Reporte de Estabilidad Financiera

Marzo de 2012

Fragility Determinants Of The Private Corporate Sector In Colombia

Juan Sebastián Lemus
Adriana Corredor-Waldron
Javier Gutiérrez Rueda

Temas de Estabilidad Financiera
Mayo de 2012, no. 66

Fragility determinants of the private corporate sector in Colombia*

Juan Sebastián Lemus*
Adriana Corredor-Waldron†
Javier Gutiérrez Rueda‡

Abstract

The aim of this paper is to identify a set of early warning indicators that effectively discriminate between firms that are more prone to default on their financial obligations from those that are less prone to do so. To fulfill this objective, we use the Discriminant Analysis methodology. We find that the strongest predictors that a Colombian real sector firm will fail to meet their financial obligations are: debt ratio and the number of banking relationships. We also use a Logit model to estimate the debtors probability of default (PD) and its distribution. The PD distribution has a positive skew and leptokurtic, suggesting a low overall PD. When performing a stress test (i.e. when a negative shock is applied to the firms' performance), we find that the PD distribution shifts to the right causing an increase in loan loss provisions and a decrease in net profits.

JEL classification: *C25, D22, G21, G33*
***Keywords*:** Discriminant analysis, default, Logit, Colombian corporate sector, credit risk, stress testing.

---

*The authors would like to thank Juan Carlos Mendoza, Dairo Estrada and the staff of the Financial Stability Department at the Banco de la República for their useful comments and suggestions. The opinions contained herein are the sole responsibility of the authors and do not reflect those of Banco de la República or its Board of Governors. All errors and omissions remain our own.

*Analyst, Financial Stability Department. E-mail: jlemuses@banrep.gov.co
†Analyst, Financial Stability Department. E-mail: acorrewa@banrep.gov.co
‡Specialized Analyst, Financial Stability Department. E-mail: jgutieru@banrep.gov.co

1. Introduction

The measurement of credit risk is an important task for financial regulators and institutions, since it makes more efficient the process of credit allocation. In this way, assessing correctly if a debtor will enter to default in the short term has a positive effect in the financial soundness of the banking sector, and hence in the macroeconomic stability. In this paper, we seek to find the variables that best predict if a Colombian private firm will fail to meet its financial obligations.

The study is developed using the information provided by the financial statements and two variables that reflect a firm's relationship with the financial system: *i)* its history, characterized by the years that a firm has been a debtor, and *ii)* the number of financial institutions that are current lenders of the firm. Therefore, the model that is proposed in this paper, intends to determine a set of early warning indicators, registered in 2010, that can be useful for identifying which debtors are most likely to default and which are less likely to do so in 2011.

The analysis of the evolution of different financial indicators for the private corporate sector is essential, since this sector concentrates an important share of the Colombian banking credit. In particular, following the *Financial Stability Report of Banco de la República* (2012)[1], the Colombian private sector held 51% of the credit establishments' assets as of December of 2011. Therefore, it can be noted that evaluating the quality of the debtors that belong to the corporate sector is a main task for supervisors; due to the fact that a systemic failure of these agents may threaten the financial stability. In this way, it is important to identify factors that allow to predict if a firm could default in the short term. Given that financial ratios are used to measure the financial soundness of a firm, it is common to study the relationship between these indicators and a firm's default status. In this way, an analysis that incorporates different categories should be performed, when assessing a firm's strengths and weaknesses. Characteristics such as liquidity, leverage, profitability, solvency, operating performance, and others are usually studied (Lennox (1999)).

In this paper, we use Discriminant Analysis (DA) for assessing the determinants that a Colombian firm will enter shortly into default. This methodology has been widely used since its introduction by Altman (1968) and its posterior dissemination. We choose this multivariate statistical method for its simplicity (in terms of implementation and interpretatoon) and its wide availability in statistical packages. Despite the existence of more precise alternatives, it is important to take into consideration that a more complex methodology limits its replication by different agents interested in monitoring the financial soundness of debtors, such as supervisors and banks. Nevertheless, DA has been proven to be successful in the literature when studying the relationship between a set of factors and the classification of a firm as defaulting or non-defaulting.

Ortega et al. (2010) estimate a discriminant function using the five financial ratios[2] that Altman (1968) suggests. They found these indicators to be useful for predicting bankruptcy of the firms that belong to a Colombian chemical conglomerate. The results suggest that two independent variables - $Sales/Total\ Assets$ and $Earnings\ before\ interest\ and\ taxes/Total\ assets$ - have the highest contribution in the estimated model for correctly discriminating firms as defaulting or non-defaulting. On the other hand, Cruz et al. (2002) applied DA to Colombian listed companies. These authors, in correspondence to Altman (1968), found that three financial indicators (return on assets, debt ratio, and long-term

[1]Banco de la República is the Central Bank of Colombia

[2]By means of an *stepwise* selection procdure, Altman (1968) finds that the following set of covariates (financial ratios) are the most useful for predicting future bankruptcy of the real sector: $x_1 = Working\ capital/Total\ assets$, $x_2 = Retained\ earnings/Total\ assets$, $x_3 = Earnings\ before\ interest$ and $taxes/Total\ assets$, $x_4 = Market\ value\ equity/Book\ value\ of\ total\ debt$, and $x_5 = Sales/Total\ Assets$

leverage) are the most appropiate for discriminating. It is important to point out, that our results take as input a broader set of firms and that we incorporate a significant higher number of financial indicators and other microeconomic ratios into the different estimations that are carried out.

Previous research for Colombia have studied the direct relationship between a firm's fragility (measured by its probability of default) and a set of microeconomic and macroeconomic variables, by means of other methodologies different than DA. Generally, the microeconomic variables are financial ratios which are calculated using annual information provided by the *Superintendencia de Sociedades*[3]. González (2010) estimates an ordered logit model which calculates a firms' probability of default, its determinants, and the aggregate credit risk of the financial system. This author finds that the inclusion of macroeconomic factors such as GDP growth, CPI (Consumer Price Index), and unemployment rate improve the explanatory power of the model.

Gutiérrez Rueda (2010) carries out an estimation of a heteroskedastic probit model. Likewise González (2010), this author controls for macroeconomic factors in the model, showing that these variables' effects are heterogeneous across the distribution of the probability of default. In particular, strong variations of the GDP growth and the CPI have the strongest effect on the probability of default. This author tests the power of its estimated model by quantifying two errors: *i)* type I error and *ii)* type II error[4]. Despite that the author's result produce a high overall classification rate (*c.a.* 80%), the type I error for the selected model is high (43%), meaning that an important number of defaulting firms were classified as non-defaulting.

According to Brezigar-Masten & Masten (2009), it is relevant for researchers to establish which of the following objectives they seek when selecting any of the previous methodologies: either to guarantee the minimization of risk exposure or to achieve the profit maximization of the credit portfolio. From the supervisor's perspective, it could be more desirable to minimize type II error, since the misclassification of defaulting firms as non-defaulting could have a negative impact on financial stability. However, the implementation of this strategy implies a tradeoff. When attempting the minimization of the type II error, usually the type I error increases, in accordance with the results presented by Gutiérrez Rueda (2010). Therefore, in this paper we intend to minimize type II error, by means of a DA function that does not produce a significantly high type I error.

The present study finds that the strongest predictors that a Colombian real sector firm will fail to meet their financial obligations are: the debt ratio and the number of banking relationships. The former has a negative contribution in the discriminant score, meaning that firms with a low debt ratio are less likely to be classified as defaulting. The latter has also a negative contribution, implying that firms that are heavily connected (measured as the number of connections with the financial intermediaries) are prone to be classified as defaulting. However, for the Colombian case the firms with a significant number of banking relationships accumulate low shares of the commercial loan portfolio. Moreover, we estimate the probability of default using a Logit model, and its distribution by means of an Epanechnikov kernel. We find that this distribution is skewed to the left suggesting that there is a low overall probability of default for the loan portfolio. When we stress firm's financial indicators, the distribution shifts to the right causing an increase in loan loss provisions and a decrease in bank's net profit.

[3]Colombia's national institution that compiles the financial statements (Balance Sheets and Income Statements) of the real sector.

[4]According to this author, type I error occurs when the model misclassifies a defaulting credit as non- defaulting. On the other hand, type II error corresponds to the situation when a non-defaulting credit is misclassified as defaulting.

This paper is organized as follows: Section 2 describes in detail the formulation of the DA technique. Section 3 presents a characterization of the database used as an input for the estimations. In Section 4, the results of the estimations and a stress test. Finally, Section 5 presents some concluding remarks.

2. The Methodology

The aim of this paper is to identify a set of informative financial indicators (IFIs) with which we can effectively discriminate between firms that are more prone to default on their financial obligations from those that are less prone to do so. We use Discriminant Analysis (DA) for this purpose. Additionally, we use the set of IFIs that are identified using DA to estimate the probability that a firm will default in their financial obligations (henceforward PD) using a Logit model. In this section we present the methodologies used for these estimations.

2.1. Discriminant Analysis

We use Discriminant Analysis (DA) to perform this assessment given that it allows us to generete non-linear classifications of groups. For our analysis we define two populations: *i)* firms that have defaulted in their financial obligations in a one year period (π_d), and *ii)* companies that have not defaulted (π_{nd}). We define that a company has made default when any of their loan's rating has been downgraded from A or B to any other rating during to the above mentioned period of time[5].

