▪ Advances in Scientific and Applied Accounting, ▪ ISSN 1983-8611 ▪ São Paulo ▪ v.14, n.2 ▪ p. 042-053 ▪ May / Aug. de 2021.
▪ DOI: 10.14392/asaa.2021140202

ASAA

Determinantes Hierárquicos da Inadimplência de Financiamento Imobiliário de Pessoa Física

Renata Barreto Lima[1], Ricardo Goulart Serra[2], Luiz Paulo Lopes Fávero[3]

[1]Banco do Brasil. São Paulo-SP. Brazil.
[2]FECAP. São Paulo-SP. Brazil.
[3]Universidade de São Paulo. São Paulo-SP. Brazil.

[1]renatabalima@gmail.com
[2]ricardo.serra@fecap.br
[3]lpfavero@usp.br

Edited by:
Orleans Silva Martins
Paulo Roberto da Cunha

Abstract

Objectives: The objectives of the article are: (i) to identify the determinants of default of real estate financing and (ii) to verify the existence of influence of the individual context (bank branch) on default.
Method: The article studied 5,113 real estate financing contracts of individuals belonging to 1,448 branches of a national financial institution in force on September 30, 2019. Due to the multilevel character of the sample (level 1: individual and level 2: bank branch) and considering that the different bank branches present different levels of performance (in different indicators, including the percentage of defaulted real estate financing contracts), the multilevel logistic model was used to the detriment of the traditional logistic model.
Results: The multilevel logistic model was superior to the traditional logistic model (18.8% of the default probability variability refers to the bank branchlevel). The negatively significant characteristics of the individual are: age, length of relationship, receiving salary at the bank and higher education level, and the positively significant characteristics are: level of education up to elementary school and the financing/income ratio. The variables gender and marital status were not significant.
Contributions: The innovations of the article are: (a) the national coverage of the sample on default by individuals, (b) the use of two unprecedented variables and (c) the use of the multilevel logistic model. The identification that individuals with the same characteristics, however coming from different bank branches, have different probabilities of defaulting leads to the knowledge of academics and professionals in the area the recommendation of hierarchical modeling for the analysis of real estate credit.

Keywords: real estate financing; default; individuals; multilevel model.

How to cite:

Lima, R. B., Serra, R. G., & Fávero, L. P. L. DETERMINANTES HIERÁRQUICOS DA INADIMPLÊNCIA DE FINANCIAMENTO IMOBILIÁRIO DE PESSOA FÍSICA. Advances in Scientific and Applied Accounting, 14(2), 029–041/042. https://doi.org/10.14392/asaa.2021140202

Received: May 09, 2021
Submitted to new round: October 07, 2021
Accepted: October 18, 2021

Introduction

The Brazilian housing problem dates back to the beginning of the 20th century, with the exodus of the rural population and small towns to large industrial centers, in search of jobs generated by manufacturing-industrial development (Pinto, 2015). Numerous government initiatives were created to address the problem: construction of housing complexes in the 1930s; Casa Popular Foundation (FCP – Fundação Casa Popular) in the 40s; National Housing Bank (BNH – Banco Nacional de Habitação) in 1964; Salary Variation Compensation Fund (FCVS – Fundo de Compensação de Variações Salariais) in 1967; Wage Equalization Coefficient (CES – Coeficiente de Equiparação Salarial) in 1969; Real Estate Financing System (SFI – Sistema de Financiamento Imobiliário) in 1997; Growth Acceleration Program (PAC – Programa de Aceleração do Crescimento) in 2007 and Minha Casa Minha Vida Program (PMCMV – Programa Minha Casa Minha Vida) in 2009. Despite numerous government initiatives, the housing deficit in Brazil was estimated at 6.355 million households, corresponding to 9.3% of the stock of private households (Fundação João Pinheiro, 2018).

Considering the high property values and the "excessive onus of urban rent" (equivalent to an expenditure greater than 30% of family income, according to the João Pinheiro Foundation [2018]), the alternative is the acquisition through real estate financing. It's interesting for financial institutions to finance the acquisition of real estate as long as the borrowers remain in good standing, despite the warranty given by the financed property (it is not in the interest of financial institutions to bear the costs related to the repossession of the warranty and its subsequent sale).

The balance of the loan portfolio with resources directed to individuals through real estate financing at regulated rates was, in November 2019, R$ 569,593, a significant increase compared to the amount of R$ 79,094 in January 2010 (Central Bank of Brazil, 2019a). Default on specifically earmarked financing was 1.6% in November 2019, less than 2.3% in March 2011, beginning of the historical series (Central Bank of Brazil, 2019b).

The identification of the determinant factors of real estate financing default would allow for the reduction of the risk of these operations and, consequently, the possibility of operating with lower interest rates and of offering a greater volume of credit, benefiting financial institutions, the economy as a whole and countless families. It is noted that the best way to allocate credit to consumers is a critical issue for companies in general, in particular, financial institutions. Despite the breadth of interests on the subject (financial, economic and social), few academic works address it, it is believed, due to the difficulty of obtaining a database.

This paper analyzes real estate financing contracts within the scope of the Minha Casa Minha Vida Program, of a large Brazilian bank with national coverage. There are 5,113 customer contracts from 1,448 bank branches effective on September 30, 2019.

Existing theories, proposed within the context of a model with a single equation, consider that there is no connection between individuals and the society in which they live, as noted by Courgeau (2003). Analyses that ignore this connection, according to the author, describe the behavior of individuals in a wrong way; and the correct analysis of the phenomena can only be carried out from the recognition of these connections. According to Khudnitskaya (2010), the nested data structure or hierarchical structure is typical in social sciences.

It is noted (1) that contracts are awarded to individuals (first level of analysis) belonging to a context, their bank branch (second level of analysis). Considering that (2) the evaluations of results, carried out by bank branches, find that individual bank branches present different results (for example, percentage of non-performing real estate contracts), it is reasonable to assume that the context of individuals can influence the determinants of default, as pointed out by Courgeau (2003).

The use of level equations allows the researcher to transcend single-equation theories. It is concluded, therefore, that it is necessary to use a tool capable of capturing (1) the two levels of analysis and (2) the influence of the bank branch on default, in this case, a multilevel logistic model. Such modeling has the benefit of disaggregating the variance of the dependent variable at the different levels of analysis, in order to clarify the influence of the different levels of analysis. Therefore, in the context of this study, such modeling allows this paper to address two questions: which characteristics of individuals are related to the probability of default and is the connection between the individual and his group (bank branch) relevant in the analysis of default?

