THE BANKRUPTCY RISK IN INFRASTRUCTURE SECTORS:
AN ANALYSIS FROM 2006 TO 2018

VANDERSON A. DELAPEDRA-SILVA

https://orcid.org/0000-0002-3937-9914


1 University of Brasília (UnB) and Bank of Brasil, Brasília, DF, Brazil.

This is an open-access article distributed under the terms of the Creative Commons Attribution License.
ABSTRACT

Purpose: This research aims to identify the probability of default of infrastructure companies considering the sector specificities of their activities. In addition, the work seeks to identify the application of structural variables of probability of default in a model in a reduced way in order to identify the significance of its use. For this purpose, we investigated 1,520 North American companies from six different sectors linked to infrastructure.

Originality/value: The analyzes carried out to identify the probability of a company going bankrupt hardly consider its sectorial particularity. Although most models bring important inputs for risk assessment, most of them do not consider this sectoral view. Then, this work has as value and originality the contribution to fill this gap and identify the existence of sectorial differences in the analysis of default risk in infrastructure companies in the North American market in the period between 2006 and 2018.

Design/methodology/approach: The study performed a logistic regression (logit model) using 11 model variables established in calculating the probability of default. It also used the variable distance to default as an explanatory variable in order to identify its ability to explain the researched phenomenon.

Findings: The study identified that, in addition to the size of the companies, the distance to default variable is the only variable that can be applied with significance in all the analyzed sectors. In addition, it was identified that companies in the oil and gas sector have less sensitivity to this variable than companies in the other sectors.

KEYWORDS

Sectoral risk. Distance to default. Infrastructure. Merton model. Credit risk.
1. INTRODUCTION

In an increasingly dynamic and evolving world, an analysis of the financial performance of companies linked to infrastructure plays a central role in guaranteeing the countries’ economic development. However, the diversity of characteristics of the sectors to which these companies are linked can compromise an effective risk analysis. In this sense, a risk analysis capable of identifying these differences is of great importance so that its access to credit is not compromised by distorted or standardized analyzes. This article aims to identify these sectoral differences in measuring the default risk of companies linked to different infrastructure sectors using a logistic regression model as a tool. As a result, it was possible to identify the significance of the distance to default variable playing the role of explanatory variable of the model. In addition, it was possible to verify that this variable has different sensitivities depending on the sector where it is used.

One of the most sensitive factors in the study of finance is the search for mechanisms to estimate the probability of default. This effort has triggered the emergence of several models aimed at solving this problem. Among them, models referenced in asset pricing techniques applied to the study of corporate liabilities were the pioneers in the task of modeling the company’s default and linking it to an economical pricing model (Chen & Wu, 2014; Lando & Nielsen, 2010).

Merton (1974) contributed to clarifying this application, creating an analogy between the capital structure of the companies and the idea of options on their assets. In the face of that, a company’s net worth could be compared to a European purchase option on its assets, in which the exercise price of these options would be the value of their debt.

Currently, two theoretical approaches are considered for measuring the probability of companies defaulting. One of them is based on the so-called structural models, in which the contributions of Black and Scholes (1973) and Merton (1974) have improved the use of structural variables of companies, especially the value of their assets, to assess the likelihood of credit risk.

The second approach is based on the so-called reduced models that allow identifying the occurrence of bankruptcy regardless of the evolution of the company’s structural data. The reduced models are based on mechanisms focused on the search for stochastic risk rates, where the default probability dynamics do not depend on the credit recovery rate (Allen & Saunders, 2002; Altman, Resti, & Sironi, 2004).
The analysis of the correlation between common dynamic latent factors between companies may explain a considerable portion of the default risk, but sectoral and macroeconomic factors are generally ignored by conventional models (Chen & Wu, 2014; Duffie, Eckner, Horel, & Saita, 2009; Jorion & Zhang, 2009).

However, according to Giesecke (2004), there has not been enough research on approaches that incorporate the interdependence of default between companies. Therefore, developing consistent models to achieve this goal is still a challenge (Chen & Wu, 2014; Escribano & Maggi, 2018).

This article intends to use several explanatory variables to answer the following research questions:

• Which variables affect the probability of default in companies of the infrastructure sector in the analyzed sample?
• How do the different infrastructure sectors react to the variables in the proposed model?

This paper is divided into five parts. Besides this introduction, we will analyze some mechanisms for credit risk measurement supported by the default probability measurement, which is provided by the KMV-Merton model, as well as the Z-score models proposed by Altman (1968), the O-score proposed by Ohlson (1980), and the Zmijewski’s (1984) model. In the third part, we will show the methodology used in the study, as well as the data sources used for the sectoral impact measurement in the calculation of credit risk of infrastructure companies. In the fourth part, the results and the analyzes will be shown, and finally, in the last part, we will write the final remarks, presenting the limitations of this study, as well as suggestions for new research opportunities.

2. DEFAULT RISK

In finance, the term “risk” is associated with the impossibility of predicting future events. Thus, risk entails the probability of results different from those which had been expected, including negative and positive cases. Because both types of results depart from expectation, they are both susceptible to being classified as risk (Crosbie & Bohn, 2003; Damodaran, 2012). Looking at this, the seminal work of Markowitz (1952), which associated the return variance as an investment risk measure, became one of the most important bases for the literature on calculating investment risk, as well as
for the emergence of new approaches on the subject (Soleimani, Golmakani, & Salimi, 2009).

The theoretical approach of credit risk can be divided into at least two distinct schools of thought (Allen & Saunders, 2002; Zhou, 2001). In the first one, the structural approach, the default probability analysis concentrates on the evolution of a company’s value. In this manner, default occurs when a company’s market value reaches a critical level that is delimited by its debt value (Duffie & Lando, 2001; Vasicek, 1984; Zhou, 2001).