Following Johnson & Wichern (1998), and Peña (2002), let $\mathbf{X}$ be a $(n \times p)$ matrix of p financial indicators associated to each of the n firms to be classified. Assuming that the financial indicators of firms in π_d differ in some degree from those in π_{nd}, we can state that there exists a set of values of $\mathbf{x}$ for which we can classify a firm to be in π_d and another set for which we can allocate a firm to be in π_{nd}. Thus, the population of firms belonging to π_d and to π_{nd} are described by the probability density functions $f_d(\mathbf{x})$ and $f_{nd}(\mathbf{x})$, respectively.

Now, consider a firm with a set of $\mathbf{x}$ values. This firm must be classified into either π_d or π_{nd}. Let p_d be the *prior* probability of π_d and p_{nd} be the *prior* probability of π_{nd}. Then, the conditional probability of correctly classifying the firm in π_d will be $\Pr(d|d)p_d$, and in π_{nd} will be $\Pr(nd|nd)p_{nd}$. However, there is a chance of misclassifying this firm. Let $\Pr(d|nd)p_{nd}$ be the conditional probability of classifying a firm as being in default when its not, and $c(d|nd)$ as the cost of misclassification. Similarly, $\Pr(nd|d)p_d$ as the conditional probability of assigning a firm as not to being in default when it is, and $c(nd|d)$ as the cost of wrongful classification. Since the cost of correctly classifying a firm is zero, we can define the expected cost of misclassification (ECM) as:

$$ECM = c(nd|d)\Pr(nd|d)p_d + c(d|nd)\Pr(d|nd)p_{nd} \tag{1}$$

The aim of any classification rule should be to have a ECM as small as possible. Accordingly, it is possible to define a classification rule as follows:

[5]In Colombia, loans have five loan rating classifications, labeled A through E. A firm rated A means that there is a high repayment probability and the firm is currently repaying its debt. The B rating indicates that there is a reduction in the probability of repayment, mainly due to risk factors or because the firm's financial obligations are overdue for less than three months. Credit rating C exhibits that a loan is overdue for more than 3 months and less than 6 months. The D shows that a loan is overdue for more than 6 months and less than year. Finally, rating E indicates default for more than a year and that there is no probability of recovering the debt.

$$R_d \quad : \quad \frac{f_d(\mathbf{x})}{f_{nd}(\mathbf{x})} \geq \frac{c(d|nd)}{c(nd|d)} \frac{p_{nd}}{p_d} \tag{2}$$

$$R_{nd} \quad : \quad \frac{f_d(\mathbf{x})}{f_{nd}(\mathbf{x})} < \frac{c(d|nd)}{c(nd|d)} \frac{p_{nd}}{p_d} \tag{3}$$

where R_d is a region where firms are classified as being in default and R_{nd} is a region where firms are classified as not being in default.

Further, we can assume that both populations are multivariate normal. Hence, $f_d(\mathbf{x})$ and $f_{nd}(\mathbf{x})$ are multivariante normal densities with mean μ_d and μ_{nd}, respectively, and covariance matrix $\mathbf{\Sigma}_d$ and $\mathbf{\Sigma}_{nd}$, in the same order.

We can first start by assuming that both populations have the same covariance matrices ($\mathbf{\Sigma}_d = \mathbf{\Sigma}_{nd} = \mathbf{\Sigma}$). In such case the joint densities of $\mathbf{X}$ for both populations are given by:

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_i)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu_i)\right] \quad \text{for } i = d, nd \tag{4}$$

Following the definition in (2), R_d and R_{nd} are given by:

$$R_d \quad : \quad (\mu_d - \mu_{nd})' \mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\mu_d - \mu_{nd})' \mathbf{\Sigma}^{-1}(\mu_d + \mu_{nd}) \geq \ln\left[\left(\frac{c(d|nd)}{c(nd|d)}\right)\left(\frac{p_{nd}}{p_d}\right)\right] \tag{5}$$

$$R_{nd} \quad : \quad (\mu_d - \mu_{nd})' \mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\mu_d - \mu_{nd})' \mathbf{\Sigma}^{-1}(\mu_d + \mu_{nd}) < \ln\left[\left(\frac{c(d|nd)}{c(nd|d)}\right)\left(\frac{p_{nd}}{p_d}\right)\right] \tag{6}$$

Given that the population values for μ_d, μ_{nd}, and $\mathbf{\Sigma}$ are unknown, we must replace them using the sample observations. Substituting $\overline{\mathbf{x}}_d$ for μ_d, $\overline{\mathbf{x}}_{nd}$ for μ_{nd}, and $\mathbf{S}$ for $\mathbf{\Sigma}$, the classification rule for any $\mathbf{x}_0$ is[6]:

Allocate $\mathbf{x}_0$ to π_d if:

$$(\overline{\mathbf{x}}_d - \overline{\mathbf{x}}_{nd})' \mathbf{S}^{-1}\mathbf{x}_0 - \frac{1}{2}(\overline{\mathbf{x}}_d - \overline{\mathbf{x}}_{nd})' \mathbf{S}^{-1}(\overline{\mathbf{x}}_d + \overline{\mathbf{x}}_{nd}) \geq \ln\left[\left(\frac{c(d|nd)}{c(nd|d)}\right)\left(\frac{p_{nd}}{p_d}\right)\right], \tag{7}$$

allocate $\mathbf{x}_0$ to π_{nd} otherwise.

Alternatively, it is possible to define a discriminant scalar variable as follows:

$$Z = \widehat{\mathbf{w}}'\mathbf{x} \tag{8}$$

where Z is the discriminant score. When $\left(\frac{c(d|nd)}{c(nd|d)}\right)\left(\frac{p_{nd}}{p_d}\right) = 1$, $\mathbf{w} = \mathbf{S}^{-1}(\overline{\mathbf{x}}_d - \overline{\mathbf{x}}_{nd})$.

Note that this classification scheme does not hold when $\mathbf{\Sigma}_d \neq \mathbf{\Sigma}_{nd}$. In such case, $\mathbf{\Sigma}$ in (4) must be replaced for $\mathbf{\Sigma}_i,\ i = d, nd$.

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_i)'\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\mu_i)\right] \quad \text{for } i = d, nd \tag{9}$$

[6] $\mathbf{S}$ is a weighted average of $\mathbf{S}_d$ and $\mathbf{S}_{nd}$. See Johnson & Wichern (1998, p. 641) for more details.

Replacing (9) in (2) and applying natural logarithm to both sides of the equation gives us the new classification regions:

$$R_d \quad : \quad -\frac{1}{2}\mathbf{x}'\left(\boldsymbol{\Sigma}_d^{-1} - \boldsymbol{\Sigma}_{nd}^{-1}\right)\mathbf{x} + \left(\mu_d'\boldsymbol{\Sigma}_d^{-1} + \mu_{nd}'\boldsymbol{\Sigma}_{nd}^{-1}\right)\mathbf{x} - k \geq \ln\left[\left(\frac{c(d|nd)}{c(nd|d)}\right)\left(\frac{p_{nd}}{p_d}\right)\right] \tag{10}$$

$$R_{nd} \quad : \quad -\frac{1}{2}\mathbf{x}'\left(\boldsymbol{\Sigma}_d^{-1} - \boldsymbol{\Sigma}_{nd}^{-1}\right)\mathbf{x} + \left(\mu_d'\boldsymbol{\Sigma}_d^{-1} + \mu_{nd}'\boldsymbol{\Sigma}_{nd}^{-1}\right)\mathbf{x} - k < \ln\left[\left(\frac{c(d|nd)}{c(nd|d)}\right)\left(\frac{p_{nd}}{p_d}\right)\right] \tag{11}$$

where

$$k = \frac{1}{2}\ln\left(\frac{|\boldsymbol{\Sigma}_d|}{|\boldsymbol{\Sigma}_{nd}|}\right) + \frac{1}{2}\left(\mu_d'\boldsymbol{\Sigma}_d^{-1}\mu_d - \mu_{nd}'\boldsymbol{\Sigma}_{nd}^{-1}\mu_{nd}\right) \tag{12}$$

Substituting μ_d, μ_{nd}, $\boldsymbol{\Sigma}_d$, and $\boldsymbol{\Sigma}_{nd}$ for their sample values, renders the following classification rule for $\mathbf{x}_0$:

Allocate $\mathbf{x}_0$ to π_d if:

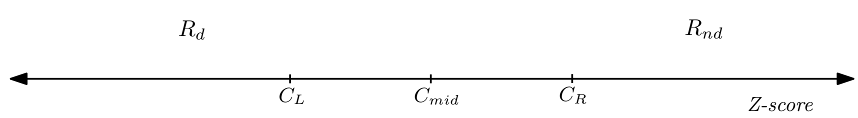
$$-\frac{1}{2}\mathbf{x}_0'\left(\mathbf{S}_d^{-1} - \mathbf{S}_{nd}^{-1}\right)\mathbf{x}_0 + \left(\overline{\mathbf{x}}_d\mathbf{S}_d^{-1} - \overline{\mathbf{x}}_{nd}\mathbf{S}_{nd}^{-1}\right)\mathbf{x}_0 - k \geq \ln\left[\left(\frac{c(d|nd)}{c(nd|d)}\right)\left(\frac{p_{nd}}{p_d}\right)\right], \tag{13}$$

allocate $\mathbf{x}_0$ to π_{nd} otherwise.

Equation (13) is commonly known as the quadratic classification rule[7].

