CHAPTER NUMBER

RISK PREVENTION IN BRAZILIAN GOVERNMENT CONTRACTS USING CREDIT SCORING

LEONARDO SALES

Abstract

Credit Scoring models are statistical applications used by financial institutions to classify would-be customers as to the possibility of becoming defaulters. This research aims to take this acknowledged experience from the private sector to a governmental context, seeking to adapt it and test its performance in the identification of bidders that will most likely not fulfill their obligations in government contracts. The results of methodologies based on different statistical techniques are compared. We hope to contribute to the preventive control of contractual risks that both public administrators and government control agencies may face.
Keywords: Government Auditing; Data Analysis; Decision Tree; Logistic Regression; Credit Scoring.

1. Introduction

The insufficient or nonexistent rendering of services is one of the main problems in contracts between Brazilian public institutions and private companies. The losses it causes are always enhanced by the direct or indirect presence of public interest in these contracts.

These situations currently represent major challenges for both public administrators and government control agencies.

On the one hand, administrators need to enter into contracts with suppliers who can fulfill their obligations and provide a quality service, while ensuring the legality of the process, in compliance to a complex and inflexible legal framework.

On the other hand, the control agencies, that are frequently trying to adopt a more preventive attitude, have the challenge of anticipating risk situations that might impact the achievement of government objectives, among them the insufficient rendering of services contracted from the private sector.

We assume in this research that it would be relevant for the government audit units to understand in advance the probability that a given contract may present problems.

Therefore, we seek to test the applicability of predictive default models used by credit institutions (usually based on multivariate data analysis) in the prevention of contract default with public agencies, defined as the private company's failure to render services.

Among the models used by financial institutions, we selected the class called Credit Scoring as the object of this research. Credit Scoring models are based on the statistical consideration of the company's characteristics (registration or historical) to calculate its probability of becoming a defaulter. These models are widely utilized by such institutions and have been generating good results in the prevention of defaults (HAND and HENLEY, 1997).

The fact that these models are based on the company's characteristics is another assumption that makes us believe in this research that the application of these techniques on government contracts is feasible. This is because it is possible to extract plenty of information about the suppliers from the bases used by the federal government and that are available to the control agencies.

We therefore sought to identify which of these characteristics contribute to an increased likelihood of suppliers failing to fulfill their obligations with the government, thus allowing the risk of problems to be known in advance.

We tested two statistical classification techniques used in Credit Scoring models: Logistic Regression and Decision Tree, using a database relating to a sample of government suppliers as input and divided it into two groups: those who have a history of default (called "defaulters") and those with a history of fulfilling obligations (called "Compliant"). The performance of the techniques will be measured by comparing what was planned to actual ratings.

The entire research was done at the Office of the Comptroller General - CGU, which is the central agency of the internal control system of the Brazilian government. We used the databases and the hardware and software structure of the Public Expenditure Observatory - ODP[1], a unit that is a reference in data analysis and production of strategic information.

The results are discussed based on the evaluation of the accuracy of each technique used to predict the situations of default and the understanding of the variables that have greater predictive power.

It is hoped that the findings of this research may contribute to the CGU in building a robust methodology for contractual risk assessment, based on data analysis, that can be applied on a large scale and that can increase the public expenditure management and its control.

The paper is structured as follows: Section 2 presents the current literature on Credit Scoring, statistical techniques used in these models, and the Brazilian Government experience in managing public contracts; Section 3 describes the motivation and objectives of this research; Section 4 shows the methodology adopted; Section 5 presents the results of the constructed

[1] ODP won the United Nations Public Service Awards 2011, in the category Advancing Knowledge Management in Government.

models; Section 6 discusses the validation of the results; finally, Section 7 contains the final conclusions.

2. Literature Review

2.1. Credit Scoring

Researchers define Credit Scoring as quantitative models adopted by financial and credit institutions to measure and manage the risk of default by the borrower (LEWIS, 1992). The correct understanding of this risk allows the institution to measure the amount of credit, its rate or term, and prevents difficulties in contracts in progress (SAUNDERS, 2000).

In general, these models are based on the multivariate analysis of statistical techniques utilized to classify the customer in different groups according to the default risk (HAND and HENLEY, 1997). For this purpose, it uses information present in registration and financial databases or historical records of customer transactions with the institution (ARAÚJO and CARMONA, 2007).

The idea of using quantitative models to assess and manage credit risk is directly related to the expansion and massification of credit that occurred worldwide in recent decades, which showed the need to standardize the procedures for the selection of clients and portfolio management (MARQUES, 2002).

Vasconcellos (2002) declares that traditional methods of credit risk assessment are based on expert opinion and are therefore subjective and individualized. Thus, it does not apply satisfactorily to large credit markets. Marques (2001) adds that although these models have evolved by aggregating scores (called credit ratings) they are still inefficient in face of large volumes of transactions, because they are slow and not standardized.

Therefore, the Credit Scoring models added standardization, speed and accuracy to the processes of selection and management of clients of these institutions. The structuring of systems such as this one allowed extending the process of credit risk analysis, making it less dependent on subjective factors.

Another factor that favored the viability of these models was the development of computers, combined with the widespread use of computerized systems that made the gathering and structuring of large databases relating to transactions and customers possible. Likewise, increasing the processing capacity of computers also enabled the use of robust statistical techniques applicable to big volumes of data (HAND and HENLEY, 1997).