The second generation of models is based on reduced-form models focused on the search for stochastic risk rates, in which the dynamics of default probability are independent of the credit recovery rate, and both bear no relation to a company’s structural characteristics. In the reduced form models, the focus has become the assessment of potential contractual loss, given a time horizon and a trust level, in which the event of default is seen as a random event.

There is also a third generation of models called temporal analysis models, which are based on the analysis of the probability of a default event occurring in a given period (Duffie, Saita, & Wang, 2007).

### 2.1 KMV-Merton

The KMV-Merton model stems from the assumption that the capital structure of a company can be equated to a series of options on its assets. In this way, the equity could be seen as a buying option on the company’s assets, and the exercise price could be represented by its debt value. In this sense, the equity of a company is shown as a function of the company’s value. Then, the value of a company in time $t$ can be given by:

$$ V_t = S_t + D_t $$  \hspace{1cm} (Equation 1)

in which:
- $V_t$ = the company’s value at time $t$;
- $S_t$ = its assets value;
- $D_t$ = its debt value.

The dynamic variation of the company’s value, in turn, can be defined by the stochastic differential equation that defines a geometric Brownian process.
\[ dV_t = \mu V_t dt + \sigma V_t dW_t \]  \hspace{1cm} (Equation 2)

In which: \( \mu \) represents the expected growth rate for the firm’s value per unit of time; \( \sigma V_t \) represents the volatility of the firm’s value over a given period; \( dW_t \) represents the standard Brownian motion.

Since the company’s equity is treated as a call option on the company’s value and the exercise price given by the face value of its obligations, the company’s equity is presented as a function of the company’s value as follow:

\[ E = V \mathcal{N}(d_1) - e^{-rT} F \mathcal{N}(d_2) \]  \hspace{1cm} (Equation 3)

In which: \( E \) expresses the company’s equity; \( F \) expresses the face value of its debt; \( r \) represents the risk-free rate; \( \mathcal{N} \) represents the accumulated standardized normal distribution function; and \( d_1 \) is given by:

\[ \ln \left( \frac{V}{F} \right) + (r + 0.5\sigma_v^2)T \]  \hspace{1cm} (Equation 4)

And the term \( d_2 \) is given by the following equation:

\[ d_2 = d_1 - \sigma_v \sqrt{T} \]  \hspace{1cm} (Equation 5)

The model also relates the volatility of the firm’s value to the volatility of its equity.

\[ \sigma_E = \left( \frac{V}{E} \right) \frac{\partial E}{\partial V} \sigma_v \]  \hspace{1cm} (Equation 6)

Once the Black-Scholes-Merton (BSM) model demonstrates that \( \frac{\partial E}{\partial V} = \mathcal{N}(d_1) \), then:

\[ \sigma_E = \left( \frac{V}{E} \right) \mathcal{N}(d_1) \sigma_v \]  \hspace{1cm} (Equation 7)

Bharath and Shumway (2008) point out that the model uses two non-linear equations (3 and 7) to express the volatility value of the company’s equity within a certain probability of default.

Unlike the total value of the company, the value of the option can be seen from the value of the company’s equity, with the variation of this
equity being estimated from past returns. At the same time, the company’s obligations can be obtained at face value of its debts.

The value of a company can be represented by the price at which all of that company’s bonds can be bought or sold. The calculation of the market value of the company’s assets is different from the book value of these assets since the market value considers their potential for future performance (Allen & Saunders, 2002). Vasicek (1984) points out that both equities can be obtained by multiplying the number of shares in the company by their price, and the value of debts can be measured by the price of debt at nominal interest rates.

With a risk-free rate, in addition to the estimates raised above, it is possible to solve equations 3 and 7 simultaneously, obtaining the values of \( V \) and \( \sigma_v \), so that you can finally calculate the distance to default (DD).

\[
DD = \frac{\ln\left(\frac{V}{F}\right) + \left(\mu - 0.5\sigma_v^2\right)T}{\sigma_v \sqrt{T}}
\]  
(Equation 8)

in which: \( \mu \) represent an estimate of the annual return on company assets.

The use of the DD has been criticized for its lack of accuracy and efficiency in the task of forecasting corporate default. This fact has put pressure on the emergence of other alternatives to achieve this goal (Bharath & Shumway, 2008; Chen & Wu, 2014).

The assumptions of Merton’s DD, such as the evolution of the company’s value through a geometric Brownian movement and the uniformity of its debts, open a way for reduced models to be able to explain the probability of default, which is based on the probability that the company’s value is less than the face value of its debt, given a time horizon (Bharath & Shumway, 2008).

Finally, Merton’s model postulates that the probability of default (PD) is the difference between the probability function of a normal distribution and the DD.

\[
PD = \mathcal{N}(-DD)
\]  
(Equation 9)

Bharath and Shumway (2008) highlight that this model employs two non-linear equations to express the volatility value of the company’s equity within a certain probability of default. The use of a risk-free rate, associated with the assumptions of the company’s equity value, allows the calculation
of the DD, a measure that expresses the distance that a company has until it falls in default.

The KMV can be understood as an extension of Merton’s model and employs the same modeling logic of the company’s asset valuation and its ability to overcome the limit value of its debts. One of the variations of the KMV-Merton model is also employed by the American credit rating agency Moody’s. The model used by this company can also be called KV model, and one of the basic differences in this version is that it is based on the measurement of several classes and debt maturity instead of considering it fixed in time (Bharath & Shumway, 2008; Kliestik, Misankova, & Kocisova, 2015).