Given the simplicity of calculating the *Z-Score*, we use this indicator as a mean of classification of firms. From the estimation of the DA model we obtain a set of two centroids. These centroids are expressed in terms of the *Z-Score* and are used as a cutoff for classification. Thus, any firm with a *Z-Score* that lays at the left of the of the first centroid (C_L) is classified as a defaulting firm and those at the right of the second centroid (C_R) as a non-defaulting firm (Figure 1). Since the centroids are not necessarily equal, there exists a gray region of classification located in between the centroids. We define a mid point in between the centroids (C_{mid}) and define the following classification rule for the gray region: every firm with *Z-Score* at the left of C_{mid} is allocated to the R_d region and those at the right to the R_{nd} region.

Figure 1: Classification Regions



Source: Own calculations.

[7] We use SPSS to estimate our model. Quadratic Discriminant Analysis (QDA) is not available in SPSS Statistics v. 20. Nonetheless, we use separate covariance matrices in the estimation, which renders identical results to QDA, "when the number of predictors does not exceed the number of groups minus 1" (International Business Machines Corp. (IBM) (2009)).

2.2. Probability of Default

In order to assess a credit risk mesure for the system, it is also desirable to obtain a probability that a firm will default in their financial obligations. To do so, we estimate a Logit model using the set of IFIs, that DA identify, as explanatory variables.

In this case, the dependent variable of the model is defined in terms of a latent variable as follows:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{in other case.} \end{cases} \tag{14}$$

where y_i is the dependent variable of the Logit model and y_i^* is the latent variable, defined as:

$$\begin{aligned} y_i^* &= f(IFIs) + \eta_i^* \quad \forall i = 1, \ldots, n \\ y_i^* &= x_i'\beta + \eta_i^* \qquad \eta_i^* \sim iid \;\; \Lambda\left(0, \sigma^2\right) \end{aligned} \tag{15}$$

where x_i $(K \times 1)$ a matrix of IFIs, β is a $(K \times 1)$ vector of unknown parameters, η_i^* is an *iid* random variable, and Λ is the Logistic cumulative distribution function.

The relation among the dependent variable of the model and the latent variable is defined as follows:

$$\begin{aligned} \Pr(y_i = 1) &= \Lambda(y_i^* > 0) \\ \Pr(y_i = 1) &= \Lambda(\eta_i^* > -x_i'\beta) \end{aligned} \tag{16}$$

Replacing $\Lambda(\cdot)$, in equation (16), for the logistic CDF[8], we get:

$$\Pr(y_i = 1) = \frac{1}{1 + e^{-x_i'\beta}} \tag{17}$$

Equation (17) gives us the probability that a firm will default in its financial obligations.

3. Data and Descriptive Statistics

To estimate the DA model we use the financial statements and the loan credit rating of each firm. The first data set is obtained from the *Superintendencia de Sociedades de Colombia* for the period 2009-2010. Each firm was described by a set of financial ratios that were calculated from the balance sheets and the income statements. The observations with missing data were eliminated, and due to the high sensitivity of DA to the inclusion of outliers, we set all ratios higher than the ninety-ninth percentile $(q_{0.99}(x_i))$ to that value. Similarly, all values lower than the $q_{0.01}(x_i)$ are truncated to that amount. Furthermore, we excluded from the database all financial companies due to their structural differences with those in the real sector.

[8] The logistic CDF is defined as $\Lambda = \frac{1}{1+e^{-x_i'\beta}}$.

The loan credit ratings are obtained from the Finance Superintendency of Colombia for period 2010-2011, and were used to construct the dependent variable. Only firms that were present in both databases are included in the study. In this way, the final sample has 13,216 observations, which correspond to 526 defaulting and 12,690 non-defaulting firms[9].

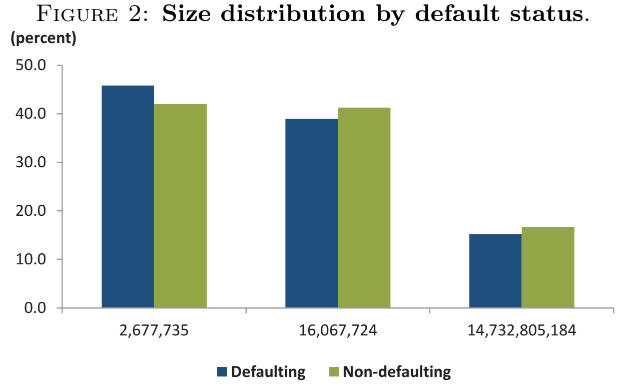
The availability of the financial statements place a restriction on the forecasting exercise. The *Superintendencia de Sociedades de Colombia* collects the data annually and it has a lag of one year. Therefore, in order to ensure that we could use our model and repplicate its results in the future, we decide to estimate the DA function with the information of one year prior to the period in which default was observed.

We calculate 37 financial ratios from the balance sheets and income statements, and classified them into four groups: activity, leverage, profitability and liquidity. The ratios have been chosen according to the international literature and previous studies for Colombia (JS). In order to select the variables to be included in the model, we perform a difference of means test. Table 1 provides the results to this test and the *t-test* for all the variables. As seen from the table, the test proves that 28 of the 37 variables are good candidates for the DA function. They belong mostly to the leverage and activity categories.

The difference of means test also shows that defaulting firms are, on average, unprofitable, highly leveraged and less liquid. Activity ratios reveal that these firms have higher inventory turnover ratio implying poorer sales, and, therefore, excess inventory. Moreover, they have higher accounts receivable turnover ratio, suggesting that they apply laxest policies in extending credit to their customers.

In addition, a firm can be classified according to its assets value[10] In Figure 2, it is observed that *c.a* 80% of the firms (without making the distinction between defaulting and non-defaulting ones) are classified as either 'small-sized' or 'medium-sized'. In addition, it is relevant to point out that on each of the three buckets of size, the proportion of defaulting firms is approximately equivalent to that observed for non-defaulting firms. Therefore, it can be concluded that size composition does not differ between the two type of firms.

FIGURE 2: **Size distribution by default status**.



Values in thousands of colombian pesos (COP).
Source: Own calculations, based on data from *Superintendencia de Sociedades* and Finance Superintendency of Colombia.

Another issue that it has been widely studied is the influence of industry; although, there is no consensus regarding the relationship between this variable and the firm's default status. Figure 3 describes

[9]These firms accumulate *c.a.* 45.3% of the credit portfolio held by the private corporate sector as of December 2010.

[10]If a firm's assets value is less or equal to 2,677,735 thousands of COP, then it is classified as 'small-sized'; if its assets values is greater than 2,677,735 thousands of COP and less or equal to 16,067,724 thousands of COP, then it is classified as 'medium-sized'; and if its assets value is greater that 16,067,724 thousands of COP, then it is classified as 'large-sized'. The highest assets value figure, for the sample of firms used in this study, corresponds to 14,732,805,184 thousands of COP.

TABLE 1: Means and Difference of Means test

No.	Name	Unit	Mean Non-defaulting	Mean Defaulting	t-value
			Activity		
X1	$\frac{Sales_t}{Sales_{t-1}}$	%	1.82	0.88	3.08*
X2	$\frac{Sales_t}{\text{Total assets}_t}$	%	1.83	1.20	7.90*
X3	$\frac{\text{Selling and administrative expenses}_t}{\text{Sales}_t}$	%	0.32	0.306	1.17
X4	$\frac{\text{Non-operating revenue}_t}{\text{Sales}_t}$	%	0.076	0.056	1.84
X5	$\frac{\text{Non-operating expense}_t}{\text{Sales}_t}$	%	0.07	0.115	-3.62*
X6	$\frac{\text{Gross income}_t}{\text{Sales}_t}$	%	0.35	0.27	7.67*
X7	$\frac{\text{Fix assets}_t}{\text{Sales}_t} * 12$	month	4.11	3.62	1.64
X8	$\frac{\text{Current assets}_t}{\text{Sales}_t} * 12$	month	8.41	10.43	-3.33*
X9	$\frac{\text{Accounts receivable}_t}{\text{Sales}_t} * 12$	month	4.39	6.36	-3.75*
X10	$\frac{\text{Inventory}_t}{\text{Sales}_t} * 12$	month	2.55	3.21	-2.86*
X11	$\frac{\text{Accounts payable}_t}{\text{Sales}_t} * 12$	month	1.26	2.07	-5.69*
X12	$\frac{\text{Accounts receivable}_t}{\text{Accounts payable + Suppliers}_t} * 12$	month	8.88	4.29	2.92*
			Leverage		
X13	$\frac{(\text{Liabilities + Shareholders equity})_t}{(\text{Liabilities + Shareholders equity})_{t-1}}$	%	1.26	1.01	3.1*
X14	$\frac{\text{Equity}_t}{(\text{Liabilities + Shareholders equity})_t}$	%	4.14	1.05	5.1*
X15	$\frac{\text{Total Liabilities}_t}{\text{Total assets}_t}$	%	0.54	0.72	-8.5*
X16	$\frac{\text{Current liabilities}_t}{\text{Equity}_t}$	%	0.43	0.52	-4.38*
X17	$\frac{\text{Financial liabilities}_t}{\text{Sales}_t}$	%	0.16	0.349	-13.47*
X18	$\frac{\text{Shareholders equity}_t}{\text{Shareholders equity}_{t-1}}$	%	2.16	1.18	4.49*
X19	$\frac{\text{Current liabilities}_t}{\text{Total liabilities}_t}$	%	0.804	0.724	6.34*
X20	$\frac{\text{Total liabilities}_t}{\text{Equity}_t}$	%	0.54	0.72	-8.504*
X21	$\frac{\text{Equity}_t}{\text{Fix assets}_t}$	%	60.84	40.12	1.46
X22	$\frac{\text{Gross Income}_t}{(\text{Liabilities + Shareholders equity})_t}$	%	0.88	0.38	22.0*