In recent studies that propose or test Credit Scoring models applied in different contexts, there is a common basis on which these systems are structured.

In general, these models are constructed based on three elements: a database with information about the customers (companies or individuals), another database with a structured record of negative experiences and one or more statistical techniques that are able to learn from this experience and predict risk situations.

The steps of its construction generally involve: 1) the initial selection of samples of good and bad cases, creating two groups generally classified as "compliant" and "defaulter"; 2) the definition of variables to be used by the model to differentiate groups; and 3) the implementation of statistical processing of learning and testing accuracy.

This last step is crucial for classification. First, the statistical tool "learns" from the base, assigning "values" to the variables and revealing those most capable of separating the groups. Then it creates a "formula" that defines the score of each company, and finally applies this formula to each case, allowing for a final classification.

The usefulness of the analysis will depend on the comparison of the generated ratings to the original classification. This will allow you to calculate the percentage of correct classifications (accuracy of the model).

The classic use of the Credit Scoring models is the selection of credit borrowers (CAOUETTE ET AL., 1998). Besides that, in the literature we can see examples of the use of these models in other stages of the credit life cycle. At times, it can be applied during the prospection of customers, when it is called Response Scoring; during the execution of the contract when it is called Behavioral Scoring; or even after a default has occurred when it is called Collection Scoring, to identify debtors that are more likely to pay the debt (SEMEDO, 2010).

As an example of traditional use of Credit Scoring, we can cite Vasconcellos (2002), who proposes a methodology for the analysis of credit loans to individuals, using transaction records and customer registration information from a popular Brazilian bank as the object of the study. In the same vein, Gevert (2011) investigates the performance of different statistical techniques for predicting defaults in bank contracts.

As for Araújo and Carmona (2007), who studied a microfinance institution, they also discuss the possibility of using Credit Scoring in the Behavioral Scoring model and compare the performances of two statistical techniques: Discriminant Analysis and Logistic Regression.

2.2. Statistical Techniques used in Machine Learning tools

As stated by Ghodselahi (2011), Discriminant Analysis and Logistic Regression are techniques historically used in the construction of score-cards. A favorable point for the use of these techniques is that they are easily found in the statistical packages that are available.

Recently, other classification techniques have been used, such as Decision Trees, Artificial Neural Networks and Support Vector Machine.

We will discuss in more detail the technical operation of Logistic Regression and Decision Trees, which are used in this research.

The Logistic Regression (LR) uses independent variables to classify elements into categorical variables and is based on a discriminant function constructed through the evaluation of the capability of each variable to distinguish the groups. Hair et al. (1998) cites the main aspects that favor the use of this technique in relation to Discriminant Analysis (another technique which, as seen, is widely used in Credit Scoring systems): the possibility of having the dependent variable be categorical and the non-dependence of this technique to the requirement of normality of the independent variables, which also need not be numerical.

The result of the logistic function (Logit) is binomial, receiving values 1 or 0, which indicate the presence or absence of a certain characteristic.

The function receives the following form (equation 1):

$$P(x) = \frac{1}{1 + e^{-(\propto + \sum \beta_i Y_i)}}$$

The elements that compose the logistic function are:

α = intercept (representing the minimum value of the function);
β = coefficient of each variable that determines its capability to move the function towards 1 or 0;
x = absolute value of a certain variable analyzed in the record;

According to Hair et al. (1998), the value P(x) is the probability of the function receiving the value 1 (indicating the presence of the studied characteristic).

The Decision Tree is a worldwide technique used to build classifiers. They consist of a simple graphical representation of the knowledge acquired through learning data analysis (SHIBBA ET AL, 2005).

Côrtes et al. (2002) compares the decision tree to a flowchart composed of the following elements: 1) nodes, which represent questions concerning attributes (predictor variables), 2) branches, which indicate answers to these questions, and 3) leaves (which are nodes without branches) that denote the distribution of records according to the response variable (dependent).

Figure 1, taken from Côrtes et al. (2002), exemplifies a Decision Tree being applied to classify customers into buyers or non-buyers, based on the following variables: salary, gender, marital status and owned property.

In this classification, the technique found the variables that best divide the two groups, calculating the most effective value (referring to the variable) to separate them. This classification generated a hierarchy of divisions that will be used to classify each new record that is analyzed. Imagine that a new potential client has a salary of 3,000 and did not own a property. In this case, following the series of divisions (branches of the tree) and passing through nodes 6 and 8, he would be classified as a non-buyer.

Logistic Regression permeates many academic studies that propose Credit Scoring models. Among them we highlight Araujo and Carmona (2007), who compare the LR with the Discriminant Analysis model in a proposal of risk management for a microfinance institution, and Semedo (2010), who investigates the usefulness of the technique in market companies in Cape Verde.

