taken from the enterprise's assets. Odeh et al [22] evaluated these methods by using data from Farm Credit System, and found that credit default predictions are really sensitive on data.

3.4. Recent approaches

One of the recent technologies that has evolved and been used are the expert systems. Not only they have been used considerably since the 1980's in financial institutions for decision making tasks, the prediction of default has also been an issue the experts systems have been used for [23]. In addition, computing intelligence techniques, such as Genetic Algorithms, Fuzzy C-Means, and Mars, have also been exploited [24] due to its capability of learning from an expert. The use of neural networks, neuro-fuzzy and fuzzy logic has also grown in recent decades,

http://dx.doi.org/10.5772/60079

because they better handle on imprecise information and there is no pure analytical model of the market [25].

Furthermore, the database containing hundreds of financial operations represent an implicit knowledge that is available for modeling and prediction. By means of data mining [14], many customer behaviors can be analyzed based on past values. Thus, more reliable and developed models can be accomplished by the use of artificial intelligence.

4. A fuzzy approach

Fuzzy Systems have already been used in a variety of problems, not only regarding risk and credit analysis, but also bankruptcy and default prediction. A Fuzzy approach combines an easy design fully based both on an expert's opinion and on data history. Zirakja and Samizadeh [8] performed a risk analysis in e-commerce (EC) activities in a more broad vision, including the projects' risk, by relying on experts' opinions to build a fuzzy decision support system (FDSS). Martin et al [24] implemented a fuzzy system to predict bankruptcy by using expert knowledge applied in fuzzy rules with a classification rate of 88% in a single model. In a hybrid model, by using neuro-fuzzy and genetic algorithm, the classification rate was 73,6% but with more input variables.

Fuzzy logic arises as a good tool to emulate expert rules since they don't require too much effort for modeling as other traditional methods do. A fuzzy system can emulate rules of type:

$$IF\left\{condition\right\}THEN\left\{consequence\right\} \tag{8}$$

where conditions and consequences are fuzzy propositions built by linguistic expressions:

1. x is Low
2. y is NOT Tall
3. x is Low AND y is Tall
4. x is Low OR y is Tall

The expressions 1 and 2 define "immediate" propositions, and the expressions 3 and 4 define combined propositions. Since they operate over fuzzy variables, they need to be defined in linguistic terms or fuzzy sets. Fuzzy sets usually take the form of membership functions.

$$\mu_A\left(x\right)=\begin{cases}0\,if\ x\,does\,not\,belong\,to\,set\ A\\ y\,if\ x\,partially\,belongs\,to\,set\ A\\ 1\,if\ x\ fully\,belongs\,to\,set\ A\end{cases} \tag{9}$$

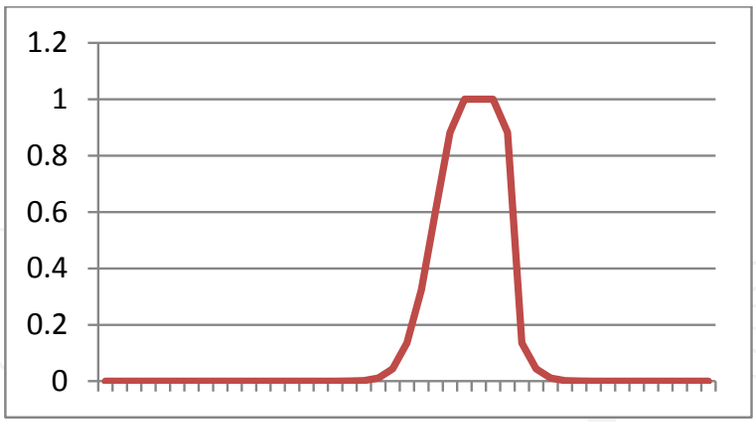


Figure 5. Example of a membership function plot.

Fuzzy expressions are built using boolean operators such as NOT, OR and AND. These expressions are combined to form relations R. A fuzzy relation is defined against two universes U and V, as U x V being a subset of the Cartesian product of those, so that R: UxV → {0,1}. That means, given A a fuzzy relation and x and y crisp values, if x

∈ U and y

∈ V then R(x,y)=1, otherwise R(x,y)=0.

Therefore, Fuzzy rules can be defined in fuzzy operations as in the equation.

$$R^{(l)} : If\ x_1\ is\ A_1^l\ AND...AND\ x_n is\ A_n^l\ THEN\ y\ is\ B^l \tag{10}$$

where

$R^{(l)}$ is a Fuzzy rule of index l

x_i is an input fuzzy variable of index i

A_i^l is an input fuzzy set of index i in a rule l

y is an output fuzzy variable

B^l is an output fuzzy set in a rule l

which in turn can be represented by membership functions

$$\mu_{R^{(l)}}(X) = \min\left(\mu_{A_1^l}(x_1),...,\mu_{A_n^l}(x_n)\right) \tag{11}$$

$$\mu_B(y) = \max_{l=1}^{m}\left[\sup_{X\in U}\min\left(\mu_{R^{(l)}}(X),\mu_{B^l}(y)\right)\right] \tag{12}$$

http://dx.doi.org/10.5772/60079

where

$\mu_R^{(l)}(X)$ is the membership function of the rule

$\mu_{A1}^{l}(x_i)$ is the membership function of the input variable of index i on fuzzy set A_i^l