Bharath and Shumway (2008) adapted the KMV-Merton model to create an alternative they called naïve, in which they bring the market value of the debt closer to its face value, and in doing so they reach the value and the market variation of the assets by means of a ponderation of the company’s equity variation.

This adjustment allowed an improvement in the performance of the Merton-KMV model when used as an explanatory variable of the default forecast; according to the authors, the KMV-Merton naïve model could be expressed as follows:

\[ \text{Naïve}\sigma_D = 0.05 + 0.25\sigma_E \]  

(Equation 10)

in which:

\[ \text{Naïve}\sigma_v = \frac{E}{E+F}\sigma_E + \frac{F}{E+F}(0.05 + 0.25\sigma_E) \]  

(Equation 11)

in which: \( \mu = r_{t-1} \).

Then, the results of the distance to default naïve can be expressed as follow:

\[ \text{Naïve } DD = \frac{\ln\left(\frac{E+F}{F}\right) + (r_{t-1} - 0.5\text{Naïve}\sigma_v^2)T}{\text{Naïve } \sigma_v \sqrt{T}} \]  

(Equation 12)

### 2.2 The reduced models of Altman (1968), Ohlson (1980), and Zmijewski (1984)

The reduced form models were an evolution of the structural models, which related the default to the fall in the market value of the company’s
assets below a certain level, usually the value of its debt. Thus, latent variables also started to be used as predictors of default (Duffie & Lando, 2001). The first modern univariate model of default prediction concluded that cash flow on total debt could be considered the most relevant explanatory variable in this forecast (Beaver, McNichols, & Rhie, 2005; (Beaver, McNichols, & Rhie, 2005; Hillegeist, Keating, Cram, & Lundstedt, 2004).

In this vein, Altman (1968) proposed a multivariate model that could explain the corporate default using the discriminant analysis method (Lando & Nielsen, 2010). Altman (1968) developed a model called Z-score, which is based on variables with the greatest significance in a model of multivariate discriminant analysis. Thus, the analysis is expressed by \( Z = v_1 x_1 + v_2 x_2 + \cdots + v_n x_n \), in which values of variables are transformed into a discriminating sequence of \( Z \) (Allen & Saunders, 2002; Hillegeist et al., 2004; Taffler, 1984).

In addition, \( v_1, v_2, \cdots, v_n \) are discriminant coefficients, and \( x_1, x_2, \cdots, x_3 \) are independent variables. The final discriminant function proposed by Altman (1968) is expressed by:

\[
Z = 0.12X_1 + 0.14X_2 + 0.33X_3 + 0.006X_4 + 0.999X_5 \quad \text{(Equation 13)}
\]

in which:
- \( X_1 = \text{working capital/total assets} \): measuring liquid assets in relation to the size of the company;
- \( X_2 = \text{retained earnings/total assets} \): measuring the profitability that reflects the company’s age and earning potential;
- \( X_3 = \text{Ebit/total assets} \): measuring operating efficiency without the impact of tax and leveraging factors. In this case, operating earnings are considered very important to analyze the long-term viability;
- \( X_4 = \text{equity/liabilities} \): measuring the market dimension of the company. In this case, the equity considered is the market value of equity;
- \( X_5 = \text{sales/total assets} \): measuring the total asset turnover.

Proposing an improvement of this model, Ohlson (1980) adopted the logistic analysis in substitution to the linear discriminant analysis previously adopted by Altman (1968). The model started to be called the O-Score model (Jayasekera, 2018; Lando & Nielsen, 2010). Ohlson’s (1980) model also made it possible to identify elements that statistically influence the probability of default of companies in the period of one year, such as the size of the company; financial structure measures; performance measures; and measures related to the company’s current liquidity. This model was based
on the observation of 105 companies that went bankrupt compared to 2,058 companies that did not. Then, the author proposed three models with different dependent variables, that is, 1. bankruptcy forecast within a year; 2. bankruptcy forecast within two years, as the company does not go bankrupt in subsequent years; and 3. bankruptcy forecast within a year or two. The explanatory variables of the models are:

- \( \text{SIZE} = \log(\text{total assets/price level index}) \)
- \( \text{TLTA} = \text{total liabilities/total assets} \)
- \( \text{WCTA} = \text{working capital/total assets} \)
- \( \text{CLTA} = \text{current liabilities/total assets} \)
- \( \text{OENEG} = 1 \) if the liability exceeds the asset, 0 if not
- \( \text{NITA} = \text{net income/total assets} \)
- \( \text{FUTL} = \text{operating funds/total liabilities} \)
- \( \text{INTWO} = 1 \) if net income was negative for the last two years, 0 if contrary; and
- \( \text{CHIN} = \frac{(\text{NI} - \text{NI}(t - 1))}{(|\text{NI}| + |\text{NI}(t - 1)|)} \), in which \( \text{NI} \) represents net income for the most recent period. In this way, the denominator would have the premise to act as a leveler and the variable, as a measure of change in the net income.

Being \( X_i \) a vector of predictor variables for the \( i^{th} \) observation, \( \beta \) a vector unknown parameter, and \( P(X_i, \beta) \) the default probability for each \( X_i \) and \( \beta \). \( P \) will represent a logarithmic function of probability between \( 0 \leq P \leq 1 \).

\[
\ln(P(X_i, \beta)) = \sum_{i \in S_1} \log(P(X_i, \beta)) + \sum_{i \in S_2} \log(1 - P(X_i, \beta)) \quad \text{(Equation 14)}
\]

in which: \( S_1 \) represents the index of firms that went bankrupt; and \( S_2 \) the number of companies that did not go bankrupt within a given time.