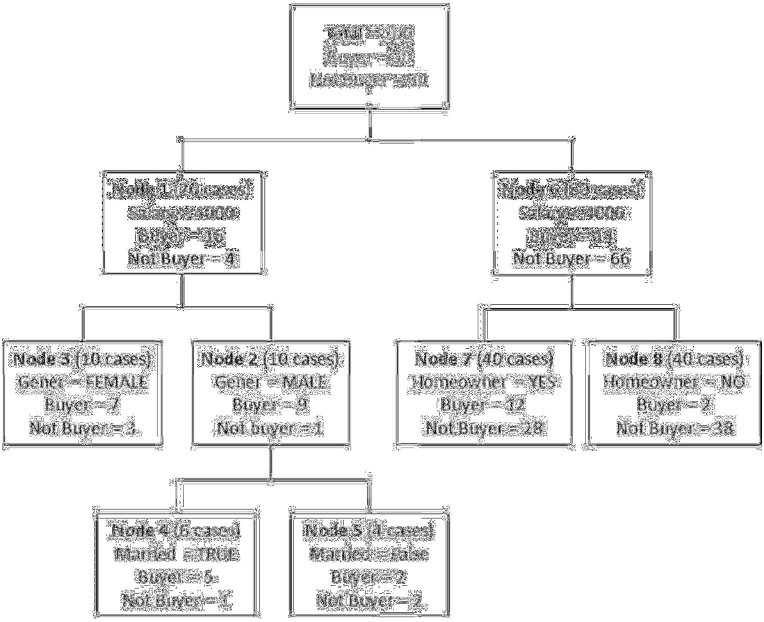

Figure 1: Decision Tree for Prospective Buyers (CÔRTES ET AL., 2002)

Vasconcellos (2002) uses decision trees to construct a model of credit risk management for individuals in a Brazilian bank.

More recent studies seek to use other techniques. Semedo (2010) compares the performance of Logistic Regression with the technique of Artificial Neural Networks. Ghodeslahi (2011) and Gevert et al (2011) test the Support Vector Machine technique, comparing its degree of accuracy with other techniques such as Neural Networks and Logistic Regression.

To the best of our knowledge this is the first research that applies Machine Learning techniques to identify companies likely to default public contracts. Nevertheless, there has been similar works in the government context. For instance, Carvalho *et. al.* (2013) use a probabilistic ontology to detect and prevent procurement fraud in the Brazilian government. The main difference is that the model created in Carvalho *et. al.* (2013) was designed by the expert instead of learned from data. Besides that, just a few real cases were used as a proof-of-concept.

2.3. Government Purchases

The entire process of buying products or hiring services in the federal government occurs according to the rules of Law n°. 8666/1993, entitled Bidding Law. Other norms complement this law such as law n °. 10.520/2005, which establishes a modality of reverse auction, and Complementary Law n °. 123, which provides privileges to micro and small companies in bidding processes. Law n°. 8.666/93 details the phases of the bidding process, the bidding modalities that are allowed, the types of contracts, aspects of the qualification of companies and also provides administrative and criminal sanctions to be applied to suppliers in case of failure to fulfill obligations.

We highlight the following parts of the Brazilian Bidding Law which defines the administrative sanctions:

> Section II
> Administrative Sanctions
> Art. 87. Due to the total or partial nonperformance of the contract, guaranteeing prior defense, the Administration may apply the following sanctions to the supplier:
> I - warning;
> II – fine, as provided in the bidding invitation or in the contract;
> III – temporary suspension in participating of bidding processes and impediment in celebrating contracts with the Administration, for a period no longer than 2 (two) years;
> IV – declaration of ineligibility to participate of bidding processes with the Public Administration while the determinant motives of the penalty perdure or until the reinstatement is executed by the authority that applied the penalty, which will be granted whenever the supplier reimburses the Administration for the resulting losses and after the period of the applied sanction, based on the previous section, has elapsed.
> (Brazil, 1993)

Article 87 states the penalties that the companies would face if they fail to fulfill their obligations with the government.

Sections III and IV (of the law) will be particularly relevant in this research. They define the companies that compose the National Registry of Ineligible and Suspended Companies - CEIS, considered in section 2.4. As foreseen in Law n °. 8.666/93, the process of acquiring and hiring in the federal government is carried out by the Integrated System for Management of General Services - SIASG. Each purchase or contract is registered in the system from the opening of the process to the commitment issuance. During the execution of contracts, public administrators can use the system to assign penalties to suppliers, if they fail to fulfill their obligations.

Established since 1994, SIASG recorded more than 400,000 contracts and nearly 4 million purchases.

Figure 2 shows the evolution of the amount of contracts published between 2000 and 2010.

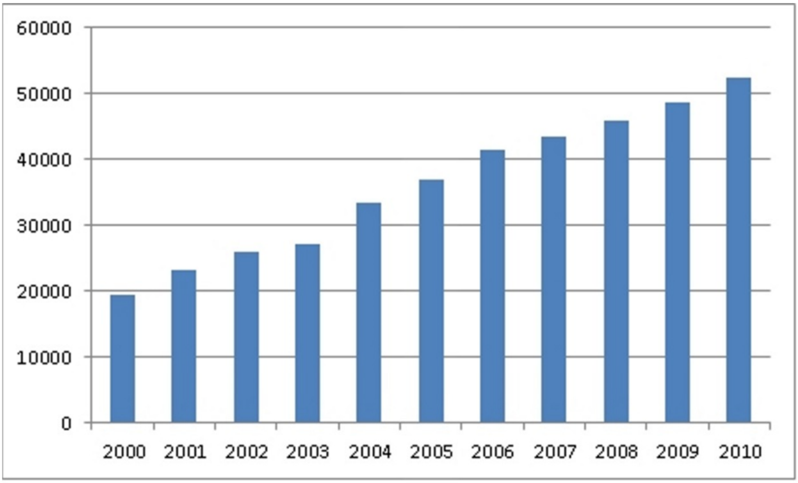

Figure 2: Amount of contracts registered in SIASG, 2000 to 2010
Source: Integrated System for Management of General Services - SIASG

2.4. Control of government acquisitions

As shown in Figure 2, only in 2010 more than 50,000 contracts were published, exhibiting a linear growth trend. This large volume of contracts is under the supervision of government control agencies, in particular the

Office of the Comptroller General - CGU, which is the central unit of the Internal Control System of the Federal Government[2].