No.	Name	Unit	Mean		t-value
			Non-default	Default	
		Profitability			
X23	$\frac{\text{Net Income}_t}{(\text{Equity})_t}$	%	0.10	-0.46	0.87
X24	$\frac{\text{Net income}_t}{(\text{Total assets})_t}$	%	0.03	-0.11	4.78*
X25	$\frac{\text{Earnings before interest and taxes (EBIT)}_t}{(\text{Total assets})_t}$	%	0.07	-0.03	6.8*
X26	$\frac{\text{Retained earnings}_t}{\text{Total assets}_t}$	%	0.054	-0.017	4.32*
X27	$\frac{\text{Gross income}_t}{\text{Equity}_t}$	%	1.75	2.71	-0.45
X28	$\frac{\text{EBIT}_t}{\text{Sales}_t}$	%	0.03	-0.03	6.11*
X29	$\frac{\text{Net income}_t}{\text{Sales}_t}$	%	0.02	-0.10	8.5*
X30	$\frac{\text{EBIT}_t}{\text{Equity}_t}$	%	0.20	-0.46	1.5
X31	$\frac{\text{Earnings before taxes}_t}{\text{Equity}_t}$	%	0.166	-0.43	0.94
X32	$\frac{\text{Sales}_t}{\text{Fix assets}_t}$	%	105.94	46.05	2.35*
		Liquidity			
X33	$\frac{\text{Working capital}_t}{\text{Total assets}_t}$	%	0.21	0.09	6.06*
X34	$\frac{\text{Current assets}_t}{\text{Current liabilities}_t}$	%	2.30	1.87	3.0*
X35	$\frac{\text{Current assets-Inventory}_t}{\text{Current liabilities}_t}$	%	1.66	1.33	2.65*
X36	$\frac{\text{Current assets}_t}{\text{Total assets}_t}$	%	0.65	0.62	3.17*
X37	$\frac{\text{Cash}_t}{\text{Sales}_t} * 365$	day	5.46	4.77	0.63
X38	Working capital	$ US	1,216,496,000	660,120,000	2.96*
		Size			
	Log(Assets)	$ COP	15.2	15.0	2.64*

Source: Own calculations.

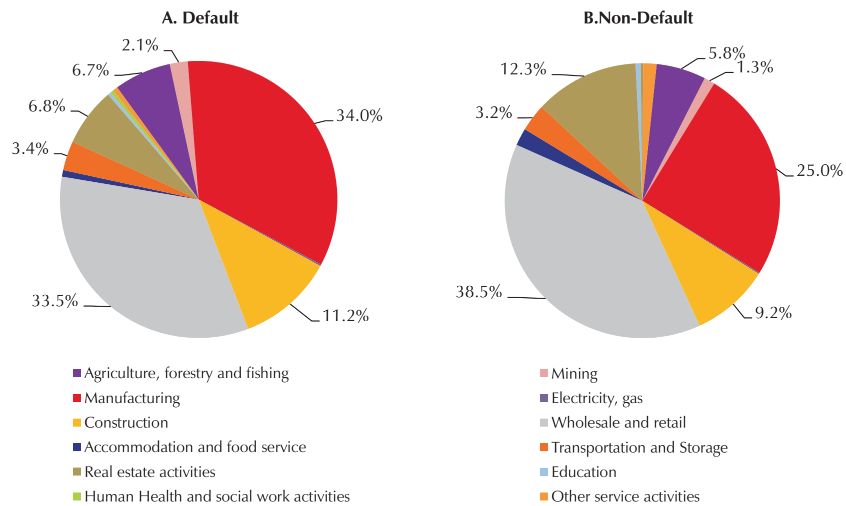
the industry composition of the database. As it is shown, both groups consists mostly of firms from Manufacturing and Wholesale and Retail trade. Other sectors like Construction, Real estate activities, and Agriculture participates between 5.9%- 12.1%. Minor differences are noted between groups composition, which may suggest that industry it is not a strong predictor of default.

When analyzing the composition of the credit portfolio stock[11] by industry and type of firm (defaulting or non-defaulting), we find that the non-defaulting firms hold 97% of this portfolio, while the defaulting ones the remaining 3%. Manufacturing[12] holds the highest participation for the two types of firms (*c.a. 50%*, as it can be observed in Figure 4. In addition, it is important to point out that this result does

[11]Corresponds to the total credit issued to the private corporate sector analyzed in this study as of December of 2010 (13,216 firms)

[12]In Figure 4 the sectors are labeled as: Manufacturing (D), Wholesale and retail trade (G), Construction (F), Agriculture, forestry, and fishing (A & B), Real estate activities (K), Mining (C), Other service activities (O), Transportation and storage (I), Accommodation and food services (H), Electricity and gas (E), Education (M), Human health and social work activities (N).

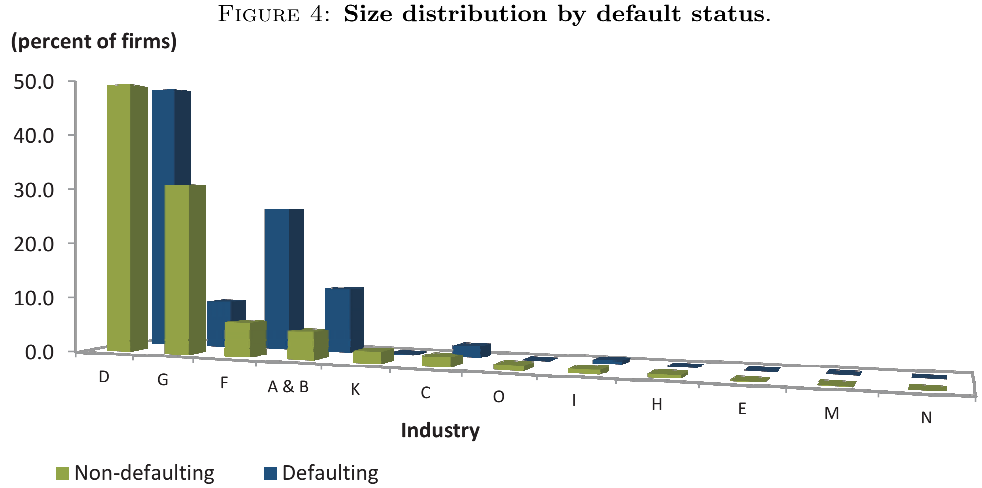
Figure 3: **Industry composition by default status**.



Source: Own calculations, based on data from *Superintendencia de Sociedades* and Finance Superintendency of Colombia.

not contradict the sectorial composition detailed on Figure 3. However, it is important to mention that while Construction is the second industry with the highest participation for the defaulting firms, for the non-defaulting ones is Wholesale and retail trade (30.9%). Finally, as it was expected, the firms that belong to the sectors with a low participation in the sectorial composition (such as Mining, Electricity and gas, Real estate activities, among others), hold a negligible participation in the credit portfolio. In this way, it can be stated that the financial system is not significantly exposed to these latter firms

Figure 4: **Size distribution by default status**.



Source: Own calculations, based on data from *Superintendencia de Sociedades* and Finance Superintendency of Colombia.

An alternative exercise for assessing if Industry effectively discriminates between defaulting firms and non-defaulting ones, is to compare the financial soundness of both types of firms when the sample is disaggregated by Industry. The indicators used in this exercise correspond to those that are widely used in the literature when applying discriminant analysis. These indicators are sales to assets ratio (X2), debt ratio (X15), return on total assets (X25), current ratio (X34), the natural logarithm of the assets, and a debtor's number of banking relationships. The results show that non-defaulting firms have a higher degree of financial soundness in comparison to that registered for defaulting firms (as it was expected). The latter firms are less liquid, less profitable, and more indebted than the former ones. It is important to point out that for a few number of sectors and for some indicators, non-defaulting firms have a better performance[13]. However, it should be noted that for every sector that is analyzed, the number of non-defaulting firms exceeds in an important way the total of defaulting firms. Therefore, one can explain these results by suggesting that the indicators' volatility of the first group is significantly less in comparison to that observed for the second one. In addition, it is worthy to mention that defaulting firms have a higher number of banking relationships in comparison to non-defaulting firms. Defaulting firms that belong to the sectors of Agriculture, forestry and fishing, Manufacturing, Construction, and Wholesale and retail trade are the agents with the highest number of connections with financial intermediaries.

The previous results are inconclusive in stating whether Industry is a discriminating variable. In this way, we perform a statistical exercise in Appendix A, with the aim of contrasting if the inclusion of Industry can improve the classification power of DA. The results show that this variable does not contributes significantly to the overall classification rate of the discriminant function.