Therefore, the objectives of this work are: (i) to identify the determinants of default of real estate financing and (ii) to verify the existence of influence of the individual context (bank branch) on default; that is, whether individuals with the same characteristics, but from different bank branches, have (or not) different probabilities of defaulting.

The article is pioneer in three aspects: (i) in the study of real estate financing with national coverage (there is no knowledge of national articles that have had access to a national database), (ii) in the definition of two variables not used in other studies (income-weighted financing, that is, financing/income and salaries, that is, receiving the salary at the bank itself) and (ii) the use of a multilevel tool in credit analysis (in general).

In addition to this introductory section, the article has: section 2 with the theoretical framework; section 3 with the methodology; data modeling and analysis (section 4); and the last section, with the final considerations.

2 Theoretical Reference

2.1 Credit

For an institution that has financial intermediation as its main activity, credit consists of making available to the client (borrower) a certain amount in the form of a loan or financing, upon a promise of payment at a future date (Silva, 2016). For Securato (2007, p.17), the term "credit" determines a "relation of trust between two (or more) parties in a given operation". Thus, in view of this trust relationship, it can be said that a credit operation involves the expectation of receiving an amount by one of the parties within a certain period of time (Brito & Assaf Neto, 2008).

2.2 Credit Risk

Credit risk is the probability that the return on capital is not as expected (Caouette, Altman, Narayanan, & Nimmo, 2009). The measurement of credit risk is the process by which the financial institution quantifies the probability of loss, if the payments of credit operations are not confirmed (Brito & Assaf Neto, 2008). It can be evaluated individually (by contract) or collectively (contract portfolio).

The credit risk analysis of individuals involves the observation of the so-called Cs of credit (character, capacity, capital, collateral and conditions), which can be done either by judgmental criteria or by statistical processes (Silva, 2016). The judgmental analysis has subjective criteria, being subordinated to the experience of the credit analyst or manager. The statistical process makes it possible to expand the range of analyses, reducing their cost. For Securato (2007), Cs of credit for individuals can be obtained by – character: based on the creditor's own registration information or that of third parties (such as Serasa and SCPC); capacity: directly related to the individual's income; capital: the applicant's personal assets; collateral: guarantees that the applicant makes available to the creditor; and conditions: macro and microeconomic factors that influence credit granting.

2.3 Credit Scoring

The most used means of risk control is the scoring system (Guimarães & Chaves Neto, 2002). Credit scoring, one of the main statistical methods, can be defined, therefore, as the process of attributing points to decision variables using statistical techniques (Amorim Neto & Carmona, 2004). Therefore, a value is assigned to each characteristic of a customer, a score is drawn up and compared with a cutoff point, deciding whether or not to grant credit. The cut-off point is established in order to find the number of bad debtors probabilistically accepted that would cause less damage to the profit generated by good debtors (Crespi Júnior, Pereira, & Kerr, 2017). These parameters used for granting credit to individuals are based on the aforementioned Cs of Credit.

Classification models are generally developed with the following statistical techniques: multiple linear regression, discriminant analysis, linear programming, genetic algorithm, survival analysis, decision tree, neural networks and logistic regression, in order to try to identify the determining factors of default (Locatelli, Ramalho, Silvério, & Afonso, 2015).

2.4 Empirical Works

Most of the studies are aimed at analyzing companies through their economic and financial indicators. Outside Brazil, the studies carried out by FitzPatrick (1932), Smith and Winakor (1935), Merwin (1942), Beaver (1966), Tamari (1966), Altman (1968), Backer and Gosman (1978) and Topa (1979) stand out, with studies that extend to the present (Yildririm, 2020). In Brazil, the most important studies (Silva, 2016) are: Elizabetsky (1976), Kanitz (1978), Matias (1978) and Silva (1982).

However, there are few works on the credit risk analysis for individuals. The Brazilian studies and the variables used in them are summarized in Table 1. The difficulty in accessing data by researchers should also be considered (eg, Amorim Neto and Carmona (2004) did not obtain access to income data).

Table 1. Variáveis utilizadas em estudos de análise de crédito de pessoas físicas

Independent variables	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	Count
Age	x	x	x	x	x	x	x	x	x	x	x	x	x	13
Income	x	x	x		x	x	x	x	x	x	x	x	x	12
Marital status	x	x			x	x	x	x		x	x	x	x	10
Education level	x	x	x	x	x			x		x	x	x		9
Gender	x	x				x	x	x		x	x	x	x	9
Banking relationship time	x	x	x		x	x				x				6
Financed amount		x				x	x		x		x		x	6
Home address/ zip code			x	x			x	x	x					5
Number of installments / term			x			x			x		x		x	5
Value of installments						x	x		x		x		x	5
Type of occupation	x			x			x			x				4
Loan type						x	x				x		x	4
Restriction (internal/ external)	x			x								x		3
Payment capacity		x								x	x			3
Residence time	x						x	x						3
Type of residence								x			x	x		3
Number of dependents											x	x		2
Time in current occupation							x	x						2

Note: I. Albuquerque, Medina and Silva (2017); II. Amorim Neto and Carmona (2004); III. Ferreira, Celso and Barbosa Neto (2012); IV. Ferreira, Oliveira, Santos and Abrantes (2011); V. Gouvêa, Gonçalves and Mantovani (2013); SAW. Guimarães and Chaves Neto (2002); VII. Jannuzzi (2010); VIII. Locatelli et al. (2015); IX. Lopes, Ciribeli, Massardi and Mendes (2017); X. Maciel and Maciel (2017); XI. Ritta, Gorla and Hein (2015); XII. Sousa, Petri and Anjo (2018); XIII. Wedge (2021).
Source: elaborated by the authors.

It can be seen in Table 1 that the five parameters most used in credit analyzes for individuals are related to the borrower's socioeconomic data (age, income, marital status, level of education and gender).

Of the researched works, only the ones by Locatelli et al. (2015) and Cunha (2021) refer to real estate credit, however, they are regional studies (without national coverage). Januzzi (2010), in turn, studies operations aimed at the renovation and expansion of properties (not their acquisition). Ritta, Gorla and Hein (2015) analyze contracts for targeted productive microcredit. Guimarães and Chaves Neto (2002) analyze credit card transactions. The other works study loans and financing in general (CDC).