Zmijewski (1984), in turn, extended Ohlson’s (1980) approach to implement a probit model as a methodological alternative for calculating risk default (Jones & Hensher, 2008; Platt & Platt, 1990). In his model, Zmijewski (1984) applies the following explanatory variables, which influence the likelihood of a company entering the standard.

\[
B = \alpha_0 + \alpha_1 \text{ROA} + \alpha_2 \text{FINL} + \alpha_3 \text{LIQ} \quad \text{(Equation 15)}
\]

in which:
- \( B = 1 \) if the company goes to Default and 0, otherwise;
- \( \text{ROA} = \text{net income/total assets} \);
FINL = total debts/total assets; and LIQ = current assets/total liabilities.

Finally, it is worth noting that the models highlighted by Altman (1968), Ohlson (1980), and Zmijewski (1984) are static models and do not consider the changes that occur in the company over time. This characteristic produces some inconsistencies in the estimation of the probability of default (Shumway, 2001).

### 2.3 Time models

According to Duffie et al. (2007), there is still a third generation of models called temporal analysis. This model was introduced to assess risk events in the area of finance by Lane, Looney, and Wansley (1986). Later, Lee and Urrutia (1996) compared a duration model with a logit model default prediction of insurance companies and concluded that these models are superior with the identification of more significant variables than the logit model.

According to Shumway (2001), most studies that estimate the probability of default do not consider the time variable in their analyzes. According to the authors, by ignoring the changes of companies over time, static models that estimate the probability of default are biased and without consistency. In this sense, Shumway (2001) developed a model that investigates several information available to calculate the default risk of companies from each point in time: an approach comparable to a multi-period logit model (Duffie et al., 2007).

Hillegeist et al. (2004) also used a discrete time model to estimate the default risk. In their model, the authors combined macroeconomic and accounting variables, together with the DD variable, for explaining the variation in default probabilities between companies.

LeClere (2002) reviewed a proportional risk model and compared the choice of time-dependent covariates with non-time-dependent covariates. The author suggests that the choice of covariates with temporal dependence substantially influences the estimation of models of this nature.

Finally, proportional risk models are very popular in research default risks, mainly due to two characteristics: the first is providing information on the length of time between a given point of origin and the occurrence of one event; and the second, in contrast to most of the other survival analysis models, the proportional risk models are semi-parametric, not requiring the specification of a single distribution for modeling the relationship between events and time (LeClere, 2002).
3. DEVELOPMENT AND INFRASTRUCTURE

Infrastructure can be understood as the set of structures and networks that connect cities and metropolitan areas to social and economic activities. Some examples of infrastructure are streets, roads, basic sanitation, telecommunications etc. (Grimsey & Lewis, 2002).

Any basic project of economic growth in a country needs to take into account alternatives for strategic investments in infrastructure. These can be designed with the intention of maintaining what has already been built or with new investments in mind. Thus, it is paramount that investments in infrastructure are designed in a way that contemplates strategic information, such as priorities, the role of the private sector, sources of financing, quality requirements, among others (Ruiz-Nuñez & Wei, 2015).

3.1 Infrastructure funding

Infrastructure projects are marked by demanding a large amount of capital, with payment terms diluted over a long-term horizon; such difficulties pave the way for partnerships between public and private agents in order for the operation of infrastructure services to be financially sustainable (Smithson & Hayt, 2001).

Until the 1990s, governments were the leading investors in infrastructure in the world. However, the lack of public capital available for funding constructions, especially in developing countries, drove the usage of private sector investments in these projects away, contributing to the reduction of risks associated with management inefficiency (Grimsey & Lewis, 2002; Kumari & Sharma, 2017).

From the year 2000 onwards, limitations caused by stricter capital regulations, in addition to changes in the macroeconomic environment and the inefficiency of public management, have encouraged greater participation of institutional investors in these projects (Della Croce & Gatti, 2014).

In this sense, Sharma and Vohra (2008) highlight the importance of private investment in infrastructure because it not only provides the large amount of capital needed in these projects but also provides more effective operational techniques, improves the capacity to meet deadlines, and offers more innovative technologies.

Alternatively, public-private partnerships (PPP) made it possible to bypass this problem since its constitution aligns public and private interests
in infrastructure projects. PPP’s, in turn, are guided by the promotion of
greater transparency, greater efficiency in the provision of infrastructure ser-
vice, and greater compliance among those responsible for these services
(Grimsey & Lewis, 2002; Mustafa, 1999).

Despite the differences in interests between public and private agents
representing a problem for the development of PPPs as tools for financing
infrastructure (Hart, Shleifer, & Vishny, 1997; Mustafa, 1999), the estab-
ishment of a cohesive institutional structure, in addition to contractual
mechanisms that mitigate these distortions, can contribute to overcoming
these impasses (Mustafa, 1999).

3.2 Sectoral influence in calculating default risk

The investigations of the sectorial influence on the calculation of default
risk have always taken into account an individual analysis of the companies,
measuring the impact of specific variables in the probability of bankruptcy,
while ignoring the factors present in the interrelation between these compa-

In this way, the calculation of relative indices of the industrial sector
(value of the divided by the average of the sector value) was an alternative
found to mitigate these distortions. Platt and Platt (1991) compared models
containing these indicators relating to models without this adjustment
and concluded that models containing indicators relative measures are more
effective than models with indicators without sector adjustment.