Table 2: Benchmark Indicators' arithmetic means by Industry

Sector	**Group**	**Number of firms**	**X2**	**X15**	**X25**	**X34**	**Log(Assets)**	**Number of banking relationships**
A & B	Non-defaulting	740	0.91	0.40	0.01	2.08	15.62	2.56
A & B	Defaulting	35	0.93	0.64	-0.06	1.25	15.95	4.31
C	Non-defaulting	159	1.33	0.51	0.08	1.64	16.48	2.61
C	Defaulting	11	0.40	1.14	-0.15	0.78	16.05	3.00
D	Non-defaulting	3176	1.29	0.48	0.06	1.87	16.02	3.93
D	Defaulting	179	0.97	0.68	0.03	1.32	15.83	5.38
E	Non-defaulting	19	1.09	0.49	0.09	1.56	16.21	3.67
F	Non-defaulting	1170	1.25	0.63	0.04	2.44	15.80	2.80
F	Defaulting	59	0.78	0.64	0.03	1.65	16.00	5.43
G	Non-defaulting	4888	2.12	0.58	0.05	1.99	15.37	3.50
G	Defaulting	176	1.50	0.77	0.00	1.84	14.89	4.54
H	Non-defaulting	258	1.17	0.38	0.06	1.49	15.42	2.75
H	Defaulting	4	0.18	0.16	-0.01	0.38	17.17	3.00
I	Non-defaulting	411	1.31	0.58	0.08	1.66	15.13	2.85
I	Defaulting	18	0.84	0.52	0.06	1.26	14.07	3.80
K	Non-defaulting	1558	1.47	0.47	0.08	2.39	15.31	2.54
K	Defaulting	36	1.89	0.49	-0.16	1.54	14.72	3.00
M	Non-defaulting	75	1.30	0.56	0.04	0.80	14.74	2.89
M	Defaulting	2	1.08	0.82	0.02	0.62	14.66	4.50
N	Non-defaulting	37	1.45	0.52	0.10	2.57	15.01	3.00
N	Defaulting	2	1.59	0.53	0.07	1.93	15.12	1.50
O	Non-defaulting	199	1.10	0.50	0.04	2.05	15.28	2.98
O	Defaulting	3	1.33	0.34	0.20	1.83	14.94	3.00
	Total	13215						

Source: Own calculations.

4. Results

In this section we present the results of the DA model. We analyze the classification accuracy of the model and identify the set of IFIs that are useful to identify defaulting firms. Additionally, we present

[13]For instance, defaulting firms that belong to the real estate activities Industry have a higher return on assets indicator (1.89) with respect to the one registered for its counterpart firms of the same sector (1.47)

the estimation of a Logit model and calculate the distribution of the PD. Finally, we perform a stress test and assess the impact that a shock on the IFIs has on the financial system.

4.1. DA Analysis

As noted in the previous section, in this study a database of 13,216 firms is taken as input for the estimation of the models. Since the sample is strongly unbalanced, *c.a.* 96 % of the firms are non-defaulting; we decide to design two different samples: *i)* A full sample that takes the data as it is and *ii)* a choice-based sample. On the second sample, we guarantee that any defaulting firm is paired up with a non-defaulting firm. The criteria for matching two firms takes into account size[14] and industry. For example, a defaulting firm that belongs to the manufacturing industry and is large, is paired up with a non-defaulting firm that has the same industry and size characteristics.

On the other hand, two variable selection approaches are compared, with the aim of testing their classification power: *stepwise* selection[15] and *benchmark* selection. The second approach simply refers to estimate a DA function incorporating the covariates that González (2010) and Gutiérrez Rueda (2010) found as the most appropriate predictors of the Probability of Default (PD) for the Colombian private corporate sector. In the previous section, these indicators are labeled as 'benchmark variables'.

Table 3 presents the estimation results for the two samples (matched[16] and full). In both approaches, we first estimate a DA function taking as input the learning sample. Then, we test for the accuracy of the estimated function by classifying out-of-sample firms (validation sample). The validation sample maintains the proportions of the population, *i.e.* 96% represent non-defaulting firms while the remaining 4% correspond to defaulting[17].

The results suggest that the full sample achieves the best overall classification rates, on average it is approximately 95.6%. However, the results for the learning sample show that a low percentage of defaulting firms are correctly classified (19% for the *stepwise* variable selection and 8.1% for the *benchmark* selection). As it was mentioned, from the supervisor's perspective it is important to formulate a credit risk model that guarantees a high correct classification rate for defaulting firms. It can be noted that the matched samples achieve this goal, regardless of the variables selection approach, the accuracy classification rate for defaulting firms is 74.6% in the learning sample and 81.9% in the validation sample,

[14] In this study, a firm's size is classified according to its assets value. Therefore, a firm can be classified as: *i)* Small if its assets value ranges between 501 and 5.001 minimum wages, *ii)* Medium if its assets value ranges between 5.001 and 30.000 minimum wages, and *iii)* Large if its assets value is greater than 30.000 minimum wages.

[15] This method incorporates step by step a predictor in the discriminant function with the aim of maximizing the Mahalanobis Distance between the two groups.

[16] With the aim of obtaining robust results, ten random matched samples are build. A DA model is estimated for each sample using the *stepwise* selection approach, in order to evaluate if the estimations within samples differ significantly. Eleven variables appear consistently in at least seven of the ten estimated models. In addition, the coefficients' sign and magnitude of these covariates are consistent throughout the estimations. Finally, we select the model that fulfills two conditions: *i)* it achieves the highest classification rates in the learning sample; and *ii)* when extrapolating its covariates coefficients into the population, a high percentage of firms are correctly classified. The model that is selected is 'Matched-I' as labeled in Table 3. This model has fourteen covariates, which eleven correspond to those that consistently appear in seven or more estimations as previously noted. In fact, when we evaluate the second condition on the population, we obtain a defaulting firms' correct classification rate of 76% and an overall correct classification rate of 75.3%. It can be noted that the quality of the results is not affected substantially. It is relevant to point out, that on average, throughout the ten samples, the defaulting firms correct classification rate is 76.9% and the overall correct classification rate corresponds to 75.7%.

[17] Given that the groups' proportions in the match sample do not correspond to those observed in the full sample, we design a validation sample that reflects the population composition. In this way, we check the results' robustness and accuracy, when classifying out-of-sample-firms that belong to an unbalanced sample.

Table 3: Prediction accuracy (percentage)

	Non-Defaulting	Defaulting	Overall*
		Stepwise	
Matched-I			
Learning sample	82.4	74.6	78.5
Validation sample	74.9	81.9	75.2
Full-I			
Learning sample	98.6	19	95.4
Validation sample	98.4	19	95.3
		Benchmark	
Matched-II			
Learning sample	76.7	69.4	73.0
Validation sample	75.5	70.5	75.3
Full-II			
Learning sample	99.7	8.1	96
Validation sample	99.3	5.7	95.5

*Overall refers to the total rate of correct classification, regardless of default status group.
Source: Own calculations.

when the *stepwise* approach is used. In addition, it is noteworthy that this approach provides better results relative to those provided by the benchmark selection.

In this way, we choose the 'Matched-I' as the best-fit model since it provides a high accuracy classification rate of defaulting firms in the learning sample (81.9%), and simultaneously an acceptable rate for the non-defaulting firms (74.6%). Table 4 shows the estimation results of the discriminant analysis of the selected model. The magnitude of the standardized coefficients indicate the relative contribution of each predictor to the function. In this sample, the debt ratio is the strongest predictor, followed by the number of banking relationships. On the other hand, the financial indebtedness ratio and the retained earnings to total assets ratio contributes the least. Unstandardized coefficients are used for calculating the *z-scores*, *i.e.* the score that will be used as an input for classifying a firm as defaulting or non-defaulting.

In addition, the sign of both coefficients (standardized and unstandardized) points out the direction of the relationship. From Table 4, it stands out that the financial ratios that measure profitability have positive coefficients. Therefore, since the centroid for the non-defaulting group is positive, the signs of the coefficients for these ratios suggest that a firm with higher profits will be more likely to be classified as non-defaulting. Likewise, since the centroid for the defaulting firms is negative and the coefficients of the leverage ratios are negative, it should be expected that a firm with higher leverage ratios will be classified as defaulting. The estimated model also includes one liquidity ratio, with positive coefficient.

These results are expected and consistent with those presented in the literature about PD prediction models. In this paper, the final estimation include new variables: *i)* credit history, *ii)* number of banking relationships, and *iii)* accounts receivable turnover ratio. From the results obtained for these variables, it can be seen firstly, that a firm that has a long history with the banking system is more likely to be classified as non-defaulting. This result could be associated with the fact that a firm that has a persistent relationship with one or more banks, is one that meets regularly its financial obligations. While a firm that has been introduced to the financial system recently could be seen as lack of creditworthiness.