Some studies are not limited to researching the determining factors of customer default, testing the same data through different statistical models, such as the works by: Amorim Neto and Carmona (2004), comparing logistic regression to discriminant analysis, concluding for the similarity of the models; Gouvêa, Gonçalves and Mantovani (2013), testing logistic regression and neural models, and Guimarães and Chaves Neto (2002), evaluating Fisher's linear discriminant function and logistic regression, being the two latter studies show equivalence among the studied models with slight advantage to the logistic regression model.

In Brazil, Albuquerque, Medina and Silva (2017) analyzed Direct Consumer Credit (CDC) and Cunha (2021) analyzed real estate financing, both using a geographically weighted logistic regression model, recording and not recording an improvement in the model compared to the regression traditional logistics, respectively. However, the works did not use a multilevel technique. Abroad, Khudnitskaya (2010) developed, based on credit card operations, a credit scorecard based on the multilevel model, with level 1: individual and level 2: microenvironment (based on common characteristics related to economic and demographic conditions in areas of residence, but not necessarily geographical). The other studies work with models with a single level.

Recently, with the increase in online credit application as a result of the pandemic, international studies have identified that financial institutions have started to use, for consumer credit analysis, information that goes beyond traditional sources of information (eg.: financial information). This information is obtained, among other sources, from mobile application data (Jiang, Liao, Xi, Wang, & Xiang, 2021). Zhou, Wang, Ren, & Chen (2021) used data on individuals' behavior (eg, telephone use).

3 Methodology

3.1 Samples and Variables

The universe is composed of 255,926 real estate financing contracts for individuals under the Minha Casa Minha Vida program of a financial institution with national coverage, with operations in force on September 30, 2019. The authors did not have access to the universe, but only one sample drawn randomly from 2% of contracts of the universe by State and Federal District, resulting in 5,125 observations distributed in 1,448 bank branches. Customers were identified by codes (from 1 to 5,125). Both procedures preserved the confidentiality of individuals (this is confidential data). Table 2 presents the variables provided by the financial institution, in addition to the client's bank branch (also coded from 1 to 1,448) and the state of origin. The two first variables in Table 2 (default and delayed) were used in two distinct models (main model – section 4.2 and sensitivity analysis – section 4.5, respectively) as dependent variables. Other variables were used as explanatory variables (first level).

Table 2. Description of dependent and independent variables

Variable	Specification
Default	- No delay or delay of less than 180 days (dummy = 0) - Delay greater than 180 days (dummy = 1)
Delayed	- No delay (dummy = 0) - Delay from 01 day (dummy = 1)
Financed Value	Financing amount, at the time of contracting, in Reais (BRL)
Income	Monthly income amount, in Reais (BRL)
Age	Account holder age, in years
Gender	- Male (dummy = 0) - Female (dummy = 1)
Marital Status	- Single (dummy = 0) - Others[a] (dummy = 1)
Education levelb - Elementary	- Illiterate and elementary school (dummy = 1) - Others (dummy = 0)
Education level - Higher	- Higherr[c] (dummy = 1) - Others (dummy = 0)
Relationship time	Customer relationship time with the bank, in years
Receives salary at the bank	- Does not receive salary at the bank (dummy = 0) - Receives salary at the bank (dummy = 1)

Notes: a. married, separated, divorced or widowed; b. three categories of education level were analyzed, therefore, requiring two dummy variables (Elementary and Higher), with the basic level of education being secondary education. c. incomplete higher education, complete higher education and postgraduate (lacto and stricto sensu).
Source: elaborated by the authors.

The sample was analyzed to check for possible absences and/or divergences. The following were excluded: nine observations due to lack of relationship time; two observations with monthly income of “R$ 0.01” and one observation with information that differs from the codes provided for the fields “Civil Status” and “Education Level”, being impossible to obtain the correct code. Thus, the analyzed sample contains 5,113 valid observations.

Assuming that the financed amount and the income individually are absolute variables, the variable "financing/ income" was created (calculated as the natural logarithm of the financed amount divided by the monthly income), in order to relativize the amount financed by the individual's income. It is noteworthy that the financed amount represents data from the past (at the time of granting) and the individual's income represents updated data (considering the need to update the customer record). It is understood that the variable created (as an innovation in this study) constitutes a better indicator of the individual's ability to pay than the absolute variables: financed amount and income, when individually used (common practice in studies in the area).

3.2 Research Hypotheses

In order to: (i) identify the determinants of default of real estate financing and (ii) verify the existence of influence of the individual context (bank branch) on default, the research hypotheses are:

Hypothesis I: there are characteristics of individuals that explain the variability in the probability of default of individuals from the same bank branch (H-I, characteristics of individuals explain differences between individuals).

Hypothesis II: there is significant variability in the probability of default of individuals from different bank branches (H-II, variability between bank branches).

The alternative hypotheses to H-I (H-Ia) and H-II (H-IIa) are their negatives.

In this sense, in addition to identifying the characteristics of individuals relevant to the analysis of credit granting, there is an interest in analyzing whether individuals with the same characteristics, but from different bank branches, have (or not) different probabilities of defaulting. Considering the research hypotheses described, the model to be used is the multilevel binary logistic model.

3.3 The Multilevel Binary Logistic Model

Logistic regression is widely used for credit risk analysis modeling, considering the probability estimation characteristic of prior classification of customers as non-defaulters and defaulters (Ritta, Gorla, & Hein, 2015). It is the technique that best provides the predictability of success in comparison to other statistical techniques, according to analyzes carried out by Guimarães and Chaves Neto (2002).

Binary logistic models differ from traditional models estimated by Ordinary Least Squares (OLS) as they focus on the analysis of probabilities of a given event Y that presents itself in a discrete and qualitative way, so that a dummy variable can be defined to characterize it (Y = 1 to characterize the occurrence of the event, and 0 otherwise).