With the emergence of this adaptation, new studies on risk modeling
of credit emerged, making it possible to combine the use of structural
explanatory variables and aggregated sector indicators to measure business
default (Izan, 1984; Platt & Platt, 1990). In addition, other alternatives
capable of capturing the correlation of common weaknesses between com-
panies also contributed to the investigation of the sectoral influence on cor-
porate performance.

In these cases, two distinct mechanisms can show the correlation of
default between companies. The first one is based on the recognition that
the financial health of any company is correlated to macroeconomic factors;
the second mechanism is based on the unmeasured direct links between
these companies and deserves greater attention due to the potential for
harmful contagion between them (Giesecke, 2004; Pu & Zhao, 2012).

Duffie and Garleanu (2001), in addition to Jarrow and Yu (2001), sug-
gest that this correlation of default can be induced by the existing intensity
between the events of default of companies, which would be able to expose the dependence on common factors of default.

Koopman, Lucas, and Schwaab (2012) identified the influence of systematic risks on the variation of the default risk. For the authors, systematic factors correspond to about 35% of the variation in the rate of insolvency of American companies, being 25% derived from sectorial weaknesses. The accumulation of systematic risks is also evident in previous periods and during financial crises.

Chen and Wu (2014) emphasize that conventional default prediction models underestimate the influence of non-observable factors on the insolvency correlations of companies. Fragilities that are observable in macro-economic and sectorial factors have a strong influence on the intensity of insolvency.

In contrast, Duffie et al. (2007) use a sensitive to variations time model to estimate the probability of default of companies in the industrial machinery and instruments sector and identify that the performance of profits sectoral measures, measured from the sector average among the companies, is only significant when used alone.

When the variable related to sectorial profits is used in conjunction with the distance to default variable and the growth of individual companies’ revenue, its performance does not play a significant role (Duffie et al., 2007).

Finally, the inclusion of sectorial effects in the calculation of companies’ default risk has increasingly highlighted the importance of this mechanism for improving models of bankruptcy prediction. On top of that, the clear existence of common default risk factors related to the industrial sector can be a way to prevent insolvency rates from being wrongly estimated, in addition to contributing to the emergence of more efficient default prediction tools.

4. METHODOLOGY

According to Jayasekera (2018), some methodological paths contributed to the prediction exercise of a company, among them, mathematical models based on neural networks, market-based models, and statistical models such as logit/probit regressions, and models of discriminant determinations (Allen & Saunders, 2002).

Therefore, this work uses the logistic regression model (logit) used in situations in which the dependent variable is binary categorical, and the other variables can be both numerical and categorical.
4.1 Logit model

The linear default probability models use previous data to explain past loan payment data and then estimate default probabilities in future loans. However, when evaluating the occurrence or non-occurrence of a given event, the estimated probabilities of default may be outside the range 0 and 1, bringing information that is not relevant to the analysis (Altman & Saunders, 1997; Saunders & Thomas, 1997).

To overcome this problem, the logit and probit models allow the dependent variable to assume a qualitative binary choice format, which indicates the occurrence or non-occurrence of a particular event, such as the default of a company (Wooldridge, 2010).

In a logistic regression, the focus is based on the logistic transformation of $\pi (x)$, given by:

$$g(x) = \ln \left[ \frac{\pi(x)}{1 + \pi(x)} \right] = \beta_0 + \beta_1 x$$

(Equation 16)

The method used to estimate parameters for logistic regression models is the maximum likelihood method, which has the objective of producing values for the parameters capable of maximizing the probability of obtaining the set of data that are observed.

4.2 Model and variables

In the making of this essay, we used data from balance sheets, income statements of the financial year, and the price of American companies’ stocks listed in the stock market and belonging to specific subgroups of infrastructure, based on the Global Industry Classification Standard (Gics). Thus, the accounting variables belonging to the models of Altman (1968) and Zmijewski (1984) and a variable pointed out by Lennox (1999) were considered the most relevant for these types of models. Hence, they were included in the model of our study.

| Figure 4.2.1 |
| VARIABLES |

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>Altman (1968)</td>
<td>SZ</td>
</tr>
<tr>
<td>Working capital/total assets</td>
<td>Altman (1968)</td>
<td>WK</td>
</tr>
</tbody>
</table>

(continue)
In addition to these accounting variables, the model included the size of the firm (SZ) (Ohlson, 1980; Lennox, 1999; Shumway, 2001), given by the natural logarithm of its assets.

The inclusion of the explanatory variable DD was justified by the works of Duffie et al. (2007) and Kealhofer (2003), who found significant dependence on the probabilities of future bankruptcy on this variable.

The variable distance to default was calculated based on the proposal of Bharath and Shumway (2008) that from a naïve model reached a higher result without the need for iterative operations between the variables of the volatility of the prices of the assets and the market value of the company. In this work, the equity volatility component \( \sigma \) was obtained by the quarterly standard deviation of the profitability of daily stock prices. The risk-free asset rate used in the DD variable was the quarterly average of the previous quarter’s (1-year treasury constant maturity rate), as suggested by Bharath and Shumway (2008).

The variable market value of equity also was based on the work of Bharath and Shumway (2008), which is expressed by the value of the asset price multiplied by the number of shares traded. In the case of this study, we used the quarterly average of stock prices of companies multiplied by the number of shares traded.
To identify specific characteristics among the sectors, five dummy variables were included, representing each of the six subgroups selected in the sample.

4.3 Database

The database used for the analysis was extracted from Bloomberg and was composed of accounting information of the balance sheet and profit and loss account of 1,520 companies and 24 variables, totaling a universe of 79,040 observations from the period of 2006 to 2018, quarterly. The initial sample information refers to North American companies belonging to six specific segments of the infrastructure sector, based on the Gics, namely: water and sanitation, electricity, renewable electric energy (ENR), logistics and transportation (LOG), oil and gas (PET), and telecommunications (TEL).