$\mu_B(y)$ is the resulting membership function of the output variable y on fuzzy set B in rule l

min is the minimum operator

max is the maximum operator

sup is the supremum operator

4.1. Fuzzy system structure

A Fuzzy System usually has:

- **Input Variables** (with their respective Fuzzy datasets);
- **Output Variables** (the diagnostics values);
- **Rule Base:** determines outputs for each combination of input fuzzy values;
- **Inference Machine:** applies fuzzy operations;
- **Fuzzy Sets:** Linguistic Terms for each Variable;
- **Crisp Values:** Numeric values taken from real world.

Figure 6 shows the structure of a basic model of fuzzy system, consisting of four components: Input Fuzzification, Rule Database, Inference Machine and Defuzzification.

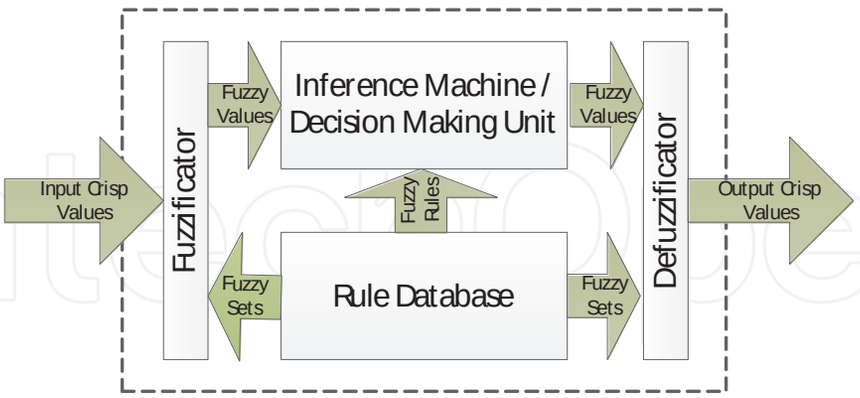


Figure 6. Fuzzy System Structure

A Fuzzy system can be defined in the following operations:

- **Input Fuzzyfication:** transform the real world crisp values into fuzzy values.
- **Fuzzy Operation**: Applies Fuzzy Operators Min or Max in input Variables according to available rules if they should be inclusive (AND) or exclusive (OR).

- **Aggregation:** These operators can group several found output values provided that several rules may have triggered.
- **Defuzzification:** transform the output found fuzzy values into real world crisp values.

In this chapter we are dealing with the application of a fuzzy system in predicting the default, so the details on these operations are beyond the scope, and for further information the reader is suggested with the references [26, 27].

4.2. Reasons to apply fuzzy logic

Fuzzy systems are relatively simple to create and deploy, and it is fully based on human experts' evaluation. Fuzzy has been applied in many fields involving decision processes which require some sort of judgment. The human mind abstracts real world variables in an imprecise manner forming semantic networks [28]. These semantic networks define relations that can be expressed with linguistic terms just as experts do. Therefore any activity requiring an expert opinion or judgment can be modeled in fuzzy logic rules without the need of an existing theoretic model to lie upon.

In small and medium-sized companies, the financial/collect department usually takes decisions regarding granting credit or not. Without any supporting tool, the decision is taken purely by an expert's experience or opinion. The same applies for predicting cash flow based on client's past financial transactions. Based on a given customer's history, it can be inferred whether this customer will pay on time or default. This kind of analysis can be performed by an expert, but as a company's portfolio grows, the task of analyzing becomes more time-consuming and then needs to be automated, and fuzzy systems emerge as a good option to automate this type of analysis [29].

5. Fuzzy system development

By taking into account all the previous information, we designed a system capable of predicting the default rate based on historical records of customers. The methodology used in this design was the same used in the work of [27], which consisted of the following procedure.

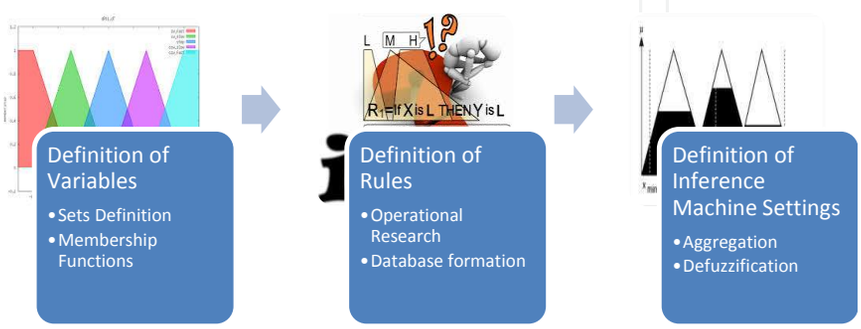


Figure 7. Fuzzy Modelling Procedure

http://dx.doi.org/10.5772/60079

According to literature, the default is influenced by many aspects of the customers, but many of them are unknown to the provider, unless they are declared. However, simple models of probability of default can be able to yield good results using statistical measures. So, to make this system more applicable, we took into account only the minimum amount of information a collection or billing department would have regarding customers' transactions. Thus in this work, we considered the database consisting only of customer invoices in the form of table.