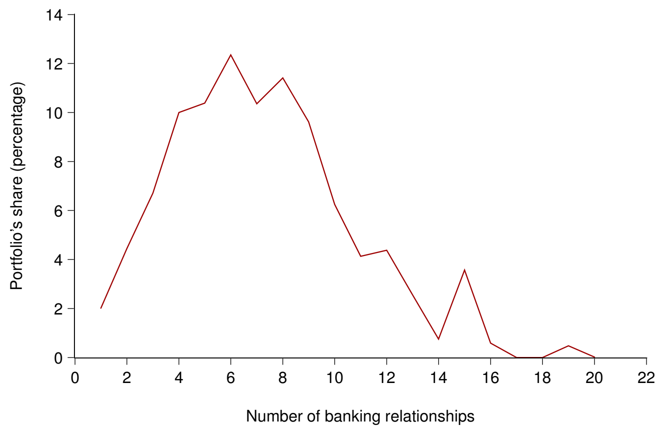
Table 4: Standardized and unstandardized discriminant function coefficients and centroids (Match sample)

Variable	**Standardized Coefficient**	**Unstandardized Coefficient**	
Credit history	0.26	0.08	
Number of banking relationships	-0.41	-0.17	
$\frac{\text{Retained earnings}_t}{\text{Total assets}_t}$	0.15	0.73	
$\frac{\text{Net income}_t}{\text{Total assets}_t}$	0.16	1.48	
$\frac{\text{Accounts receivable}_t}{\text{Sales}_t} * 12$	-0.22	-0.04	
$\frac{\text{Accounts receivable}_t}{\text{Accounts payable + Suppliers}_t} * 12$	-0.18	-0.03	
$\frac{\text{Current assets}_t}{\text{Total assets}_t}$	0.25	0.96	
$\frac{\text{Earnings before taxes}_t}{\text{Equity}_t}$	0.17	0.41	
$\frac{Sales_t}{Sales_{t-1}}$	0.19	0.40	
$\frac{Sales_t}{\text{Total assets}_t}$	0.26	0.21	
$\frac{\text{(Liabilities + Shareholders equity)}_t}{\text{(Liabilities + Shareholders equity)}_{t-1}}$	0.27	0.65	
$\frac{\text{Total liabilities}_t}{\text{Total assets}_t}$	-0.56	-2.49	
$\frac{\text{Financial liabilities}_t}{\text{Sales}_t}$	-0.10	-0.52	
Log(Assets)	0.18	0.13	
Constant	N/A	-1.96	
Functions at groups centroids			
Non-defaulting			0.77
Defaulting			-0.77

Source: Own calculations.

However, the coefficient sign for the number of banking relationships shows that a firm is more likely to be classified as defaulting, if it is heavily connected with the financial intermediaries. This result turns out to be relevant in the scenario where the financial sector credit is concentrated on few debtors. Insofar that it could have a negative impact on the financial stability since if at least one of these heavily connected debtors enter into default, a significant amount of the allocated resources could be lost. Nevertheless, as shown in Figure 5 the debtors with a high number of banking relationships (eight or more) accumulate low shares of the Colombian credit portfolio. However, it is important to point out that the debtors with the highest shares on the credit portfolio (on average 11%) tend to have 5 to 8 banking relationships. Finally, it is noteworthy to mention that the accounts receivable turnover ratio has a negative sign, suggesting that, *ceteris paribus*, a firm with laxest policies in extending credit to customers is more likely to be classified as defaulting.

FIGURE 5: Number of Banking relationships vs. Portfolio's share



Source: Own calculations, based on data from *Superintendencia de Sociedades* and Finance Superintendency of Colombia.

4.1.1. Robustness check

As it was mentioned in the Data section, the analysis made in this paper uses only the latest available information, that is, the one compiled for the period 2010-2011. However, an important step of the model's validation, is to test its ability to classify out-of-sample firms. In particular, the following exercise seeks to test the model's prediction accuracy, through different years. This is done by computing the *z-score* for each firm, with the estimated unstandardized coefficients (see Table 4)[18]. Table 5 presents the estimation results for each year. As it can be seen, the mean overall prediction rate is 71.2%, the maximum is 75.5% and the minimum is 58.9%. It is noteworthy that in seven of the ten analyzed periods, the model's prediction accuracy is higher than 70%. As for the defaulting firms, the results shows that, on average, the model correctly identifies them at a rate of 63.8%. It is important to point out that for the years when a high rate of default was observed (2001 *c.a.* 15%, 2002 *c.a.* 10%) the model obtains high correct classification rates for the defaulting firms.

TABLE 5: Prediction accuracy (percentage)

	Non-Defaulting	**Defaulting**	**Overall**
2010	73.4	65.1	73.0
2009	74.7	59.3	73.9
2008	58.6	63.8	58.9
2007	77.4	57.3	76.4
2006	76.0	58.7	75.5
2005	75.6	57.3	75.0
2004	72.3	65.9	72.0
2003	68.9	65.3	68.7
2002	68.1	71.8	68.4
2001	69.6	73.3	70.1

Source: Authors' calculations.

[18] The cut-off point is calculated using the historical median value of each variable, and then computing the *z-score* with the unstandardized coefficients. Only the information for the 'defaulting firms' is taking into account.

4.2. Probability of Default

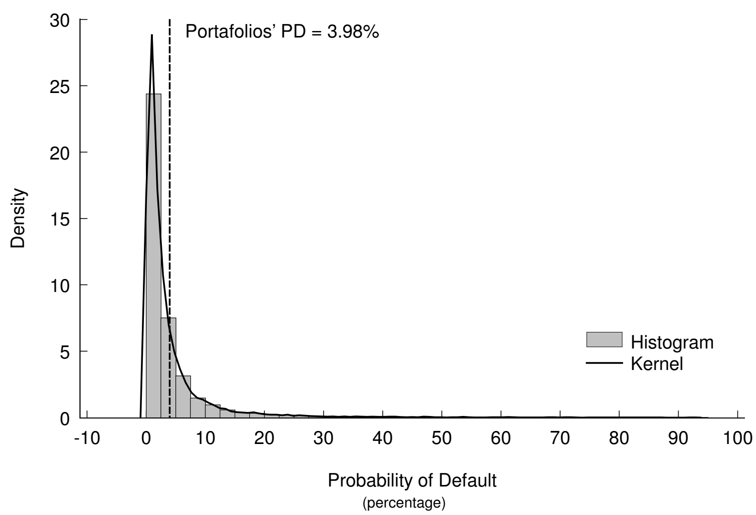
As we already stated, it is important to assess an indicator that gives us an idea of the credit risk exposure of the financial system to the companies that are analyzed in this paper. To do so, we use the set of IFIs that are identified with the DA model, and estimate a Logit model as a function of these variables. Afterwards, we calculate the probability of default for each firm and for the total loan portfolio of these firms.

Table 6 presents the results of the estimated coefficients and the marginal effects of the variables that are included in the Logit model. As it is expected, the results are similar to those found in discriminant analysis[19]; however, there are some variables that are not statistically different from zero. The results show that the variables related with profitability, indebtedness, and number of banking relationships are the ones that contribute the most to explain the probability of default. Thus, a firm with high profitability, low indebtedness and few banking relationships is less likely to default in its financial obligations than a firm with low profitability, high indebtedness and many banking relationships.

In contrast to DA, the duration of the firm's banking relationship, the ratio of accounts receivable to accounts payable plus suppliers, and retained earnings are not significantly different from zero when explaining the PD; nonetheless, the first two have the expected sign.

Using the results of the Logit, we estimated the probability of default for each firm and then its distribution using a Epanechnikov kernel. Figure 6 displays the PD distribution. As can been seen in the figure, the distribution is clearly skewed to the left and the density is highly concentrated in the $[0\%, 5\%]$ interval, meaning that the portfolio's probability of default is relatively low. The average PD is 3.98% and the expected loss is approximately US$854.2. By industry, Fishing and Transportation have the highest PDs, 6.7% and 4.6%, respectively; whereas, Accommodation and Food Services have the lowest, (3.2%).

FIGURE 6: Probability of Default Distribution



Source: Own calculations.

[19]It is important to note that the coefficient sign must be read carefully. The coefficient sign of the Logit have an opposite sign to those of DA due to the difference in the location of the defaulting and non-defaulting regions. In DA default is located at the left, while in the Logit is at the right.

Table 6: Logit Estimation of the Probability of Default and Marginal Effects

Variable	**Coefficient**	**Std. Err.**§	**dy/dx**	**Std. Err.**
Duration of the banking relationship	-0.0290	0.0186	-0.0005	0.0003
Number of banking relationships	0.2767***	0.0180	0.0047***	0.0004
$\frac{\text{Retained earnings}_t}{\text{Total assets}_t}$	0.0033	0.2809	0.0001	0.0048
$\frac{\text{Net income}_t}{\text{Total assets}_t}$	-3.5980***	0.5213	-0.0616***	0.0098
$\frac{\text{Accounts receivable}_t}{\text{Sales}_t} * 12$	0.0315**	0.0092	0.0005**	0.0002
$\frac{\text{Accounts receivable}_t}{\text{Accounts payable + Suppliers}_t} * 12$	-0.0017	0.0071	0.0000	0.0001
$\frac{\text{Current assets}_t}{\text{Total assets}_t}$	-1.0525***	0.2173	-0.0180***	0.0038
$\frac{\text{Earnings before taxes}_t}{\text{Equity}_t}$	-0.3485**	0.1115	-0.0060**	0.0020
$\frac{Sales_t}{Sales_{t-1}}$	-0.6100**	0.1889	-0.0105**	0.0032
$\frac{Sales_t}{\text{Total assets}_t}$	-0.3298***	0.0768	-0.0057***	0.0013
$\frac{(\text{Liabilities + Shareholders equity})_t}{(\text{Liabilities + Shareholders equity})_{t-1}}$	-1.1839***	0.1795	-0.0203***	0.0030
$\frac{\text{Total liabilities}_t}{\text{Total assets}_t}$	2.4209***	0.2939	0.0415***	0.0051
$\frac{\text{Financial liabilities}_t}{\text{Sales}_t}$	1.1023***	0.2453	0.0189***	0.0043
Log(Assets)	-0.3180***	0.0466	-0.0054***	0.0008
Constant	1.9079**	0.6751		
Log-likelihood	-1739.7			
Wald-test	851.44***			
Pseudo R2	0.2132			

§ Bootstrap standard errors, 1000 reps.