The CGU states in its audit procedure manual that when auditing a governmental unit, each audit unit must select the contracts following criteria of materiality, severity, relevance and operational capability of the area to be audited. Then, the contracts have to be separated into three groups: exemption, unenforceability and other modalities.

The CGU also acts preventively. In 2009, it created a system to previously analyze calls for bids that looks for signs of favoritism. Usually, favoritism is related to the requirements that the supplier must fulfill in order to be hired.

Also in 2009, the CGU created the National Registry of Ineligible and Suspended Companies – CEIS[3], fulfilling one of the goals of the National Strategy Against Corruption and Money Laundering – ENCCLA, The purpose of the registration was to consolidate a list of companies that suffered sanctions by the agencies and the government entities of the Public Administration (Municipal, State and Federal).

These sanctions of the Declaration of Ineligibility and Suspension (to compete in the bidding process) are foreseen in the Law 8.666/93 and derive from the non-execution or partial execution of the contract and the existence of proven fraud at any stage of the contract or the bidding process itself.

The CEIS currently has more than 5,000 records.

3. Motivation and Objectives

We understand that the construction of statistical methodologies of data analysis, that focus on risk management and prevention on a large scale, will innovate how the CGU operates when auditing bidding processes and contracts.
There are gaps in the risk assessment process of government bidding processes by this control agency. This is because the supervision of

[2] Federal Law nº. 10.180/2001 (BRAZIL 2001), assigned to CGU the role of as central agency of the Internal Control System of the Federal Government.
[3] More information in http://www.portaldatransparencia.gov.br/ceis/SaibaMais.asp

purchasing processes always occurs in a precise manner (even when it is preventive, as in the previous analysis of calls for bids), in a number of processes that are limited to the operational capability of the audit units and the fragmentation of work by government structure. The subjective nature of assessments, as in the study of calls for bids, allows only a few of them to actually be analyzed.

This context of ad-hoc analysis which depends on expertise also exists in banking institutions and has motivated the credit risk management departments to develop Credit Scoring models. Nowadays, such models work supporting the activity of the professional, who develops subjective analysis only for contracts of great materiality, attributing the largest mass of analysis to statistical tools.

This research focuses on identifying, based on statistics and data analysis, enterprises that are likely to default public contracts. Thus, one of the expected contributions of this research is to enhance the risk management in these contracts.

The specific goals of this research are to: 1) Test the applicability and the performance of Credit Scoring models to predict the defaults caused by the supplier, and 2) Understand the characteristics of suppliers where they themselves contribute to being more prone to defaulting.

4. Methodology

In this research, we tested the performance of two statistical classification techniques, often used in Credit Scoring models: Logistic Regression and Decision Tree.

To be applied in the context of bidding processes, we use a database that represents two distinct groups: one of companies supposedly most likely to breach a contract and another with companies considered less prone to this type of situation. The database was also divided into groups of learning and testing.

We sought to identify and understand the variables that could best separate these two groups using classification techniques. Finally, we compared the results of each technique, calculating each accuracy percentage.

We divided the suppliers into two groups, which we call "Compliant" and "Defaulter".

Just for the fact that the company is registered at the CEIS, it is enough to be classified as a “defaulter”. The fact that the supplier was registered in the CEIS means that it has committed serious errors in the execution of any contract.

To be classified as "compliant" not only must the company not be registered at the CEIS, but it also has to have had a minimum of five contracts completed with the government over the past five years.

The database used contains 2,000 companies, where 1,000 are considered to be "compliant" and 1,000 are considered to be "defaulter". Each group contains 500 companies selected for learning and 500 for testing databases.

The 1,000 companies in the "defaulter" group were selected randomly from the database of the CEIS, after excluding the following cases: individuals, state-owned companies, businesses without contracts with the federal government and companies punished exclusively by municipalities or states. After these exclusions, our universe was reduced from 4,100 to 1,222 companies.

The construction of the group "Compliant" was much simpler. First, since we started from a very large universe of companies, we selected those who signed contracts this year (2011). We consider that this measure would reduce the universe without the model being bias. From the new universe of 4,011 companies, we selected 1,000 also randomly.

Table 1 represents the construction of the database, showing a step by step procedure for each category.

One of the techniques adopted to build the models was the Logistic Regression because this technique is one of the most used models in Credit Scoring and, moreover, it is less restrictive than the assumptions of the database used, not requiring the normality of distribution from the predictor variables nor the exclusive presence of numeric values.