Id	Client	Description	Value	Due Date	Payment
23129	Incs. Co.	Maintenance apr/12	99.00	20/04/12	24/04/12
23137	Sol. Llc.	Development apr/12	1342.00	25/04/12	23/04/12
23144	White Ss.	Fin. SSAS fee apr/12	49.90	11/04/12	09/06/12
...	...	...	...	...	...

Table 2. Accounting Records Database

5.1. Fuzzy variables

According to the database depicted in table 2, we defined the following input variables for the fuzzy system.

- Average Payment Delay (APD)
- Amount Owed (AO)
- Maximum Payment Delay (MPD)
- Maximum Amount Owed (MAO)
- Time as a Client (TC)
- Number of Default Cases (NDC)

A formal definition of each variable is outlined in the following equations:

$$APD_i = E\left[PD_i^j - DD_i^j \right] \tag{13}$$

$$MPD_i = \max\left[PD_i^j - DD_i^j \right] \tag{14}$$

$$AO_i = \sum_j^N AO_i^j \ \forall \ PD_i^j \ is\ null\ and\ DD_i^j < t \tag{15}$$

$$MAO_i = max\left[AO_i \right] \ \forall PD_i^j > DD_i^j \tag{16}$$

$$TC_i = t - \min\left[D_i^j \right] \tag{17}$$

$$NDC = \sum_{j}^{N} 1 \; \forall PD_i^j > DD_i^j \; and \; DD_i^j < t \tag{18}$$

where

PDij is the Payment Date of the Invoice j of the Client i

DDij is the Due Date of the Invoice j of the Client i

N is the number of issued Invoices

t is the current Date

For specific purposes of this work, a default is considered to be when an invoice is not paid before the due date.

Upon consultation with experts in the collect department, we found the following terms for each of the input variables.

Input Variables		**Linguistic terms**
APD	Average Payment Delay	Short, Middle, Long
AO	Amount Owed	Low, Middle, High
MPD	Maximum Payment Delay	Short, Middle, Long
MAO	Maximum Amount Owed	Low, Middle, High
TC	Time as a Client	New, Known, Old Known
NDC	Number of Default Cases	Low, Middle, High

Table 3. Fuzzy Input Variables and their Linguistic Terms

The output variables are the values we want to predict, namely when and how much is customer going to pay. That can be express in two ways: Expected Amount/Date of receipt; Probability of Receiving a certain amount within a period of time. Since here we are considering only internal factors, this kind of prediction is through the cycle (TTC). According to Basel II Parameters [10], the simplest approach to estimate the probability of default is logistic regression, taking historical database as a basis for estimation.

Thus, given a date, the probability distribution of payment can be expressed by the following equation:

$$PDR = A^{*}e^{(B^{*}EDR)-(C/NDC)-(D^{*}EAR)-1} \tag{19}$$

http://dx.doi.org/10.5772/60079

where

PDR is the Probability of Receipt or Payment

EDR is the Expected Date of Receipt (in days)

EAR is the Expected Amount to Receive (in monetary units)

NDC is the Number of Default Cases

A, B, C and D are coefficients

Upon experiments and linear regression we found the coefficients to be.

$$\begin{aligned} A &= 13.338 \\ B &= 1.6231 \\ C &= 0.875 \\ D &= 22.81 \end{aligned} \tag{20}$$

It can be seen that there is a relation between the next payment date and the probability. So the output variables were chosen to be the next payment and expected amount to be paid.

Output Variables		Linguistic terms
EAR	Expected Amount to Receive	None, Little, Enough, Integral
EDR	Expected Date of Receipt	Near, Reasonably Near, Far, Never

Table 4. Expected Output Variables

However, from these outputs the probability of receiving over time t can also be derived, according to the equation.

$$PP(t) = \int_0^t PDR \, dEDR = \frac{t}{EDR} e^{-\frac{t}{EDR}+E} \tag{21}$$

where

PP(t) is the probability of payment over time t

PDR is the probability distribution function

EDR is the expected Date of Receipt

E is the remaining part of the probability distribution function PDR, independent of EDR

Thus, we can state the variables PPW and PPM with parameter values for t of 7 and 30, respectively. The probabilities can also be defined in fuzzy sets.

Output Variables		Linguistic terms
PPW	Probability of Payment in a Week	Null, Very Low, Low, Medium, High, Very High
PPM	Probability of Payment in a Month	Null, Very Low, Low, Medium, High, Very High

Table 5. Probability Output Variables

Likewise, the expected date of payment can be derived from the quantile equation, which is the inverted probability density function.

$$ER\left(PP\left(t\right)\right) = \frac{-\ln\left(1 - PP\left(t\right)\right)}{\frac{1}{1+EDR} + E} \tag{22}$$

where ER is the Expected date of payment resulted from the probability distribution PP(t).