Source: Own calculations.

4.3. Stress test

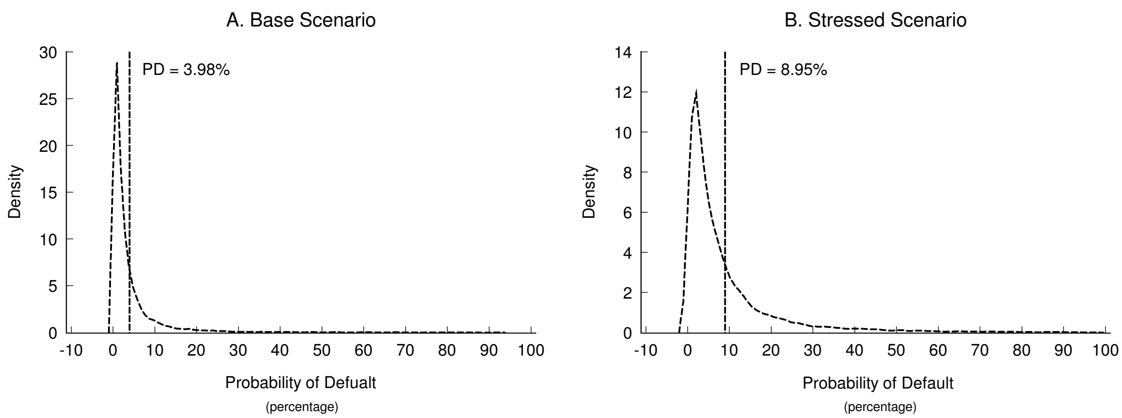
In this section, we make a stress test exercise in order to assess the effect that could have a negative shock to firm's performance. The purpose of the exercise is to analyze the effect of such shock on the portfolio's PD, the bank's expected loss, loan loss provisions, and net income.

We design the stress scenario by adding or subtracting one standard deviation (σ) to a set of the IFIs used to estimate the Logit model. Taking into account the high dispersion that exists among industries, we calculate the standard deviation for each variable by industry. Then, we stress each of the firm's variables by adding or subtracting the specific variable's standard deviation of the industry to which the firm belongs. In Table 12 of Appendix B, we present the mean of each variable by industry before stressing and in Table 13 the mean by industry after stressing the variables. It is important to note, that this exercise is hypothetical and that the results must be analyzed with caution.

Figure 7 presents the results of the exercise on the PD. A shock in the firms performance shifts the PD distribution to right; thus, increasing the likelihood of a default. The average portfolio's PD increases from 3.98% to 8.98%. On the one hand, Mining Industry, and Electricity and Gas Industry are the ones that undertake the most severe shock. The PD of the former increases from 3.6% to 7.5%; while the latter increases from 3.7% to 13.6%. In the other, Agriculture Industry and Education Industry are the ones that suffer the less from the shock. The PD of the former increases from 3.7% to 6.3%; while the PD of the formers goes from 4.0% to 6.7%.

Table 7 presents the effects of the shock on Bank's Balance Sheet and Income Statements. Bank's expected loss increases from $1,659,403 million pesos before the shock to $1,993,001 million pesos after the shock[20]. This implies an increase of loan loss provisions equal to $273,599 million pesos[21], which renders a reduction of net profits by 4%.

FIGURE 7: Probability of Default Distribution



Source: Own calculations.

TABLE 7: Effect of the Shock on Bank's Balance Sheet and Income Statements

	Baseline Scenario	**Stressed Scenario**
Average PD	3.98%	8.95%
Expected loss§	$1,659,403	$1,993,001
Δ *Loan loss provisions*§		$273,599
Bank's net profit§	$6,860,696	$6,587,090
Δ *net profit*		-4.0%

§ Million pesos.

Source: Own calculations.

We also performed this exercise for the whole commercial loan portfolio using the results of the increase in the PD. Thus, we assume the PD to be constant for all loans and calculate the increase in the expected loss given the shock in the PD. The results show that the shock increases the expected loss by 124.8% (Table 8). Such increase renders an accrual in loan loss provisions of $3.3 billion pesos, which causes a depletion of net profits by 48.4%.

[20] The expected loss was calculated as the product of the outstanding loan amounts and the loss given default.

[21] The increase of loan loss provision is calculated as the product of the probability of default and the expected loss.

Table 8: Effect of the Shock on Bank's Balance Sheet and Income Statements

	Baseline Scenario	**Stressed Scenario**
Average PD	3.98%	8.95%
Expected loss	$2,658,525	$5,978,051
Loan loss provisions		$3,319,526
Bank's net profit	$6,860,696	$3,541,170
Δ net profit		-48.4%

§ Million pesos.

Source: Own calculations.

5. Concluding Remarks

In this paper we identify a set of informative financial indicators (IFIs) which may effectively discriminate between firms that are more prone to default on their financial obligations from those that are less prone to do so. For this objective, we apply the multivariate statistical method known as *Discriminant Analysis* (DA). Our research focuses on the loans issued by the Colombian credit establishments to the private corporate sector. Due to the availability and restrictions of the information that could potentially be useful for our analysis, the databases that are used as input for the different estimations correspond to those registered between 2009 and 2011.

We analyze the relationship between financial indicators that are widely used in accountability and corporate finance literature and some new variables (such as the number of banking relationships, duration of the banking relationship, among others) with the likelihood of a firm being classified as defaulting or non-defaulting. This exercise uses as input the financial ratios observed in $t-1$ and the firms' default status registered in period t, to find the most optimal set of IFIs that will be useful for predicting a firm's future distress.

In order to perform the estimations of this paper, we used two kind of samples: *i)* data as it is and *ii)* choice-based sample. As it has been stated, the objective of this paper is to formulate a credit risk model that maximizes the prediction accuracy of defaulting firms. In such way, we select the model that uses a perfectly matched sample and a *stepwise* variable selection procedure. Similarly to the researched literature, we find that a defaulting firm is characterized by having high leverage ratios, low profitability, and liquidity. In addition, we obtain that the number of banking relationships, duration of the banking relationship, and the accounts receivable turnover ratio are good predictors for correctly classifying defaulting firms.

Additionally, we use a Logit model to estimate the probability of default using the IFIs as covariates. Then, we use the Epanechnikov kernel to calculate the probability of default distribution. We find that the distribution is skewed to the left and is highly concentrated at the left tail, suggesting that there is a low probability of default for the loan portfolio. Furthermore, we perform a stress test in order to assess the effects of a shock in firm's performance to the financial system. The exercise shows that the shock shifts the probability of default distribution to the right and that the average probability of default more than doubles. Hence, there is an increase in loan loss provisions and bankť's net profit reduces by 4%.

This paper assessed the most important loan portfolio in Colombia. For further research it would also be important to analyze retail and mortgage loans since they represent almost 40% of the Colombian financial institutions' total loan portfolio.

References

Altman, E. I. (1968), 'Financial ratios, discriminant analysis and the prediction of corporate bankruptcy', *The Journal of Finance* **XXIII**, 589–609.

Brezigar-Masten, A. & Masten, I. (2009), 'Comparison of parametric, semi-parametric and non-parametric methods in bankruptcty prediction', *IMAD Working Paper Series* **XVIII**, 18.

Cruz, J., Villarreal, J. & Rosillo, J. (2002), *Finanzas Corporativas: Valoración, Política de Financiamiento y Riesgo*, Thomson.

Financial Stability Report of Banco de la República (2012), Technical report, Banco de la República.

González, Á. (2010), 'Determinantes del riesgo de crédito comercial en colombia', *Banco de la República, Temas de Estabilidad Financiera* p. 38.

Gutiérrez Rueda, J. (2010), 'Un análisis de riesgo de crédito de las empresas del sector real y sus determinantes', *Banco de la República, Temas de Estabilidad Financiera* p. 37.

International Business Machines Corp. (IBM) (2009), 'Quadratic discriminant analysis in SPSS '. *http://www-01.ibm.com/support/docview.wss?uid=swg21479748

Johnson, R. & Wichern, D. (1998), *Applied Multivariate Statistical Analysis*, 4th edn, Prentince Hall.

Lennox, C. (1999), 'Identifying failing companies: a re-evaluation of the Logit, Probit and DA approches ', *Journal of Economics and Business* **46**(4), 347–364.

Ortega, J., Martínez, J. & Valencia, J. (2010), 'El modelo de calificación crediticia z-score: Aplicación en la evaluación del riesgo crediticio de hb fuller colombia ltda.', *Revista MBA EAFIT* pp. 102–111.

Peña, D. (2002), *Análisis de datos multivariantes*, McGraw-HILL/INTERAMERICANA DE ESPAÑA, S. A. U.

Appendix A. Assessing if sector may be used as a discriminating variable

At a first glance, the *defaulting* firms sectorial composition is not significantly different with respect to that registered for *non-defaulting* firms. From Figure 1 of the text, it is observed that Wholesale and Retail trade, Manufacturing, and Construction are the economic sectors with the highest participation[22]. However, the sectors' ordering of the first group of firms does not necessarily match with that observed for the second one. For example, the sector with the highest participation in the defaulting firms group is Manufacturing (34%), while Wholesale and Retail trade has the highest participation for the non-defaulting firms (38,5%). On the other hand, the sectors' participation is different for the two groups of firms.