Thus, in the logistic model, the probability of occurrence of an event is given, according to Fávero and Belfiore (2017), by:

$$\rho_i = 1/1+e^{zi} \quad (1)$$

where p is the probability of occurrence of the event and Z is the logit, defined by:

$$Z_i = \alpha + \beta 1.X_{1i} + \beta_2.X_{2i} + \ldots + \beta_k.X_{ki} \quad (2)$$

where a is the constant, β_j (j = 1, 2,..., k) are the estimated parameters for explanatory variables, X_{ji} are the j explanatory variables, for each i observation of the sample. The logit parameters are estimated by maximum likelihood, from the maximization of the natural logarithm of the likelihood function, defined by:

$$LL = \sum_{i=1}^{n} \left\{ \left[(Y_i).ln\left(\frac{e^{Z_i}}{1+e^{Z_i}}\right) \right] + \left[(1-Y_i).ln\left(\frac{1}{1+e^{Z_i}}\right) \right] \right\} \tag{3}$$

In general, a binary logistic model with two levels can be defined, in which the first level offers the explanatory variables $X_1, \ldots, X_Q$ for each individual i (i = 1, ..., n), and the second level, the explanatory variables $W_1, \ldots, W_S$ referentes a cada grupo ou contexto j (j = 1, ..., J), for each group or context j (j = 1, ..., J), invariant for observations belonging to the same group:

Level 1

$$p_{ij} = \frac{1}{1+e^{-(b_{0j} + b_{1j}.X_{1ij} + b_{2j}.X_{2ij} + \ldots + b_{Qj}.X_{Qij})}} \tag{4}$$

where p_{ij} represents the probability of occurrence of the event of interest for each i observation belonging to a certain group j and b_{qj} (q = 0, 1, ..., Q) refer to the level 1 coefficients (Raudenbush & Bryk, 2002; Rabe-Hesketh & Skrondal, 2012).

Level 2

$$b_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs}.W_{sj} + u_{qj} \tag{5}$$

where γ_{qs} (s = 0, 1, ..., S_q) refer to level 2 coefficients and u_{qj} are the random level 2 effects, normally distributed, with zero mean and variance τ_{qq}. Besides, any error terms regardless of u_{qj} have zero mean and variance $\pi^2/3$.

According to Fávero and Belfiore (2017), the main advantage of multilevel models over traditional regression models (GLM - Generalized Linear Models) is the possibility of considering the natural nesting of data, thus allowing individual heterogeneities to be identified and analyzed and among groups to which these individuals belong, making it possible to specify random components at each level of analysis.

In other words, multilevel models correct for the fact that observations in the same group are not independent and therefore, compared to traditional models, lead to unbiased estimates of standard errors of parameters. As Steenbergen and Jones (2002), Arceneaux and Nickerson (2009), and Hair and Fávero (2019) emphasize, if researchers are interested in testing whether group-level covariates moderate individual-level effects, multilevel models seem to be the most appropriate choice. And a likelihood ratio test can be designed to verify the adequacy of the multilevel estimation to the data structure, as well as providing a comparison with estimates from, for example, traditional GLM models.

According to Courgeau (2003), within a model structure with a single equation, there seems to be no connection between individuals and the society in which they live. In this sense, the use of level equations allows the researcher to "jump" from one science to another: students and schools, families and neighborhoods, companies and countries. Ignoring this relationship means elaborating incorrect analyses about the behavior of individuals and, equally, about the behavior of groups. Only the recognition of these reciprocal influences allows the correct analysis of the phenomena.

This is in line with what Mathieu and Chen (2011) call a multilevel paradigm, which refers to a way of thinking: considering management phenomena in context and looking for variables that drive not only the focal unit of analysis, but also the levels above and below. This approach usually implies the development of multidisciplinary theories and investigations, which is the spirit articulated by Hitt, Beamish, Jackson and Mathieu (2007) when discussing the construction of theoretical and empirical bridges through contexts in multilevel modeling. And this is the objective of the present work, which seeks to relate different levels, such as individual attributes and contextual conditions of bank branches, on the probability of default on real estate financing.

4 Modeling and Data Analysis

4.1 Descriptive Statistics

The analysis considers the bank branch as the second level. However, for reasons of confidentiality of the individuals and as described in 3.1, the authors had access to a random sample of 2% of contracts, taken from the universe by State and Federal District, with the identification of individuals and bank branches by an index (from 1 to 5,125 for individuals and from 1 to 1,448 for bank branches). Therefore, one can only identify the contracts analyzed by State and Federal District, as shown in Table 3.

Table 3. Distribution of Contracts by State and Federal District

State/Federal district	%	State/Federal district	%	State/Federal district	%
AC	0,04	MA	1,15	RJ	0,76
AL	1,23	MG	6,59	RN	1,62
AM	0,06	MS	1,56	RO	0,59
AP	0,04	MT	1,17	RR	0,04
BA	1,47	PA	0,96	RS	1,66
CE	2,50	PB	2,03	SC	1,76
DF	51,75	PE	2,58	SE	0,76
ES	1,17	PI	0,68	SP	10,21
GO	2,97	PR	4,19	TO	0,43

Source: elaborated by the authors.

Table 4 presents the descriptive statistics of the discrete variables (all first level, since they are linked to individuals).

Table 4. Descriptive statistics of discrete variables

Variável	Especificação	Frequência	(%)
Default	- no delay, or less than 180 days delay - delay greater than 180 days	4.954 159	96,89 3,11
Delayed	- without delay - delayed (from 01 day)	3.747 1.366	73,28 26,72
Gender	- male - female	2.724 2.389	53,28 46,72
Marital status[a]	- single - married - widowed - separated - divorced	3.625 1.020 44 52 372	70,90 19,95 0,86 1,02 7,28
Education level[b]	- illiterate - elementary School - high school - incomplete higher - Graduated - postgraduate - master's degree - doctorate degree	2 483 3.396 374 830 23 2 3	0,04 9,45 66,42 7,31 16,23 0,45 0,04 0,06
Receives salary at the bank	- no - yes	4.195 918	82,05 17,95

Notes: a. represented by a dummy variable where 0 = single and 1 = married, widowed, separated and divorced; b. represented by two dummy variables (because there are three categories, having secondary education as the basic level of education), one dummy for illiterate and elementary education and the other dummy for incomplete higher education, complete higher education and postgraduate degrees (lacto and stricto sensu).
Source: elaborated by the authors.

It can be seen in Table 4 that 3.11% of the contracts are in default (delays greater than 180 days), while 26.72% are late (delays greater than 01 day). The sample is composed of men (53.28%) and women (46.72%) in a similar percentage, with the vast majority being single (70.90%), with high school (66.42%) and people who don't receive salaries at the bank (82) 0.05%).

Table 5 presents descriptive statistics for continuous variables.

Table 5. Descriptive statistics of continuous variables

Variable	Average	Median	Default deviation	Minimum	Maximum
Age	35,10	33,00	8,98	19	79
Relationship time	7,64	6,00	4,92	0	33
Financing	91.052,32	91.040,00	19.927,84	18.665,84	198.842,70
Income	2.257,79	1.800,00	1.552,95	400,00	22.616,33
Financing/ Income (ln)	3,82	3,88	0,51	0,00	5,53

Source: elaborated by the authors.