Table 1 - Procedure for the construction of the base

	"Defaulters" Group	**"Compliant" Group**
Initial universe	All companies who were punished were registered at the CEIS (4,100 companies)	All suppliers with no record in CEIS, with at least five contracts completed in the last five years (21,743 companies)
Selection criteria	Excluded: individuals, state-owned companies, businesses without contracts with the federal government and companies punished exclusively by municipalities or states.	Selected: companies with contracts signed in 2011.
New universe	1,222 companies	4,011 companies
Sample	Random	Random
Main database	1,000 companies	1,000 companies
Learning database	500 companies	500 companies
Test database	500 companies	500 companies

This flexibility is also present in the Decision Tree. According to Meira et al. (2009), several variables in these models, numerical or categorical, could be elaborated simultaneously without damaging its performance and reliability. Moreover, this technique can graphically represent the functionality of the prediction model, which is an advantage over regression models.

To enable the construction of the models we used the software STATISTICA, from Statsoft, Inc.

The response variable used, called the dependent variable, is the group ("compliant" and "defaulter"), and will be obtained by the models through the analysis and consideration of the independent variables (predictors).

Table 2 describes the predictors variables used in the models.

Table 2 – Predictor variables

Variable	Description	Source
Days until 1st contract	Number of days elapsed between the opening of the company and the first contract.	*SIASG* and Federal Revenue database
Campaign donations in 2010	Amount donated to political campaigns in 2010.	Superior Electoral Court
Incidents monitored by the CGU	Number of times in which the company was selected in some monitoring trail by the CGU.	*CGU*[4]
No. of partners	Number of shareholders in the company.	Federal Revenue database
No. of employees	Number of employees.	Database of Ministry of Labor and Employment
Median age of partners	Median age of the shareholders in the company.	Federal Revenue database
Median age of employees	Median age of employees.	Federal Revenue database and of the Ministry of Labor and Employment
Median salary of partners	Median salary of the shareholders in the company.	Federal Revenue database and of the Ministry of Labor and Employment
Median salary of employees	Median salary of employees.	Database of Ministry of Labor and Employment

These variables were selected by the domain expert based on his experience and on the databases available. The rationale behind these variables is as follows:

---

[4] CGU has an intelligence unit, called Public Expenditure Observatory, which makes use of information technology to monitor government spending at a distance. This research generates alerts that represent suspicious transactions (purchases, for example). More comments on the ODP can be found at http://www.cgu.gov.br/ODP.

Days until 1st contract: According to auditing and inspections performed by CGU, the longer the company is active in the private market before the first contract with the government, the better their performance in public contracts.

Campaign donations in 2010: In Brazil, there are a lot of cases of corruption involving companies that have donated to politic parties. Even though, these companies tend to have a worse performance when executing public contracts.

Incidents monitored by the CGU: One of CGU's tasks is to monitor automatically public contracts and update the alerts database (e.g., bidders of the same procurement that have an owner in common). According to the experts, the higher the number of alerts for a specific company the more likely it is to presents problems.

No. of partners and No. of employees: Also according to the experts, the higher the number of partners and employees the better the quality of the company.

Median age of partners and employees: Although there was no previous knowledge on how the age of the partners would impact the performance of the company, the experts believed that it would have a negative correlation between the two variables.

Median salary of partners and employees: Again, according to the experts, better companies tend to pay better (higher) salaries to their partners and employees.

The Logistic Regression function was built in two stages. First, using the learning database, all selected variables were used and had their coefficients of discrimination (using the Maximum Likelihood Estimation method[5]) calculated to identify those with more discriminating power. Then, only the variables that were statistically significant (at a significance level of 5%) were used together in the construction of the function.

After having the best four coefficients recalculated (relative to the variables from the first round), we had enough parameters to build the

[5] According to Portugal (1995), the Maximum Likelihood method is a procedure, as well as the Least Squares, which allows for the estimation of statistical parameters and testing hypotheses

classification function. This function was applied to each of 1,000 records of the test database, giving us the result according to the classification model.

To validate the results, as per usual in such work, we built a classification matrix that compares each prediction to the actual value of the dependent variable, thus achieving an accuracy rate (percentage).

The significance level of accuracy was tested according to the Press' Q test[6].

Using the same variables and the same test database, the decision tree was built using the method of recursive partitioning, which is traditional for this technique.

The partitioning algorithm used was CART - Classification and Regression Trees[7], which divides the distribution of each variable in order to improve the separation of groups. This division is based on the Gini coefficient (HOFFMAN, 1980). Partitioning creates the conditions ("if, then") for each variable and distributes them in a tree hierarchically configured according to their discriminating power.

Once the tree was constructed, the software itself classified the test database records.

Again, this classification was validated by a comparison matrix with the actual values, when we obtained a new accuracy rate.

5. Results

5.1. Logistic Regression

Table 3 presents the estimated parameters obtained in the first step of the Logistic Regression.

---

[6] Details on the use of Press 'Q test in Hair Jr. et al. (1998). The equation is described in Section 6.1.

[7] The CART is a binary partitioning algorithm often used in the construction of Decision Trees. For more information, see Berry et al suggest. (1997).