These results suggest that sector is a variable which could allow to effectively discriminate and classify a Colombian private corporate sector firm as *defaulting* or *non-defaulting*. With the aim of contrasting this hypothesis, two discriminant analysis models are estimated in order to evaluate the impact that a firm's sector has in the model classification. Since we are analyzing twelve sectors, dummy variables must be incorporated in order to include the sector variable into the estimations. The first estimation incorporates the forty microeconomic variables[23] which are analyzed in this study using the *stepwise* selection procedure. Alternatively, the second estimation takes into account these variables and eleven additional *dummy* variables which model a firm's sector[24].

On Table 9, the results for the two estimated models are presented. The second model (Model A-2) has a a slightly higher overall accurate classification rate (20 basis points higher) in comparison to that obtained for the first model (Model A-1)[25]. The accurate classification rate for defaulting firms and for the overall sample of firms in the sample are identical for the two estimations. In this way, these results suggest that the incorporation of the *dummy* variables do not guarantee a significant improvement in the classification capability of the model.

Table 9: Prediction accuracy (percentage)

	Non-Defaulting	**Defaulting**	**Overall**
Model A-1	98,7	18,6	95,5
Model A-2	98,7	18,8	95,5

Source: Authors' calculations.

The estimations of the two models have in general the same variables, as it can be observed in Table 10[26]. Even though the coefficients' magnitude of the variables that are present in both models are

[22] A sector participation is calculated as the ratio of the number of firms that belong to that sector in particular adjusted by the total number of firms in the group

[23] Thirty seven financial ratios, duration of the banking relationship, number of banking relationships, number of creditors to which a firm has a delinquency loan.

[24] It can be noted that, at most eleven *dummy* variables should be incorporated into the estimations, in order to avoid multicollinearity. Analogously, this estimation is done using the *stepwise* selection approach.

[25] 'Model A-1' and 'Model A-2' correspond respectively to the first and second estimations as referenced on the previous paragraph.

[26] It is important to point out that 'Model A-1' incorporates the variables $\frac{Non-operating\ revenue_t}{Sales_t}$ (X4) and $\frac{Gross\ Income_t}{Equity_t}$ (X27) in the estimation, while 'Model A-2' does not. On the other hand, the latter model incorporates into the final estimation the variables $\frac{Gross\ income_t}{Liabilities\ +\ Shareholders\ equity_t}$ (X22) and $\frac{Selling\ and\ administrative\ expenses_t}{Sales_t}$ (X3), which are not incorporated in the output estimation of the former model. In addition, the second model includes the estimated coefficients for the *dummy* variables that characterize the sectors of Wholesale and Retail and Mining

different, sign is preserved. In addition, the estimation of 'Model A-2' incorporates the *dummy* coefficients for the sectors of Mining ($Dummy_1$) and Wholesale and Retail trade ($Dummy_2$); suggesting that these sectors can be good predictors for effectively discriminating defaulting firms from non-defaulting. Finally, the canonical correlation coefficients[27] for the discriminating functions and Wilks Lambda[28] for the estimated models, corroborate the poor results presented in Table 9: the covariates of the final estimated discriminant functions for 'Models A-1 and A-2' are inadequate for effectively discriminating between the two types of firms. In this way, we conclude that a firm's sector is a variable that do not contributes significantly for classifying a Colombian real sector firm as defaulting or non-defaulting.

TABLE 10: Standardized coefficients for models A-1 and A-2

Model A-1		**Model A-2**	
Variable name/number	**Standardized Coefficient**	**Variable name/number**	**Standardized Coefficient**
Vencimientos	0.126	Vencimientos	0.124
Number of banking relationships	0.389	Number of banking belationships	0.394
X24	-0.162	X24	-0.216
X4	-0.155	X3	-0.323
X5	0.274	X5	0.194
X6	-0.227	X7	-0.105
X7	-0.127	X9	0.137
X9	0.138	X11	0.154
X11	0.145	X38	0.092
X38	0.091	X42	0.146
X35	0.161	X36	-0.271
X36	-0.299	X29	-0.266
X29	-0.177	X1	-0.131
X1	-0.120	X2	-0.148
X2	-0.110	X13	-0.149
X13	-0.151	X22	0.131
X14	0.164	X14	0.132
X33	0.346	X33	0.359
X17	0.282	X17	0.287
Log(Total Assets)	-0.426	Log(Total Assets)	-0.435
		$Dummy_1$	0.080
		$Dummy_2$	-0.068

Source: Authors' calculations.

TABLE 11: Wilks Lambda and Canonical Correlation for the Discriminant functions of Models A-1 and A-2

	Estimated Wilks Lambda ($\hat{\Lambda}$)	**Estimated Canonical Correlation**
Discriminant function for Model A-1	0.903*	0.311
Discriminant function for Model A-2	0.902*	0.313

Source: Authors' calculations. * Significant at 1% confidence level (p-value $\cong$ 0)

[27] As stated by Peña (2002), "canonical correlation is used when a set of multivariate variables can be divided into two homogeneous groups, ... since it is expected to study the relationship between both set of variables (pp. 489)". The low values obtained for the discriminant functions, in the two estimated models, suggest that the discriminating variables in each function are not appropriate for differentiating into the groups.

[28] Following Peña (2002), Wilks Λ, can be defined as: $\Lambda = \frac{|S_1|}{|S_1|+|S_2|}$, where S_1 corresponds the within-group sum-of-square matrix and S_2 denotes the between-group-sum-of-square-matrix. Λ is used for contrasting if the mean vectors (centroids) of the groups are different or not. From Table 11, it can be observed that the estimated $\hat{\Lambda}$ for the discriminant function of each of the two models are highly significant given that the probability of rejecting the null hypothesis for the groups' mean vectors equality test ($\mu_{defaulting} = \mu_{non-defaulting}$) is negligible (*c.a.* 0%). However, the estimated Wilks lambdas are 0.903 and 0.902, for the 'Models A1 and A2' respectively; indicating that the within-group variability is low and that the two groups (defaulting and non-defaulting firms) are not significantly different.

Appendix B. Stressed Scenario

Table 12: Variable Means in the Baseline Scenario

Industry	**x26**	**x24**	**x9**	**x12**	**x36**	**x31**	**x1**	**x2**	**x13**	**x15**	**x17**
A	0.008	0.008	3.758	4.278	0.352	0.043	1.052	0.901	1.115	0.409	0.249
B	0.017	0.019	2.323	2.006	0.490	0.011	1.000	1.234	1.141	0.466	0.250
C	0.093	0.045	4.512	2.779	0.466	0.180	1.385	1.379	1.396	0.522	0.148
D	0.045	0.028	3.814	2.589	0.602	0.120	1.073	1.335	1.146	0.517	0.188
E	0.050	0.061	3.745	4.282	0.665	0.235	1.037	1.975	1.271	0.583	0.221
F	0.087	0.038	5.667	6.468	0.703	0.188	1.287	1.366	1.193	0.600	0.190
G	0.079	0.035	3.335	3.005	0.763	0.171	1.083	2.026	1.150	0.582	0.144
H	0.030	0.036	2.335	2.024	0.399	0.187	1.077	1.589	1.159	0.495	0.166
I	0.030	0.040	5.634	4.688	0.621	0.184	1.102	1.455	1.174	0.573	0.169
K	0.082	0.058	5.359	6.751	0.612	0.224	1.168	1.787	1.183	0.494	0.152
M	-0.010	0.026	2.409	5.477	0.345	0.105	1.087	1.520	1.102	0.526	0.158
N	0.082	0.056	5.672	6.581	0.585	0.219	1.128	1.463	1.163	0.508	0.130
O	0.032	0.043	4.109	4.664	0.480	0.166	1.101	1.539	1.201	0.476	0.157

Source: Own calculations.

Table 13: Variable Means in the Stressed Scenario

Industry	**x26**	**x24**	**x9**	**x12**	**x36**	**x31**	**x1**	**x2**	**x13**	**x15**	**x17**
A	-0.158	-0.067	-1.362	-6.472	0.103	-0.198	0.677	0.007	0.699	0.152	0.028
B	-0.253	-0.053	0.741	0.208	0.290	-0.312	0.848	0.565	0.877	0.227	-0.005
C	-0.126	-0.075	-1.656	-0.885	0.232	-0.231	0.585	0.136	0.704	0.288	-0.039
D	-0.152	-0.055	0.283	-1.734	0.373	-0.161	0.738	0.508	0.756	0.308	0.005
E	-0.014	-0.005	1.410	1.523	0.401	0.021	0.680	-0.108	0.605	0.381	0.015
F	-0.059	-0.037	-0.889	-4.895	0.472	-0.082	0.433	0.282	0.576	0.382	-0.019
G	-0.097	-0.049	-0.196	-3.303	0.543	-0.133	0.714	0.447	0.741	0.372	-0.017
H	-0.148	-0.040	-0.903	-1.350	0.118	-0.120	0.778	0.279	0.725	0.238	-0.014
I	-0.154	-0.055	0.152	-3.056	0.345	-0.152	0.685	0.199	0.711	0.362	-0.016
K	-0.100	-0.044	-1.058	-4.688	0.314	-0.125	0.660	0.006	0.653	0.255	-0.045
M	-0.181	-0.032	-2.008	-5.417	0.076	-0.151	0.815	0.515	0.801	0.300	0.000
N	-0.125	-0.030	-0.451	-6.871	0.319	-0.030	0.796	0.014	0.703	0.289	-0.023
O	-0.170	-0.045	-1.002	-3.702	0.181	-0.120	0.704	0.139	0.716	0.256	-0.032

Source: Own calculations.