Table 5 indicates that the individual is, on average, 35.1 years old, has 7.6 years of relationship with the bank, average monthly salary of R$2.3 thousand and average financing of R$91.0 thousand. The average of the natural logarithm of the financing divided by the monthly income is 3.82 (equivalent to a financing 45.77 times greater than the monthly income).

Individuals are distributed in 1,448 bank branches, some of which are highly concentrated (the three bank branches with the highest number of observations represent 3.93%; 3.03% and 2.97% of the total, compared to the expected value below 0.1%, equivalent to 1/1,448). Such concentration reinforces the need to use a multilevel model, since, otherwise, it can inflate possible bank branch influences (and its characteristics) in the analysis, by attributing them to individuals.

4.2 Multilevel Model

The main object of study is default (default dependent variable). In this section, the analyses are performed considering default as the dependent variable. In section 4.5 the sensitivity analyses are carried out with the delayed dependent variable.

Considering the variables defined in Table 2 (where income and financing were combined in the variable finance/income, which is the natural logarithm of the division of financing by income) and Equation 4, the first-level model of the probability of default is (Equation 6):

Level 1: $p_{ij} = \dfrac{1}{1+e^{-Z_{ij}}}$ (6)

Where Z_{ij} is the logit of individual i belonging to branch j and is calculated by Equation 7:

$$Z_{ij} = b_{0j} + b_{1j}.idade_i + b_{2j}.sexo_i + b_{3j}.estado\ civil_i + b_{4j}.fund_i + b_{5j}.superior_i + b_{6j}.tempo\ de\ relatcionamento_i + b_{7j}.proventista_i + b_{8j}.financ/renda_i \quad (7)$$

Where b_{qj} (q = 0, 1, ..., Q) refer to level 1 coefficients, later defined by Equation 8:

Level 2: $b_{qj} = \gamma_{q0} + u_{q0}$ (8)

Where γ_{q0} refers to the level 2 coefficient and u_{q0} is the random level 2 effect, normally distributed, with zero mean and variance τ_{qq}. Besides, any error term independent of u_{q0} show zero mean and variance $\pi^2/3$.

4.2.1 Step One: Multilevel Null Model

The null model is the first step of the analysis (Table 6). With it, the objective is: (i) to verify the relevance of the multilevel modeling and (ii) to identify the origin of the variability of the variable under study (default), at the levels of analysis (individual and bank branch).

Table 6. Null multilevel model (default dependent variable)

Coefficient	Estimation	Default error (p-value)
Constant	-3,9069	0,1835 (0,000)
Random effect	0,7632	0,3224
LR test	16,02	(0,0000)

Source: elaborated by the authors.

From the random effect shown in Table 6 and the variance of the error term of $\pi^2/3$ (Fávero & Belfiore, 2017, p. 928), the intraclass correlation (according to the nomenclature of Raudenbush and Bryk, 2002) is calculated in 18 .8%, through Equation 9. Therefore, it can be said that 18.8% of the variance in the probability of default is due to the influence of the bank branches to which the individuals belong (the complement referring to the portion of the variance due to the influence of individuals).

$$rho = \frac{\tau_{00}}{\tau_{00} + \frac{\pi^2}{3}} = \frac{0{,}763226}{0{,}763226 + \frac{\pi^2}{3}} = 0{,}188 \quad (9)$$

These results indicate that there is variability between individuals and variability between bank branches (H-II), considering that the variance can be broken down into two significant portions: the portion originating from the individual level and the portion originating from the bank branch level. Therefore, individuals with the same characteristics from different bank branches have different default probabilities. The LR (likelihood-ratio) test corroborates the superiority of the multilevel logistic model compared to the traditional logistic model, reinforcing what had already been observed about H-II (significance of 0.000).

4.2.2 Step Two: Final Multilevel Model

The model presented in Table 7 includes all variables available at the individual level (second and last step of the analysis of this article), including their financing contract, except for the financing amount and income variables, replaced by the finance/income variable (calculated as the financing/income amount) for the reasons already explained in item 3.1. Given that the characteristics of the branches were not provided by the financial institution (second level of analysis) and that the tests indicated that there was no significant randomness of the angular coefficients, the model in Table 7 is the final model.

Table 7. Final multilevel model (default dependent variable)

	All variables		Meaningful variables only	
Variable	Estimation	Default error (p-value)	Estimation	Default error (p-value)
Age	-0.0312	0.0115 (0.007)	-0.0316	0.0110 (0.004)
Relationship time	-0.0555	0.0228 (0.015)	-0.0553	0.0228 (0.015)
Elementary	0.4756	0.2558 (0.063)	0.4908	0.2545 (0.054)
Higher	-0.8816	0.4684 (0.060)	-0.8915	0.4681 (0.057)
Financing/ income	0.5963	0.2026 (0.003)	0.5784	0.2004 (0.004)
Receives salary at the bank	-2.4367	0.7183 (0.001)	-2.4438	0.7182 (0.001)
Gender	-0.1027	0.1698 (0.545)		
Marital status	-0.0172	0.2087 (0.934)		
Constant	-4.44626	0.9092 (0.000)	-4.4324	0.9077 (0.000)
Random effect	0.4815	0.2763	0.4825	0.2759
LR test	7.96	(0.0024)	8.03	(0.0023)

Source: elaborated by the authors

The positively significant variables of individuals, that is, which increase the probability of default, ceteris paribus, are: elementary (level of education; low education indicates poor financial decisions) and finance/income ratio (high commitment of income with the payment of financing higher indicates greater chance of default). Other authors did not use this last variable (proposed in this study), having opted to use the financing amount and income separately.

The negatively significant individual variables, that is, which decrease the probability of default, ceteris paribus, are: age (older people may have greater responsibility or less propensity to risk), in line with Ferreira, Celso and Barbosa Neto (2012) and Gouvêa, Gonçalves and Mantovani (2013), contrary to Locatelli et al. (2015) and Maciel and Maciel (2017) and partially in line with Jannuzzi (2010); relationship time (greater commitment to the banking relationship), shared by Ferreira et al. (2011), Locatelli et al. (2015) and Ferreira, Celso and Barbosa Neto (2012); higher (education level; high education indicates better financial decisions), in line with Ferreira et al. (2011), Ferreira, Celso and Barbosa Neto (2012), Maciel and Maciel (2017) and Sousa, Petri and Anjo (2018) and contrary to Locatelli et al. (2015) and people who receive salary at the bank (salary is deposited into the account managed by the bank). The last variable was not considered by the other authors.