5.2. Fuzzy set limits

We defined the fuzzy set limits upon querying against a huge database containing over 5 years of financial records, in such way that each set should have the same number of clients belonging to it. To that end, we had to rearrange the database to group the results per client.

Client	Average Delay	Maximum Delay	Amount Owed	Maximum Owed	Number of Cases	Time as Client
Incs. Co.	5.666	18	0.00	150.00	24	235
Sol. Llc.	0.2222	12	230.00	15500.00	4	346
White Ss.	18.5426	128	57.50	6500.50	16	1448
...	...	...	...	...	...	...

Table 6. Accounting Database records grouped by each client

We defined the Gaussian function as a membership function for each set, on input and output. After querying the dataset, we defined the sets' limits as can be show in the table and figures.

$$\mu_A\left(x;c;\sigma\right) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \tag{23}$$

where c is the center of the function, and σ is the variance. Then, we defined as the set's limits c±σ.

http://dx.doi.org/10.5772/60079

Variable	Low/Short/New		Middle/Known		High/Long/Old Known	
	Inf.Lim.	Sup.Lim.	Inf.Lim.	Sup.Lim.	Inf.Lim.	Sup.Lim.
APD	0	13.91	13.91	34.91	34.91	222
AO	0	272.66	272.66	3723.53	3723.53	50084
MPD	0	100.45	100.45	200.45	200.45	1009
MAO	0	845.00	845.00	5593.75	5593.75	50084
TC	0	234	234	1304	1304	2233
NDC	1	3	3	10	10	38

Table 7. Sets' limits defined upon database querying

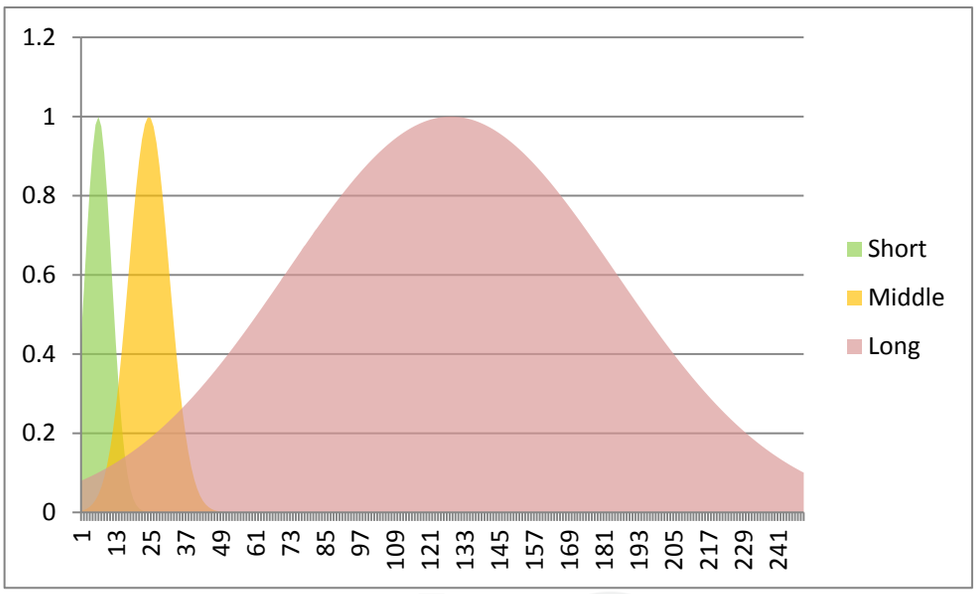


Figure 8. Fuzzy Sets' Plots of Variable Average Payment Date

5.3. Fuzzy rules

As performed in the work of [27], we have built the fuzzy rules upon querying the database shown in table 6 for each combination of the input sets. That would give 729 rules. But before querying a database, we cut some combinations that would never happen in practice or could be intuitively disposed. Some examples are the following rules:

If APD is long and MPD is short and...

If AO is high and MAO is short and...

If TC is new and NDC is high and...

By cutting infeasible rules, the rules database has been reduced to 288 rules. The outputs for each rule, both for expected date and amount of receipt, have been determined upon querying the history database. Nevertheless, some situations never happened, so we had to decide the output for these rules by asking the experts. That procedure trimmed down the rule database to only 53 rules

For a given rule, we found an average difference between the due date and the payment date.

$$APD_i = E\left[PD_i^j - DD_i^j \right] \ \forall DD_i^j > t \tag{24}$$

$$STD_i = \sqrt{E\left[\left(PD_i^j - DD_i^j \right)^2 \right] - APD_i^2} \tag{25}$$

where

APDi is the average payment date for the client i;

STDi is the standard deviation for the difference between due dates and payment dates of the client i.