Table 3 - Estimated parameters in the first stage of Logistic Regression

Variable	β	*Wald* test	P-value
Days until 1st contract	0,000015	0,31086	0,577153
Campaign donations in 2010	0,000004	2,87676	0,089867
Incidents monitored by the CGU	0,000110	1,08392	0,297822
No. of partners	0,018996	1,03650	0,308637
No. of employees	0,000056	0,35190	0,553037
Median age of partners	-0,038611	28,78896	0,000000
Median age of employees	-0,017014	6,05829	0,013841
Median salary of partners	-0,000179	8,34594	0,003866
Median salary of employees	-0,000654	11,71519	0,000620

As shown in Table 3, the last four variables were significant according to the Wald test (Hair Júnior. et al - 1998, states that there is significance for the variables with p-value less than 0.05).

As already mentioned, the coefficients β determine the influence of the predictor variable on the dependent variable ("Defaulter" or "Compliant" group). The four main predictors were the age of partners, the age of employees, the salaries of partners and the salaries of employees. The negative sign of the coefficient in all of them indicates that the variable inversely contributes to the likelihood of the value of the Logit function being equal to 1 (presence of the value "Defaulter"), which supports the rationale behind each variable. In other words, the smaller the values of variables, the closer the result of a function will be to 1. It is possible to confirm, for example, that the younger the age of the partners the higher the probability that the company will default contracts, using the same reasoning for the other three predictors.

Besides the "direction" of influence, we must also analyze its "intensity", characterized by the absolute value of the coefficient. The closer the value is to 1, the greater the intensity of the influence of the predictor variable. In this particular case, it is clear that the age (of partners and employees) has the most impact in predicting the class.

Nevertheless, the other variables did not have the expected outcome. Since their p-value was high, it was not possible to validate the expected correlation between these variables.
The next step, in order to optimize the adjustment of the model, was to rebuild it using only the significant variables (the median age and salary of members and employees). The second step resulted in new parameters for the selected variables, as described in Table 4:

Table 4 - Parameters estimated in the second stage of logistic regression

Variable	**β**	***Wald* test**	**P-value**
Median age of partners	-0,037118	32,86038	0,000000
Median age of employees	-0,014772	4,71279	0,029939
Median salary of partners	-0,000174	7,64116	0,005705
Median salary of employees	-0,000654	11,76702	0,005705

Thus, the Logit function was assembled using the coefficients of these four variables plus the intercept coefficient (equal to 2.688939).

5.2. Decision Tree

The tree generated from the learning base is presented in Figure 3.
As shown in the Figure 3, the decision tree represents a set of rules that determines the classification of each record. These rules will be applied to the test base to validate the model.

The first division of the tree is determined by the variable of number of employees. The algorithm determined the cutoff to be at the value of 8.5. Below this value the company is classified as "Compliant". Above it a new division occurs. Here, we can see that the algorithm has separated 305 companies ("Defaulters") and went in search of classifying the others.

The second division involves the median salaries of partners, where R$ 3,237.74 is the cutoff value. Above that, the company receives a rating of "Compliant" and below it new divisions occur.

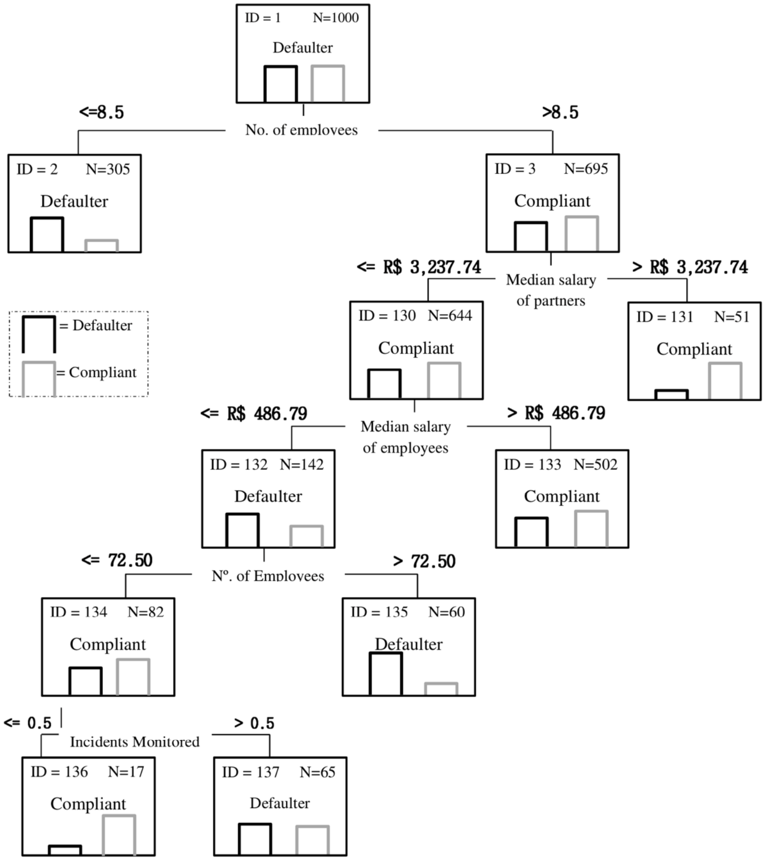

Figure 3: Results - Decision Tree

The tree is dismembered up until the last condition that divides the remaining 82 companies in "Compliant" to a number of monitoring instances less than 0.5, and "Defaulters" for instances above this value.