Gas utilities were excluded from the database, companies whose activities were diverse within these sectors. The justification for this exclusion was based on the low representativeness of these activities in infrastructure sectors, as well as the fact that there were companies with mixed information about their activity. The period analyzed was from the first quarter of 2006 to the fourth quarter of 2018, with a periodicity of 48 quarters.

The lack of temporal constancy in the data, with many information losses throughout the quarters, compromised the longitudinal analysis of the sample. Therefore, it was necessary to make some adjustments to the database, such as summarizing the information over time and applying the average of each of the variables of interest per company.

5. RESULTS AND DISCUSSION

The data presented 404 companies with missing values in at least one of the variables. Similarly, the water and sanitation, and electric energy sectors did not present any bankruptcy events throughout the period, being removed from the observations. In sum, the final analysis had a set of 1,066 companies.

Extreme values were identified in some variables and, to avoid discarding sampled data, as well as the incidence of outliers, the values of the variables
were truncated at the ninety-ninth and first percentiles, as suggested by Shumway (2001).

**Figure 5.1**

SAMPLE ANALYSIS

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNE</td>
<td>50</td>
<td>4.69%</td>
</tr>
<tr>
<td>LOG</td>
<td>98</td>
<td>9.19%</td>
</tr>
<tr>
<td>PET</td>
<td>842</td>
<td>78.99%</td>
</tr>
<tr>
<td>TEL</td>
<td>76</td>
<td>7.13%</td>
</tr>
<tr>
<td>Default</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>100</td>
<td>9.38%</td>
</tr>
<tr>
<td>No</td>
<td>966</td>
<td>90.62%</td>
</tr>
</tbody>
</table>

*Source: Elaborated by the author.*

It should be noted that the Water and Sanitation and the Energy sectors had 27 and 65 companies, respectively. In addition, the intention to separate Energy companies into two distinct sectors (energy and renewable energy) was in order to improve the detail of the analysis. However, within these sectors, there was no classification capable of allowing greater detail regarding the generation, transmission, distribution of energy, or its source.

**Figure 5.2**

DESCRIPTIVE VARIABLES

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>Average</th>
<th>SD</th>
<th>Min.</th>
<th>1° Q</th>
<th>2° Q</th>
<th>3° Q</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZ</td>
<td>1,066</td>
<td>1.86</td>
<td>1.55</td>
<td>-1.81</td>
<td>0.66</td>
<td>2.2</td>
<td>3.13</td>
<td>4.68</td>
</tr>
<tr>
<td>WK</td>
<td>1,066</td>
<td>-19.03</td>
<td>111.2</td>
<td>-934.39</td>
<td>-0.33</td>
<td>-0.01</td>
<td>0.11</td>
<td>0.58</td>
</tr>
<tr>
<td>RE</td>
<td>1,066</td>
<td>-163.92</td>
<td>910.56</td>
<td>-7571.15</td>
<td>-5.19</td>
<td>-0.46</td>
<td>0.0</td>
<td>0.66</td>
</tr>
<tr>
<td>EB</td>
<td>1,066</td>
<td>-1.96</td>
<td>10.4</td>
<td>-83.31</td>
<td>-0.12</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>SL</td>
<td>1,066</td>
<td>0.2</td>
<td>0.4</td>
<td>0</td>
<td>0.03</td>
<td>0.08</td>
<td>0.19</td>
<td>2.93</td>
</tr>
<tr>
<td>NI</td>
<td>1,066</td>
<td>-2.53</td>
<td>13.54</td>
<td>-109.04</td>
<td>-0.15</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.39</td>
</tr>
<tr>
<td>LA</td>
<td>1,066</td>
<td>20.79</td>
<td>113.75</td>
<td>0.07</td>
<td>0.4</td>
<td>0.59</td>
<td>1.02</td>
<td>938.85</td>
</tr>
<tr>
<td>LI</td>
<td>1,066</td>
<td>3.83</td>
<td>8.16</td>
<td>0.04</td>
<td>0.85</td>
<td>1.51</td>
<td>3.01</td>
<td>57.29</td>
</tr>
</tbody>
</table>

(continue)
Figure 5.3 shows how the default is distributed across sectors. It was verified that the Petroleum sector presented the highest percentage of bankruptcy (10.1%), while the Renewable Energy sector presented the lowest (4%). However, there was no significant difference (p-value = 0.460) between the percentages of bankruptcy by Fisher’s exact test.

In the comparative analysis between the companies that failed and did not fail, the highlights are the variables: SZ, EB, LA, and EQ.

In Figure 5.3, it is possible to verify how the quantitative variables are presented with respect to bankruptcy, as well as the p-value of the non-parametric Mann-Whitney test.

It is verified that at least 50% of the failed companies presented the variable SZ with values below or equal to 2.63, while among the companies that did not fail, at least 50% presented this variable with numbers below or equal to 1.93.
In addition, for the variable EB, it was verified that at least 50% of the companies that failed presented values below or equal to -0.02, while among companies that did not fail, 50% presented this variable with numbers below or equal to -0.01.

The variable LA also showed significance when the samples were compared. At least 50% of the companies that failed presented the variable LA with values below or equal to 0.73, while in the companies that did not fail, at least 50% presented the variable LA with values below or equal to 0.58.

Finally, at least 50% of the companies that failed have had the variable EQ with values below or equal to 2.87, while in the companies that did not fail, at least 50% showed this variable with numbers below or equal to 6.43.