After querying the database for any given rule, we built out a histogram of each fuzzy output variable corresponding to that rule. Table 8 shows a histogram found for the following rule:

"if **APD** is *Middle* and **AO** is *Low* and **MPD** is *Long* and **MAO** is *Low* and **TC** is *Old Known* and **NDC** is *Low*"

	None/Near	**Little/Reasonably Near**	**Enough/Far**	**Integral/Never**
Expected Date of Receipt	3	2	1	0
Expected Amount of Receipt	0	0	1	5

Table 8. Histogram for a given rule

The output set was chosen as the one that the rule result better fits into, which is *Near* for **Expected Date of Receipt** and *Integral* for **Expected Amount of Receipt**.

5.4. Database preparation

In order to have a separation between the rules database development and the validation, we defined distinct periods for querying and for validation. The database in the form shown in table 6 has been cut into these two periods, 2 years and a half each, forming a new database grouped by period. This database is shown in table 9.

Client	Period	Average Delay	Maximum Delay	Amount Owed	Maximum Owed	Number of Cases	Time as Client
Incs. Co.	1	5.666	18	0.00	150.00	24	235
Incs. Co.	2	7.183	18	120.00	250.00	33	622
Sol. Llc.	1	0.2222	12	230.00	15500.00	4	346
Sol. Llc.	2	0.3333	12	230.00	15500.50	4	733
...	...	...	...	...	...	...	...

Table 9. Database split into 2 periods

The validation period was replicated multiple times in order to perform a continuous validation from first until the last date of the period. For each date, a snapshot of the database of table 6 was taken in order to simulate the Fuzzy Prediction.



Figure 9. Schema of database snapshots taken for every date in the periods for simulation

Current Date	Id	Client	Description	Value	Due Date	Payment
...	...	...	...	...	...	...
23/04/12	23129	Incs. Co.	Maintenance apr/12	99.00	20/04/12	
23/04/12	23137	Sol. Llc.	Development apr/12	1342.00	25/04/12	23/04/12
23/04/12	23144	White Ss.	Fin. SSAS fee apr/12	49.90	11/04/12	
...	...	...	...	...	...	...
24/04/12	23129	Incs. Co.	Maintenance apr/12	99.00	20/04/12	24/04/12
24/04/12	23137	Sol. Llc.	Development apr/12	1342.00	25/04/12	23/04/12
24/04/12	23144	White Ss.	Fin. SSAS fee apr/12	49.90	11/04/12	
...	...	...	...	...	...	...

Table 10. Database with time dimension

5.5. Further fuzzy settings

The fuzzy system was implemented using Mamdani [26] as the inference machine, because of its simplicity in processing the rules and values and ease to be implemented in this case. Then it was deployed on an important Brazilian Financial Accounting System whose aim was to infer how much of the accounts receivable could be received within a week or a month, and what would be the default rate. The fuzzy system was set up as follows.

Input Fuzzy Sets	*Gaussian*
Output Fuzzy Sets	*Triangle*
Implication Method	*AND/OR*
Aggregation Method	*Product*
Defuzzification Method	*Centre of gravity*

Table 11. Fuzzy Settings

6. Results and simulation

After defining and validating the rules database, we performed a simulation of prediction the default rate in a period of 2 years and a half. Since we have a probability as an output, we had to apply the Monte Carlo method to generate random numbers and get real results from the simulations and confront them against the real values [30].

6.1. Simulation procedure

We used the database shown in table 10 to perform simulations on any record for every invoice which was supposed to be paid. The fuzzy system would give an expected date and amount to be received. So we applied for a given record some calculations using equations 19, 21 and 22 to get probabilities of payment within a day, a week and a month. With the Monte Carlo method, we have gotten a number of random values to be applied in the probability distribution as shown in the equations 21 and 22. If that random number would be greater less than the corresponding probability value, calculated in the equations 21 and 22, it means the debt has been paid.

The algorithm for the simulation was defined as follows.

Then, we performed simulations to predict:

- the default rate of customers
- the future revenue within a short period (a month)
- the revenue within a long term (a year)