The variables referring to gender and marital status were not significant, being excluded from the final model. While the former was not significant in Ferreira, Celso and Barbosa Neto (2012), Lopes et al. (2017), Maciel and Maciel (2017) and Ritta, Gorla and Hein (2015), marital status was not statistically significant for Ferreira et al. (2011) and Sousa, Petri and Anjo (2018).

On the other hand, the variable referring to gender was significant for Ferreira et al. (2011), Locatelli et al. (2015), Gonçalves et al. (2013), Jannuzzi (2010) and Sousa, Petri and Anjo (2018), who found a higher probability of default for males. Marital status was shown to be statistically significant for Gouvêa, Gonçalves and Mantovani (2013) – according to them, being single increases the probability of default, and for Locatelli et al. (2015), Ferreira, Celso and Barbosa Neto (2012), Lopes et al. (2017), Jannuzzi (2010), Maciel and Maciel (2017) and Ritta, Gorla and Hein (2015) – according to them, being married (or in some cases, not single) increases the probability of default.

Therefore, considering that there are individual variables that explain the probability of default (with statistical significance), the H-I hypothesis (individual characteristics explain differences between individuals) was also corroborated.

4.3 Traditional Logistic Model

Table 8 presents the coefficients of the traditional logistic model (only with the significant variables, obtained through the stepwise procedure), which disregards the nested aspect of the data and for the default dependent variable. It is observed that these are the same variables and with the same signs as those presented in Table 7, for the multilevel logistic model. The marital status and gender variables were discarded, as they did not have statistical significance in explaining the probability of default of individuals.

Table 8. Traditional logistic model (default dependent variable)

Variable	Estimation	Default error (p-value)
Age	-0,0310	0,0108 (0,004)
Relationship time	-0,0552	0,0223 (0,013)
Elementary	0,4949	0,2474 (0,045)
Higher	-0,9540	0,4621 (0,039)
Financing/income	0,6051	0,1938 (0,002)
Receives salary at the bank	-2,4762	0,7164 (0,001)
Constant	-4,2571	0,8733 (0,000)

Source: elaborated by the authors

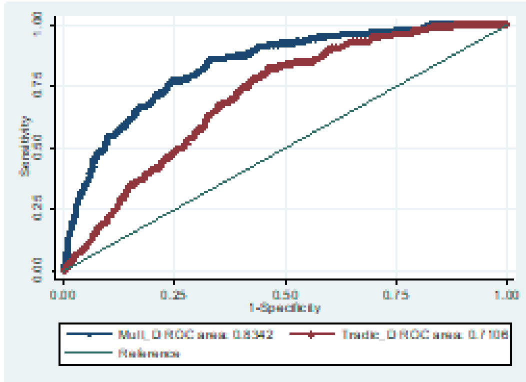
To confirm the result of the LR test shown in Table 7, which indicates the superiority of the multilevel logistic model over the traditional logistic model, it is observed that the log likelihood (LL) of the multilevel model is -658.2227, compared to -662. 4228 of the traditional model.

4.4 ROC (Receiver Operating Characteristic) Curve

According to Fávero and Belfiore (2017), the ROC curve is widely used in credit risk and default probability management models. It measures the behavior of the trade-off between specificity and sensitivity. Specificity concerns the percentage of correct answers for a given event, considering the observations that are not an event, given a cutoff. Sensitivity, on the other hand, corresponds to the hit rate of a given event, from a cutoff, considering only the observations that really are events. The cutoff, in turn, is a cutoff point defined by the researcher (or at the discretion of an organization) so that observations are classified according to their calculated probabilities (Fávero & Belfiore, 2017). According to Khudnitskaya (2010), in applications for retail financial institutions, the ROC curve shows the trade-off between the benefits obtained by the creditor by correctly classifying defaulters and the costs by incorrectly classifying non-defaulters.

Following what was done by Khudnitskaya (2010), the ROC curve was used in this work to confirm the predictive accuracy of the scoring models estimated from the delimited area, not intending to determine cutoff values. The area under the ROC curve of the traditional logistic model (Tradic_D ROC) is 0.7106, while the area for the multilevel model (Mult_D ROC) is 0.8342 (Figure 1).

Figure 1. ROC curve between the multilevel models and the traditional logistic model.



Source: Stata.

The test for equality of areas under the ROC curves was developed from a Stata code developed by Cleves (2002), from an algorithm suggested by DeLong, DeLong and Clarke-Pearson (1988). The chi-square test indicates that the areas are different, at a significance level of 0.000, which suggests that there is a significant difference between the areas and allows us to conclude, again, that there is superiority of the multilevel logistic model over the traditional logistic model.

4.5 Sensitivity Analysis

As a sensitivity analysis, the dependent variable was changed from (a) default, which considers delays from 180 days (159 individuals, corresponding to 3.11% of the sample) to (b) delayed, which considers delays from 1 day (1,366 individuals, corresponding to 26.72% of the sample).

4.5.1 Multilevel Model

The random effect of the null model (first step of the analysis, whose table was removed from the present study) for the delayed dependent variable is 0.3350, indicating that 9.24% of the variance comes from the bank branch level. The multilevel model proved to be preferable to the traditional model (significance 0.0000 of the LR test).

Table 9 presents the final multilevel model (second and last step of the analysis in this article) for the delayed dependent variable:

Table 9. Final multilevel model (delayed dependent variable)

	All variables		Meaningful variables only	
Variable	Estimation	Default error (p-value)	Estimation	Default error (p-value)
Age	-0,0084	0,0043 (0,052)	-0,0095	0,0041 (0,020)
Relationship time	-0,0406	0,0080 (0,000)	-0,0407	0,0080 (0,000)
Elementary	0,3392	0,1136 (0,003)	0,3390	0,1129 (0,003)

Higher	0,0161	0,1308 (0,902)		
Financing/income	0,5233	0,0778 (0,000)	0,5258	0,0770 (0,000)
Receives salary at the bank	-1,0721	0,1222 (0,000)	-1,0728	0,1221 (0,000)
Gender	0,0046	0,0687 (0,947)		
Marital status	-0,0622	0,0810 (0,442)		
Constant	-2,4387	0,3434 (0,000)	-2,4251	0,3411 (0,000)
Random effect	0,1944	0,0617	0,1960	0,0618
LR test	29,19	(0,0000)	29,87	(0,0000)

Source: elaborated by the authors

The final multilevel model for the delayed dependent variable excludes the same variables as the model for the default dependent variable (Table 7), but additionally excludes the superior variable, as it is not significant. All results presented in section 4.5.1 corroborate the results in section 4.2.