6. Validation of models

The scoring matrix of Table 5 shows the degree of accuracy of the Logit model:

Table 5 - Classification Matrix - Logistic Regression

Accuracy Matrix	**Classification of the model**	
Real classification	**Defaulter**	**Compliant**
Defaulter	68%	32%
Compliant	43%	57%
Total accuracy	**62,5%**	

Press' Q test can be used to measure the significance of the model. The formula is as follows (equation 2):

$$Q = \frac{[N - (nk)]^2}{N(k-1)}$$

In this equation, "N" is the number of records in the test database, "n" is the number of correct classifications and "k" is the number of groups.

We obtained the value of 62.5 with the use of the test, which is a result considered by Hair Jr et al. (1998) to be the minimum acceptable to differentiate the correctness of the model than what would be obtained by the criterion of chance (randomness).

This result indicates that the model adds value to the classification, having acceptable predictive capabilities, however, it needs improvement. A positive point is that, as can be seen in the classification matrix, the performance to predict "Defaulter" companies was 68% better than that obtained in the classification of "Compliant", which is more interesting for the predictive control of bidding processes.
Another way to evaluate the consistency of the Logit model is the dispersion curve of the residuals, reproduced in Figure 4.

The shape of the curve indicates the approach of residual dispersion to the normal curve, indicating that the model is suitable for the data.

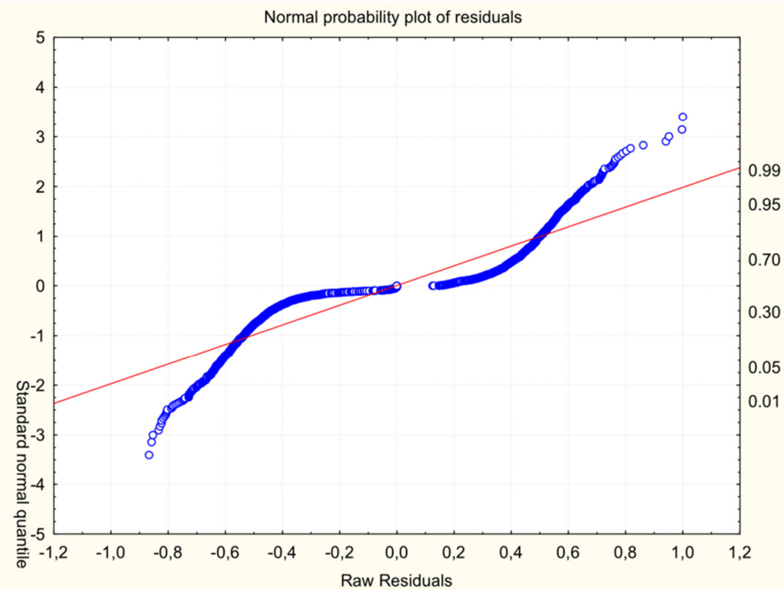

Figure 4 - Logit function. Dispersion Curve of Residuals

The results of the test base on the decision tree model are shown on Table 6.

Table 6 - Classification Matrix - Decision Tree

Accuracy Matrix	**Classification of the model**	
Real classification	**Defaulter**	**Compliant**
Defaulter	56%	44%
Compliant	30%	70%
Total accuracy	**63%**	

Press' Q test, once again showed an index of 67.6, slightly above that obtained with the Logistic Regression but still not an excellent result.

7. Conclusions

In order to achieve the overall goal of this research, to contribute to building a predictive model for risk management in government contracts with private companies, we tested the applicability of statistical techniques for the selection of clients used by financial institutions.

Both models used for classification arose from an identical set of variables, selecting those most capable of separating two groups of government suppliers: those most likely to fail to execute the contracts (called "Defaulters") and those least likely ("Compliant"). The criterion of maximum likelihood of Logistic Regression prioritized the characteristics of salary and age of partners and employees. Whereas the CART algorithm of the Decision Tree also selected the characteristics of salary of partners and employees, but did not include aspects linked to their ages in the final model. Instead, characteristics of the number of employees and incidents monitored trails were added.

The prevalence of issues linked to the "size" of companies such as major discriminating of good and bad suppliers in both models (quantity, age and salary of employees and partners) became evident. This indicates that the base used to select the group "Defaulter," the National Registry of Ineligible and Suspended Companies - CEIS, is more populated with smaller companies", which may indicate that these companies have made commitments to the government without having enough structure and experience to fulfill them.

Besides these characteristics, the variable of incidents recorded in audits made by the CGU (monitoring trails) appears in the tree model, which gives the audit a different perspective, more focused on results.

Moreover, understanding the ability of variables to determine the potential default of suppliers is, in our opinion, the most important and immediate contribution to the context of public audit, as well as the need for administrators of public contracts to foresee risky situations.

As a possibility in the medium term, our research demonstrated the feasibility of using multivariate statistical techniques, commonly used for the evaluation of credit with financial institutions, for the prediction of problems by government administrators with their suppliers. Based on the use of two statistical techniques and only nine variables, we achieved a more representative performance than mere randomness.

However, we consider the models constructed here to be a starting point and not a completed solution. The prediction results were only slightly above the acceptable and they indicate a great path to search for effective forecasts. We foresee that gains in accuracy can be obtained from testing more techniques, and, especially, increasing the amount of variables.