The other predictors did not present a significant difference between the companies that failed and did not fail. The correlation matrix presented in Figure 5.4 highlights the existence of multicollinearity among several variables. It was found variables a with high positive correlation, as for example, WK and RE, besides NI and EB, and variables with high negative correlation, as for example, LA and WK.

The model initially proposed, expressed in Figure 5.5, underscores the evidence that several predictor variables were not significant to explain bankruptcy (p-value ≥ 0.05).

In addition, the variance inflation factor (VIF) result, greater than 10, underscored the multicollinearity problem among the model variables.

Regarding the odds ratio (OR) results, it was verified that the variable SZ was significant, which means that the increase in one unit of this variable while maintaining the other constant variables increases the chance of failure by 41.3%.

The coefficient of the variable DD was multiplied by 100 to facilitate interpretation. The results indicate that the increase in 100 units, also kept constant, decreases the chance of bankruptcy by 6.6%.
The bankruptcy risk in infrastructure sectors: An analysis from 2006 to 2018

Figure 5.4
MATRIX CORRELATION

<table>
<thead>
<tr>
<th></th>
<th>SZ</th>
<th>WK</th>
<th>RE</th>
<th>EB</th>
<th>SL</th>
<th>NI</th>
<th>LA</th>
<th>LI</th>
<th>CF</th>
<th>DD</th>
<th>EQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZ</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WK</td>
<td>0.31</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>0.32</td>
<td>0.90</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EB</td>
<td>0.34</td>
<td>0.74</td>
<td>0.73</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SL</td>
<td>-0.08</td>
<td>-0.16</td>
<td>-0.19</td>
<td>-0.29</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NI</td>
<td>0.34</td>
<td>0.80</td>
<td>0.77</td>
<td>0.93</td>
<td>-0.23</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LA</td>
<td>-0.32</td>
<td>-0.99</td>
<td>-0.91</td>
<td>-0.76</td>
<td>0.16</td>
<td>0.82</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LI</td>
<td>-0.14</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.06</td>
<td>-0.11</td>
<td>0.06</td>
<td>-0.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td>0.35</td>
<td>0.50</td>
<td>0.56</td>
<td>0.60</td>
<td>-0.14</td>
<td>0.58</td>
<td>-0.52</td>
<td>-0.02</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>-0.16</td>
<td>-0.23</td>
<td>-0.20</td>
<td>-0.20</td>
<td>0.07</td>
<td>-0.20</td>
<td>0.22</td>
<td>0.08</td>
<td>-0.12</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>EQ</td>
<td>-0.17</td>
<td>-0.09</td>
<td>-0.08</td>
<td>-0.13</td>
<td>0.00</td>
<td>-0.10</td>
<td>0.10</td>
<td>0.00</td>
<td>-0.23</td>
<td>0.07</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Source: Elaborated by the author.

Figure 5.5
DESCRIPTIVE VARIABLES

<table>
<thead>
<tr>
<th>Variable</th>
<th>Initial model</th>
<th>β</th>
<th>IC</th>
<th>OR</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td></td>
<td>-2.405</td>
<td>[-2.92; -1.92]</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>SZ</td>
<td></td>
<td>0.346</td>
<td>[0.16; 0.54]</td>
<td>1.413</td>
<td>0.0</td>
</tr>
<tr>
<td>Sector = PET</td>
<td></td>
<td>-1.005</td>
<td>[-2.85; 0.23]</td>
<td>0.366</td>
<td>0.178</td>
</tr>
<tr>
<td>Sector = ENR</td>
<td></td>
<td>-0.429</td>
<td>[-1.28; 0.3]</td>
<td>0.651</td>
<td>0.281</td>
</tr>
<tr>
<td>Sector = LOG</td>
<td></td>
<td>-0.559</td>
<td>[-1.66; 0.32]</td>
<td>0.572</td>
<td>0.259</td>
</tr>
<tr>
<td>Sector = TEL</td>
<td></td>
<td>0.006</td>
<td>[-0.01; 0.02]</td>
<td>1.006</td>
<td>0.4</td>
</tr>
<tr>
<td>WK</td>
<td></td>
<td>0</td>
<td>[0.00; 0.00]</td>
<td>1</td>
<td>0.547</td>
</tr>
<tr>
<td>RE</td>
<td></td>
<td>-0.066</td>
<td>[-0.19; 0.00]</td>
<td>0.936</td>
<td>0.109</td>
</tr>
<tr>
<td>EB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continue)
From the results obtained, the Backward method was applied in order to remove the variables with the multicollinearity problem. Only the variables SZ, EB, and DD, remained as significant.

The p-value of the Hosmer-Lemeshow Test presented a value greater than 0.05 in both cases, demonstrating that both models are adequate.

The remaining fixed all the other variables; it was verified the increase in one unit of the variable SZ. Thus, the chance of default also increases by 46.8% in the adjusted model.

In addition, after adjusting the model, the EB variable became significant, indicating that its increase in one unit, while remaining the other variables fixed, makes the chance of bankruptcy increase by 2.3%.