4.5.2 Traditional Logistic Model

Table 10 presents the coefficients of the traditional logistic model for the delayed dependent variable, as a complement to the sensitivity analysis.

Table 10. Traditional logistic model (delayed dependent variable)

Variable	Estimation	Default error (p-value_
Age	-0,0086	0,0040 (0,029)
Relationship time	-0,0404	0,0077 (0,000)
Elementary	0,3404	0,1086 (0,002)
Financing/income	0,5539	0,0735 (0,000)
Receives salary at the bank	-1,0715	0,1193 (0,000)
Constant	-2,4590	0,3277 (0,000)

Source: elaborated by the authors

The log likelihood of the traditional logistic model (–2,809.7434) is lower than the final multilevel model (-2,794.8075). The ROC curve of the traditional logistic model has an area (0.6991) smaller than that of the multilevel model (0.7664), being statistically different (0.000). Therefore, also for the delayed dependent variable, it is concluded that the multilevel logistic model is superior.

5 Final Considerations

A sample of 5,113 real estate financing contracts for individuals was studied under the Minha Casa Minha Vida Program, from clients of a nationwide financial institution, on September 30, 2019.

It is understood to be one of the first Brazilian studies on the subject. Other authors mostly study corporate credit, with few credit works for individuals, especially when it comes to real estate financing. In the latter group, no work was found in Brazil, with national coverage. It was found only the work by Locatelli et al. (2015), with operations in Minas Gerais, Espírito Santo and Rio de Janeiro. To relativize the financing variable, it was used, and apparently for the first time, the variable financing divided by income. The 'receives salary at the bank' variable (clients who receive a salary at the same credit institution) was also used for the first time.

Another innovation of the present work is the use of the multilevel technique, in line with the characteristics of the database, which presents individuals from 1,448 bank branches throughout Brazil, configuring the hierarchical character of the sample. The inhomogeneity of the distribution of individuals across the bank branches means that, if there is some influence of the bank branch on the variability of the default probability, the incorrect treatment of the nesting can interfere in the result by inflating the influence of the bank branch level by the number of clients of the respective bank branches.

In this sense, the objective of the work is to identify (i) the determinants of default in real estate financing operations and (ii) the influence of the individual's context (bank branch).

The tests indicate the superiority of the multilevel logistic model over the traditional logistic model (significance of the LR test of 0.000). Thus, it can be concluded that individuals with the same characteristics, but from different bank branches, have different default probabilities.

The correct treatment of the data, through hierarchical modeling, produced a model whose area under the ROC curve, widely used in credit risk management and default probability models (Fávero & Belfiore, 2017), was superior in relation to the traditional logistic model (area of 0.8356 for the former compared to an area of 0.7052 for this one, with a statistically significant difference of 0.000).

The final hierarchical logistic model identified the following variables with a positive influence on default: (a) financing/income ratio and (b) education of the individual to be up to elementary school (higher probability of default compared to individuals with higher education). On the other hand, the variables that negatively influence default are: (a) age, (b) length of relationship with the bank, (c) receiving at the bank and (d) higher education (incomplete, complete or postgraduate, master's and doctoral degrees; lower probability of default compared to individuals with secondary education). The variables gender (male and female) and marital status (single and others) were not significant in the analyzed sample.

As a sensitivity analysis, instead of analyzing default (characterized by delay of more than 180 days and which accounted for 3.11% of the sample), delayed was used as a dependent variable (characterized by delay of more than 1 day, representing 26.72% of the sample). The results were maintained, both in terms of the superiority of the hierarchical model over the traditional model, and for the significant variables (except for the fact that the superior variable is not significant for the delayed).

As a limitation of the work, there was no access to the characteristics of the bank branches, therefore, it is not possible to identify which of them may be related to default and even which of them moderate the effect of individual characteristics on the probability of default. Future works can explore these relationships as well as expand the analysis to a third level, with the grouping of bank branches (by geographic criteria – capital or countryside, state, etc. – or by some characteristic such as size).

Considering the results obtained, it is recommended that credit analysis for granting real estate financing be conducted through multilevel models that consider the nesting of individuals in their respective bank branches.

References

Albuquerque, P. H. M., Medina, F. A. S., & Silva, A. R. da. (2017). Regressão Logística Geograficamente Ponderada Aplicada a Modelos de Credit Scoring. *Revista Contabilidade e Financas*, 28(73), 93–112. doi:10.1590/1808-057x201703760.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–608. doi: 10.2307/2978933.

Amorim Neto, A. A., & Carmona, C. U. D. M. (2004). Modelagem do risco de crédito: Um estudo do segmento de pessoas físicas em um banco de varejo. *REAd-Revista Eletrônica de Administração*, 10(40), 1–23.

Arceneaux, K., & Nickerson, D. W. (2009). Modeling certainty with clustered data: a comparison of methods. *Political Analysis*, 17(2), 177-190. doi: 10.1093/pan/mpp004.

Backer, M., & Gosman, M. L. (1978). *Financial reporting and business liquidity*. Nova Iorque: National Association of Accountants.

Banco Central do Brasil - Departamento de Estatísticas. (2019a). Saldo da carteira de crédito com recursos direcionados - Pessoas físicas - Financiamento imobiliário com taxas reguladas Recuperado de https://dadosabertos.bcb.gov.br/dataset/20611-saldo-da-carteira-de-credito-com-recursos-direcionados---pessoas-fisicas---financiamento-imob. Acesso em 1 de novembro de 2019.

Banco Central do Brasil - Departamento de Estatísticas. (2019b). Inadimplência da carteira de crédito com recursos direcionados - Pessoas físicas - Financiamento imobiliário total. Recuperado de https://dadosabertos.bcb.gov.br/dataset/21151-inadimplencia-da-carteira-de-credito-com-recursos-direcionados--pessoas-fisicas--financiame. Acesso em 1 de novembro de 2019.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111. doi: 10.2307/2490171.

Brito, G. A. S., & Assaf Neto, A. (2008). Modelo de classificação de risco de crédito de empresas. *Revista de Contabilidade e Finanças*, 19(46), 18–29. doi:10.1590/S1519-70772008000100003.

Caouette, J. B., Altman, E. I., Narayanan, P., & Nimmo, R. W. J. (2009). Gestão do risco de crédito: *O grande desafio dos mercados financeiros globais* (2a ed.). Rio de Janeiro: Qualitymark.