As for the use of classification models, such as Credit Scoring, we must emphasize that these cannot be used by the government in exactly the same way that they are used by financial institutions. This is because public administrators cannot base themselves on a predictive model to reject a supplier that wants to be hired by the government. The requirements to participate in bidding processes and employment are only defined by law.

We understand, however, that information coming from such prediction is interesting for the government for two reasons. First of all, for administrators, it is interesting to know their suppliers and the risks they offer to execute the services under their responsibility. This way, administrative effort can be focused on those contracts that are being conducted by potentially problematic suppliers.

Secondly, regarding audit activity, besides the obvious usefulness for the selection of contracts to be audited, there is the potential of expansion of the models focused on suppliers to a larger process of risk management in the provision of rendering of government services.

BIBLIOGRAPHY

Araújo, E.A. and C.U.M. Carmona 2007, "Desenvolvimento de Modelos Credit Scoring com Abordagem de Regressão Logística para a Gestão da Inadimplência de uma Instituição de Microcrédito", *Contabilidade Vista & Revista* 18 (3), 107-131

Berry, M.J.A. and G. Linoff 1997, *Data Mining Techniques - For Marketing, Sales and Custumer Support*, Indianapolis: Computer Publishing

Brazil 2001, "Lei nº 10180 de 06 de fevereiro de 2001. Organiza e disciplina os Sistemas de Planejamento e de Orçamento Federal, de Administração Financeira Federal, de Contabilidade Federal e de Controle Interno do Poder Executivo Federal, e dá outras providências", http://www.planalto.gov.br/ccivil_03/leis/LEIS_2001/L10180.htm

Brazil 1993, "Lei nº 8666 de 21 de junho de 1993. Referente à regulamentação do art. nº 37, inciso XXI, da Constituição Federal, instituindo normas para licitações e contratos da administração pública", http://www.planalto.gov.br/ccivil_03/leis/LEIS_2001/L10180.htm

Caouette, J., E. Altman and P. Narayanam 1997, *Gestão do Risco de Crédito: O Grande Desafio dos Mercados Financeiros Globais*, Rio de Janeiro: Qualitymark

Carvalho, R.N., S. Matsumoto , K.B. Laskey, P.C.G. Costa, M. Ladeira and L.L. Santos 2013, "Probabilistic ontology and knowledge fusion for procurement fraud detection in brazil", *Uncertainty Reasoning for the Semantic Web II, Lecture Notes in Computer Science* 7123, 19-40

Côrtes, S.C., R.M. Porcaro and S. Lifshitz 2002, *Mineração de Dados: Funcionalidades, Técnicas e Abordagens*, Rio de Janeiro: PUC-RioInf

Gevert, V.G., A.C.L. Silva, F. Gevert and V.T. Alves 2010, "Modelos de Regressão Logística, Redes Neurais e Support Vector Machine (SVM) na Análise de Crédito a pessoas jurídicas", *Revista de Ciências Exatas e Naturais* 12 (2), 269-293

Ghodselahi, A. 2011, "A Hybrid Support Vector Machine Ensemble Model for Credit Scoring", *International Journal of Computer Applications* 17 (5), 1-5

Hair, J.F., R.L. Tatham, R.E. Anderson and W. Black 1998, *Multivariate Data Analysis*, New Jersey: Prentice Hall

Hand, D.J. and W.E. Henley 1997, "Statistical Classification Methods in Consumer Credit Scoring: a review", *Journal of the Royal Statistical Society* 160 (3), 523-541

Hoffman, R. 1980, *Estatística para Economistas*, São Paulo: Pioneira

Lewis, E. 1992, *An Introduction to Credit scoring*, San Rafael: Fair Isaac

Marques, L.F.B. and G.O. Kloeckner 2002, "Gerenciamento do risco de crédito: um modelo de mercado vs. normas do Banco Central do Brasil", *Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração*, Salvador: ANPAD

Meira, C.A.A., L.H.A. Rodrigues and S.A. Moraes 2009, "Modelos de alerta para o controle da ferrugem-do-cafeeiro em lavouras com alta carga pendente", *Pesquisa Agropecuária Brasileira* 44 (3), 233-242

Portugal, M.S. 1995, *Notas Introdutórias Sobre o Princípio da Máxima Verossimilhança: Estimação e Testes de Hipóteses*, Porto Alegre: DECON/UFRGS

Saunders, A. 2000, *Medindo o Risco de Crédito – Novas Abordagens para o Value at Risk e Outros Paradigmas*, Rio de Janeiro: Qualitymark

Semedo, D.P.V. 2010, "Credit Scoring: Aplicação da regressão logística vs redes neuronais artificiais na avaliação do risco de crédito no mercado Cabo-verdiano", http://hdl.handle.net/10362/4041

Shibba, M.H., L.R. Santos, J.A. Quintanilha and H.Y. Kim 2005, "Classificação de imagens de sensoriamento remoto pela aprendizagem por árvore de decisão: uma avaliação de desempenho", *Anais do XII Simpósio Brasileiro de Sensoriamento Remoto*, Goiânia: INPE

Vasconcellos, M.S. 2002, "Proposta de Método para Análise de Concessões de Crédito a Pessoas Físicas", http://www.teses.usp.br/teses/disponiveis/12/12138/tde-18112004-125442/pt-br.php

INDEX