The adjusted model also showed that the increase of 100 units of the variable DD, with the other variables remaining fixed, makes the chance of bankruptcy decrease 6.5%.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Initial model</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>IC</td>
<td>OR</td>
<td>p-value</td>
</tr>
<tr>
<td>SL</td>
<td>0.188</td>
<td>[-0.44; 0.72]</td>
<td>1.207</td>
<td>0.52</td>
</tr>
<tr>
<td>NI</td>
<td>0.043</td>
<td>[-0.01; 0.15]</td>
<td>1.043</td>
<td>0.244</td>
</tr>
<tr>
<td>LA</td>
<td>0.005</td>
<td>[-0.01; 0.02]</td>
<td>1.005</td>
<td>0.513</td>
</tr>
<tr>
<td>LI</td>
<td>-0.047</td>
<td>[-0.12; 0.00]</td>
<td>0.954</td>
<td>0.111</td>
</tr>
<tr>
<td>CF</td>
<td>0.01</td>
<td>[-0.04; 0.06]</td>
<td>1.01</td>
<td>0.686</td>
</tr>
<tr>
<td>DD/100</td>
<td>-0.069</td>
<td>[-0.12; -0.02]</td>
<td>0.934</td>
<td>0.007</td>
</tr>
<tr>
<td>EQ</td>
<td>0</td>
<td>[0.00; 0.00]</td>
<td>1</td>
<td>0.445</td>
</tr>
</tbody>
</table>

Source: Elaborated by the author.
Finally, after the identification of the second model, it was verified the differences between PET and non-PET sectors. For this, the final model was adjusted for each of the sectors, comparing the point and interval estimates of the regression coefficients.
Figure 5.7

**OIL X NON-OIL COMPANIES**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Oil companies</th>
<th>Non-oil companies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>IC</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-2.729</td>
<td>[-3.22; -2.29]</td>
</tr>
<tr>
<td>SZ</td>
<td>0.403</td>
<td>[0.20; 0.62]</td>
</tr>
<tr>
<td>EB</td>
<td>-0.022</td>
<td>[-0.04; 0.00]</td>
</tr>
<tr>
<td>DD/100</td>
<td>-0.054</td>
<td>[-0.11; -0.01]</td>
</tr>
<tr>
<td>Hosmer-Lemeshow</td>
<td>0.299</td>
<td></td>
</tr>
<tr>
<td>Bigger VIF</td>
<td>1.688</td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>0.667</td>
<td></td>
</tr>
</tbody>
</table>

*Source: Elaborated by the author.*

The small number of companies in each sector made the specific analysis of each one unfeasible from a statistical point of view. However, the separation between the two groups (oil companies and non-oil companies) made it possible to obtain more secure and clear information.

6. CONCLUSION

This study aimed to identify the existence of sectorial differences in the prediction of default risk of American infrastructure companies based on logistic regression with a binary dependent variable.

We verified that the sectorial separation for the estimation of default probability might contribute to the identification of specific causes of this probability that are linked to sectorial idiosyncrasies. The variable distance to default showed that it has good applicability for sectorial analysis.

The study showed that it is possible to explain the default probability of each sector separately. Despite the limitations of this study, especially in terms of the small number of default events by sector, it can contribute to the creation of new research that takes into account sector specificities when calculating default risk, such as events related to seasonality, for example.

Another way to explore these results is to use the DD variable as a dependent variable in modeling the probability of default. It was identified that this variable has different sensitivities according to each sector.
Future studies could include the existence of more significant variables since such adjustment could make possible the increase of interest events. Finally, investigating the default probability of companies from different infrastructure sectors can contribute to the creation of specific mechanisms of corporate risk analysis in a more detailed manner, avoiding that poor performance in certain indicators penalizes companies in sectors that are not sensitive to these indicators.

RISCO DE FALÊNCIA EM SETORES DE INFRAESTRUTURA:
UMA ANÁLISE DE 2006 A 2018

RESUMO

Objetivo: Esta pesquisa tem como objetivo identificar a probabilidade de inadimplência de empresas de infraestrutura, considerando as especificidades setoriais de suas atividades. Além disso, o trabalho procura identificar a aplicação de variáveis estruturais de probabilidade de inadimplência em um modelo de forma reduzida, a fim de identificar a significância de seu uso. Para isso, investigamos 1.520 empresas norte-americanas de seis setores diferentes vinculados à infraestrutura.

Originalidade/valor: As análises realizadas para identificar a probabilidade de uma empresa falir dificilmente consideram sua particularidade setorial. Como resultado, embora a maioria dos modelos traga insumos importantes para a avaliação de riscos, a maioria deles não considera essa visão setorial. Nesse sentido, este trabalho tem como valor e originalidade contribuir para o preenchimento dessa lacuna e identificar a existência de diferenças setoriais na análise de risco de inadimplência em empresas de infraestrutura no mercado norte-americano, no período entre 2006 e 2018.

Design/metodologia/abordagem: O estudo realizou uma regressão logística (modelo logit) usando 11 variáveis do modelo estabelecidas no cálculo da probabilidade de inadimplência. O estudo também utilizou a variável distância até o default como variável explicativa, a fim de identificar sua capacidade de explicar o fenômeno pesquisado.

Resultados: O estudo identificou que, além do tamanho das empresas, a distância até o default é a única variável que pode ser aplicada com significância em todos os setores analisados. Além disso, constatou-se que as
empresas do setor de petróleo e gás têm menos sensibilidade a essa variável do que as empresas dos outros setores.

**PALAVRAS-CHAVE**


**REFERENCES**


Vanderson A. Delapedra-Silva, master from the Faculty of Science and Letters, Paulista State University “Júlio de Mesquita Filho” (Unesp).
Vanderson A. Delapedra-Silva is now a Ph.D. student in finance at the Faculty of Administration, Accounting, Economics and Public Management of University of Brasília (UnB), and advisor in Bank of Brazil.
Correspondence concerning this article should be addressed to Vanderson A. Delapedra-Silva, Saun, s/n, Quadra 5, Lote B, Edifício Banco do Brasil, Asa Norte, Brasília, Distrito Federal, Brazil, CEP 70040-912.
E-mail: vanderson.economia@gmail.com