http://dx.doi.org/10.5772/60079

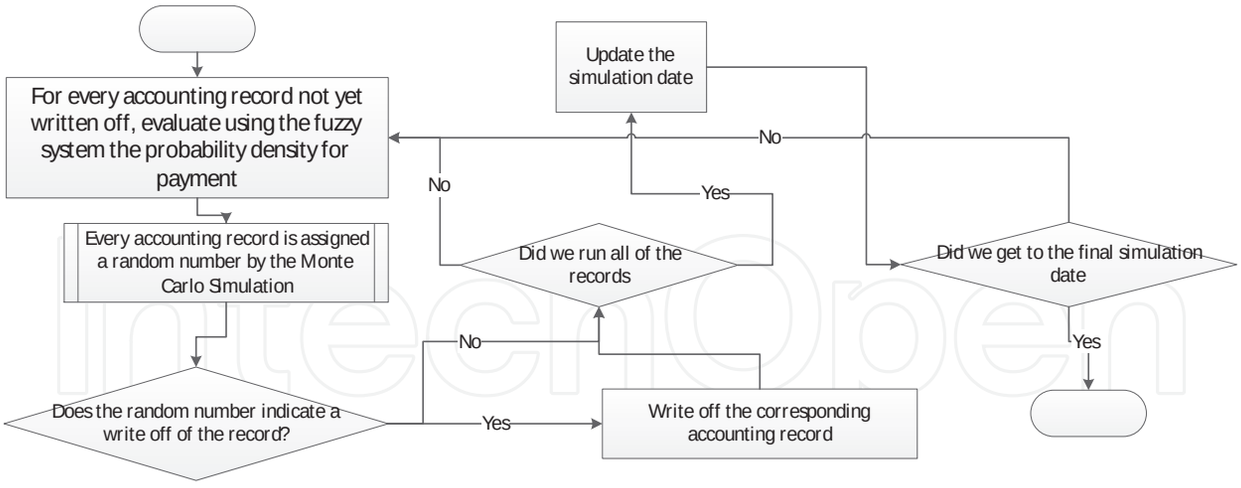


Figure 10. Flowchart of the simulation

6.2. Prediction of the default rate

The default rate is assumed to be the percentage of invoices that are delayed or paid after the due date:

$$DR(t) = \frac{\sum AO_i \ \forall PD_i > DD_i \ and \ DD_i < t}{\sum AO_i \ \forall DD_i < t} \tag{26}$$

where DR(t) is the default rate at the date t.

Then we applied the Fuzzy system to give an expected percentage of invoices that were about to be paid after due date, and compared to what happened in fact. We repeated the experiments 100 times, in order to have more accurate values. The results are outlined in table 12.

Id	Client	Amount Received	Due Date	Payment Date	Amount Estimated	Estimated Payment Date
...	...	...	...	...	...	...
19027	York S.	195.85	03/12/12	03/12/12	**190.0±15.0**	46±40 days after
19028	Bennetts	314.25	20/12/12	01/02/13	**330.0±20.0**	12±4 days after
19033	Houth	160.00	11/10/12	10/10/12	**150.0±10.0**	2±1 days after
19041	York S.	195.00	19/09/12	03/12/12	**190.0±15.0**	**46±40 days after**
19045	N. Shots	149.00	06/09/12	13/09/12	180.0±20.0	**7±5 days after**
19048	Net Sol.	1081.22	11/11/12	12/11/12	**1050.0±100**	6±1 days after
19062	Yamada	225.00	09/12/12	05/04/13	200.0±20.0	68±43 days after
...	...	...	...	...	...	...

Table 12. Some results from predictions. Correct estimations are bolded.

The following plots show how the default rate is predicted from the Fuzzy system over time. One data series is the actual default rate per month, and the other is the average prediction after simulating a 100 times.

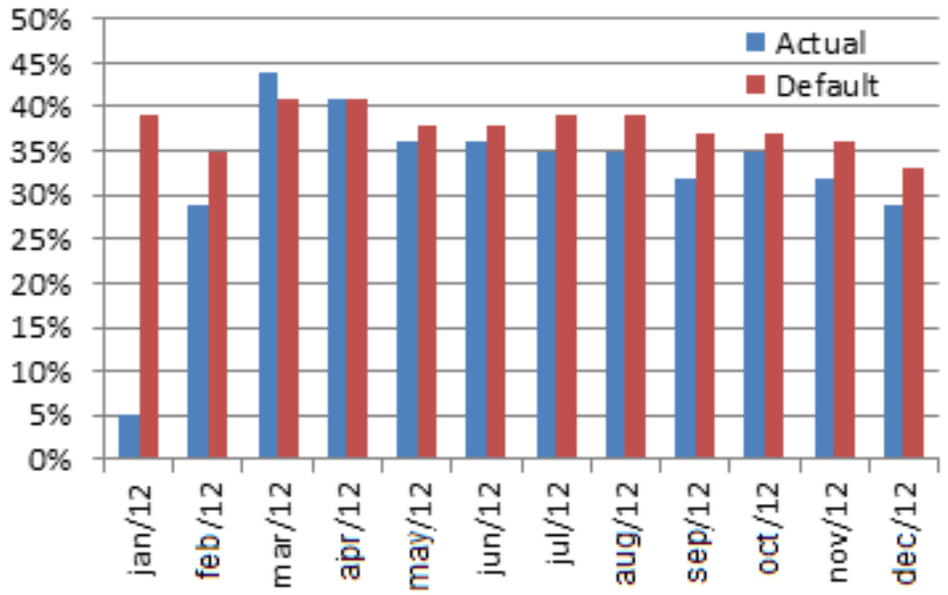


Figure 11. Prediction of the Default Rate

By applying this procedure, one can infer the revenue over a period, using the expected amount to be paid as the output value in addition to the expected date of payment. The expected revenue is then set to be:

$$PER_i = PEPR_i + DR_i * ER_i \tag{27}$$

where

PERi is the Predicted revenue for the period i

PEPRi is the Predicted revenue from past periods of period i

DRi is the default rate for period i

ERi is the original expected revenue for period i

6.3. Forecasting revenue

We processed the results from the default rate prediction and built several snapshots out of the simulations to forecast the revenue over each period by taking into account recent real records up to the current period.

The results are outlined in the plots shown in figure 12.

http://dx.doi.org/10.5772/60079

6.4. Long term simulations

To calculate how much the enterprise expects to receive in the long term, we performed random simulations on the probabilities given by the fuzzy system for the whole period. By checking the history for each client until the present moment in the simulation, and an estimative of when this client will pay is obtained. For validation we compared the predicted default rate against the real default rate that happened in the period. The system was validated with a 12 month period simulation using past values for 100 times. This strategy was able to give a prediction of the default rate with an 80% accuracy.

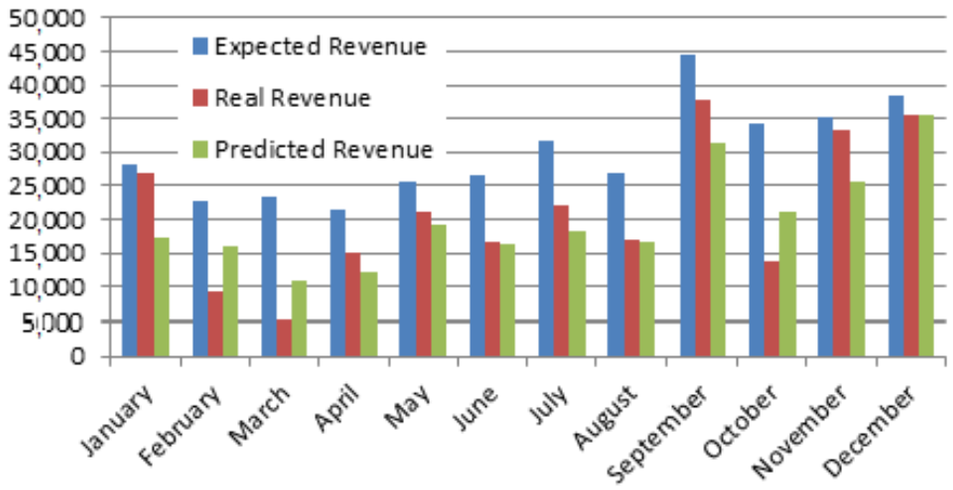


Figure 12. Bar charts comparing the expected, real and predicted revenue for one year

As can be seen, the fuzzy system has learned the expert's knowledge, therefore acting as a process expert and releasing them from the task of analysing, judging, and change the chosen values, then becoming able to do other activities.

7. Discussions and conclusion

The results produced by this initiative show how the default issue can be addressed by the use of Fuzzy Systems. The default in the economy is a serious problem, and although this problem cannot be solved easily, the facility to predict it can prevent bad clients to buy services for which it is not able to pay. Moreover, the Fuzzy System can be used to infer and forecast a more accurate cash flow, instead of traditional approaches.

One important advantage of this fuzzy system to forecast defaults is that it needed just a little piece of information to predict when a given customer would default a payment and under which probability. The simulation using quantitative techniques such as Monte Carlo method turned out a good estimation because of the stochastic nature of this process. Many models of the probability of default rely on statistical methods to infer the probabilities. This is an

interesting option when there is little data available on the customers to forecast default or bankruptcy by taking into account TTC probabilities.

The system has been applied in an Accounting System having aided financial analysts with predictions on cash flow and liquidity. One drawback of this system though is the lack of good predictions on new clients' transactions, but even in these cases the predictions are within the margin established by the fuzzy sets. However these results can be improved by performing risk and credit analysis or taking into account more information from the clients in the fuzzy system.

Author details

Fábio M. Soares*, Olavo S. Rocha Neto and Hevertton H. K. Barbosa

*Address all correspondence to: fabio@exodus.eti.br

Exodus Sistemas, Brazil

References

[1] Fonseca, R.A., Francisco, J.R. de S., Amaral, H.F., Bertucci, L.A., (2013) Impact of Defaults in Commercial Sector, Revista da Faculdade de Administração e Economia, v. 4, n.2, pp.39-60.

[2] de Severigny, A., Renault, O, (2004) The Standard & Poor's Guide to Measuring and Managing Credit Risk, McGraw-Hill, New York, US ISBN 0-07-141755-9.

[3] Duffie, D., Singleton, K.J., (2003), Credit Risk: Pricing, Measurement, and Management. Princeton University Press. Princeton, US, ISBN 0-691-09046-7.

[4] CNC (2013) Pesquisa Nacional de Intenção de Consumo das Famílias, http://www.cnc.org.br/central-do-conhecimento/pesquisas/pesquisa-nacional-de-intencao-de-consumo-das-familias-icf-janeir-0 , accessed in June 2014.

[5] Serasa Experian (2013), Economic Indicators and Default in Brazil, http://noticias.serasaexperian.com.br/indicadores-economicos/inadimplencia-das-empresas (accessed in June 2014).

[6] Katchova, A.L., Barry, P.J., (2005) Credit Risk Models and Agricultural Lending, American Journal of Agricultural Economics 87, no 1 2005, pp 194-205.

[7] Camargos, M.A., Camargos, M.A.S., Silva, F.W., Santos, F.S., Rodrigues, P.J., (2004) Determining factors of the default in processes of credit concession to micro and small businesses in the State of Minas Gerais (in Portuguese), Revista de Administração Contemporânea vol 14. no.2, ISSN 1982-7849, online version: http://www.scie-

lo.br/scielo.php?pid=S1415-65552010000200009&script=sci_arttext accessed in June 2014.

[8] Zirakja, M.H., Samizadeh, R., (2011) Risk Analysis in E-commerce via Fuzzy Logic, Int. Journal Management Business, Summer 2011, pp 99-112.

[9] Annibal, C.A. (2009), Default in the Brazilian Banking Sector: an assessment of its measures (In Portuguese), Central Bank of Brazil – Working Papers n. 192, ISSN 1519-1028

[10] Engelmann, B., Rauhmeier, R., (2006) The Basel II Risk Parameters, Springer Berlin Heidelberg New York, ISBN-10 3540-33085-2, June 2006.

[11] Simkovic M., Kaminetzky, B., (2010) Leveraged Buyout Bankruptcies, the Problem of Hindsight Bias, and the Credit Default Swap Solution (http://ssrn.com/abstract=1632084). Columbia Business Law Review, Vol. 2011, No.11, pp. 118, 2011.

[12] Herzog, A.L., (2001) Blow-by-blow (In Portuguese), Revista Exame, 7 mar. 2001.

[13] Kaplan, R.S., Norton, D.P., (1997) Why does business need a balanced scorecard?, Journal of Cost Management, May/June 1997, pp 5-10.

[14] Hsih, N.C., Chu, K.C. (2009) Enhancing Consumer Behavior Analysis by Data Mining Techniques. Int. J. of Information and Mangement Sciences, vol 10, N. 1, pp 39-53, March 2009.

[15] Preisler, A.M., (2003) Risk and Credit Analysis for Micro and Small Businesses – An guiding proposal (in Portuguese), Master Degree Dissertation in Production Engineering, Postgraduate program in Production Engineering, Federal University of Santa Catarina, Florianopolis, 180f, 2003.

[16] Jarrow, R.A., Turnbull, S., (1995) Pricing Derivatives on Financial Securities Subject to Credit Risk, Journal of Finance, vol 50, March 1995.

[17] Merton, R., (1976) Option Pricing when Underlying Stock Returns are Discontinuous, Journal of Financial Economics, 3, Jan-Mar 1976, pp. 125-144.

[18] Aguais, S.D., et. al, (2004) Point-in-Time versus Through-the-Cycle Ratings, in M.Ong (ed), The Basel Handbook: A Guide for Financial Practitioners (London:Risk Books).

[19] Altman, E. I., (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, Journal of Finance, 23: pp 589-609. doi: 10.1111/j.1540-6261.1968.tb00843.x.

[20] Basel II, (2003) Overview of the New Basel Capital Accord, Consultative Document of Bank for International Settlements, Issued for comment by 31 July 2003

[21] Jacobson, T., Roszbach, K., (2003) Bank Lending Policy, Credit Scoring and Value-at-Risk, Journal of Banking and Finance, 27 (2003), pp 615-633.

[22] Odeh, O.O., Featherstone, A.M., Sanjoy, D., (2006) Predicting Credit Default in an Agricultural Bank: Methods and Issues, Southern Agricultural Economics Association Annual Meeting, Orlando,FL, Feb 5-8, 2006.

[23] Kuan, C.M., Liu, T., (1995) Forecasting exchange rates using feedforward and recurrent neural networks, vol 10 issue 4 Oct-Dec 1995, pp 347-364.

[24] Martin, A., Gayathri, V., Saranya, G., Gayathri, P., Venkatesan, P., (2011) A hybrid model for bankruptcy prediction using genetic algorithm, fuzzy C-means and mars, International Journal on Soft Computing (IJSC), Vol 2 No 1, Feb. 2011. DOI: 10.5121/ijsc.2011.2102.

[25] Kastens, T.L., Featherstone, A.M., (1996) Feedforward Backpropagation Neural Networks in prediction of Farmer Risk Preferences, American Journal of Agricultural Economics 78 (May 1996), pp 400-415.

[26] Mamdani, E. H., Assilian, S. (1975) An experiment in linguistic synthesis with a fuzzy logic controller. Int. J. Man-Machine Studies, vol. 7, p. 1-13.

[27] Pereira, V.G.,De Oliveira, R.C.L.,Soares F.M. (2012) Fuzzy Control Applied to Aluminum Smelting. Fuzzy Logic Concepts Theories and Applications, ch. 13, Intech Open, Rijeka, Croatia, pp. 253-278.

[28] Zadeh, L.A., (1978) Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems, Vol. 1, pp 3-28.

[29] Shang, K., Hossen, Z. (2013), Applying Fuzzy Logic to Risk Assessment and Decision Making, Research Report sponsored by Casualty Actuarial Society, Canadian Institute of Actuaries. Available at: https://www.soa.org/Files/Research/Projects/research-2013-fuzzy-logic.pdf (accessed in June, 2014)

[30] White, D., (1995) Application of Systems Thinking to Risk Management: A Review of the Literature, Management Decision, vol. 3 (10), pp 35-45.

IntechOpen