Cleves, M.A. (2002). From the help desk: comparing areas under receiver operating characteristic curves from two or more probit or logit models. *The Stata Journal*, 2(3), 301-313. doi:10.1177/1536867X0200200307.

Courgeau, D. (2003). *Methodology and epistemology of multilevel analysis*. London: Kluwer Academic Publishers.

Crespi Júnior, H., Perera, L. C. J., & Kerr, R. B. (2017). Gerenciamento do ponto de corte na concessão do crédito direto ao consumidor. *Revista de Administração Contemporânea*, 21(2), 269–285. doi.: 10.1590/1982-7849rac2017170025.

Cunha, M. A. (2021). *Finanças e regionalidade: um modelo de credit scoring com uso da regressão logística geograficamente ponderada no projeto minha casa minha vida em Minas Gerais*. (Dissertação de Mestrado) Departamento de Administração, Universidade Federal de Uberlândia, Uberlândia, MG, Brasil. doi: 10.14393/ufu.di.2021.73.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3), 837–845. doi: 10.2307/2531595.

Elizabetsky, R. (1976). *Um modelo matemático para decisões de crédito no banco comercial*. (Dissertação de Mestrado) Departamento de Engenharia da Produção, Escola Politécnica da Universidade de São Paulo, São Paulo, SP, Brasil.

Fávero, L. P., & Belfiore, P. (2017). *Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®* (1a ed.). Rio de Janeiro: Elsevier.

Ferreira, M. A. M., Celso, A. S. dos S., & Barbosa Neto, J. E. (2012). Aplicação do modelo logit binominal na análise do risco de crédito em instituições bancárias. *Revista de Negócios*, 17(1), 41–59. doi: 10.7867/1980-4431.2012v17n1p38-55.

Ferreira, M. A. M., Oliveira, L. M. de, Santos, L. M. dos, & Abrantes, L. A. (2011). Previsão de risco de crédito para definição do perfil de clientes de um banco de varejo. *Revista de Negócios*, 16(2), 47–64.

FitzPatrick, P. J. (1932). A comparison of ratios of successful industrial enterprises with those of failed firm. *Certified Public*

Accountant, 6, 727–731.
Fundação João Pinheiro. (2018). Déficit habitacional no Brasil 2015. In *Estatísticas & Informações - Fundação João Pinheiro*. Recuperado de http://www.fjp.mg.gov.br/index.php/indicadores-sociais/deficit-habitacional-no-brasil

Gouvêa, M. A., Gonçalves, E. B., & Mantovani, D. (2013). Análise de risco de crédito com o uso de regressão logística. *Revista Contabilidade Vista & Revista*, 10(20), 96–123. doi:10.5007/2175-8069.2013v10n20p139.

Guimarães, I. A., & Chaves Neto, A. (2002). Reconhecimento de padrões: metodologias estatísticas em crédito ao consumidor. *RAE Eletrônica*, 1(2), 1–14. doi:10.1590/s1676-56482002000200006.

Hair Jr., J. F., & Fávero, L. P. (2019). Multilevel modeling for longitudinal data: concepts and applications. *RAUSP Management Journal*, 54(4), 459-489. doi: 10.1108/RAUSP-04-2019-0059.

Hitt, M. A., Beamish, P. W., Jackson, S. E., & Mathieu, J. E. (2007). Building theoretical and empirical bridges across levels: multilevel research in management. *Academy of Management Journal*, 50, 1385-1399. doi: 10.5465/amj.2007.28166219.

Jannuzzi, F. C. K. (2010). Um estudo sobre as variáveis que impactam a inadimplência no crédito concedido para projetos imobiliários (Dissertação de Mestrado). Universidade Estácio de Sá, Rio de Janeiro, RJ, Brasil.

Jiang, J., Liao, L., Xi, L., Wang, Z., & Xiang, H. (2021). Deciphering big data in consumer credit. *Journal of Empirical Finance*, 62, 28-45. doi: 10.1016/j.jempfin.2021.01.009.

Kanitz, S. C. (1978). Como prever falências. São Paulo: McGraw do Brasil.

Khudnitskaya, A. S. (2010). *Improved Credit Scoring with Multilevel Statistical Modelling*. (Tese de Doutorado). Universidade Técnica de Dortmund, Alemanha. doi: 10.17877/DE290R-257.

Locatelli, R. L., Ramalho, W., Silvério, R. A. de O., & Afonso, T. (2015). Determinantes da inadimplência no crédito habitacional direcionado a classe média emergente brasileira. *Revista de Finanças Aplicadas*, 1(1), 1–30.

Lopes, M. G., Ciribeli, J. P., Massardi, W. D. O., & Mendes, W. D. A. (2017). Análise dos indicadores de inadimplência nas linhas de crédito para pessoa física: um estudo utilizando modelo de regressão logística. *Estudos Do CEPE*, (46), 75–90. doi:10.17058/cepe.v0i46.11099.

Maciel, H. M., & Maciel, W. M. (2017). Análise da Inadimplência Bancária: Um Estudo de Caso da Região Metropolitana de Fortaleza. *Conexões - Ciência e Tecnologia*, 11(3), 12–23. doi:10.21439/conexoes.v11i3.867.

Matias, A. B. (1978). *Contribuição às técnicas de análise financeira: um modelo de concessão de crédito*. (Dissertação de Mestrado). Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, SP, Brasil.

Mathieu, J. E., & Chen, G. (2011). The etiology of the multilevel paradigm in management research. *Journal of Management*, 37(2), 610-641. doi: 10.1177/0149206310364663.

Merwin, C. L. (1942). *Financing small corporations in five manufacturing industries, 1926 - 1936*. Nova Iorque: National Bureau of Economic Research Inc.

Pinto, E. G. F. (2015). Financiamento imobiliário no Brasil: Uma análise histórica compreendendo o período de 1964 a 2013, norteada pelo arcabouço teórico pós-keynesiano e evolucionário. *Economia e Desenvolvimento*, 27(2), 276–296. doi:10.5902/141465092110

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata: categorical responses, counts, and survival* (Vol. II) (3ª ed.). College Station: Stata Press.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: applications and data analysis methods* (2ª ed.). Thousand Oaks: Sage Publications.

Ritta, C. de O., Gorla, M. C., & Hein, N. (2015). Modelo de regressão logística para análise de risco de crédito em uma instituição de microcrédito produtivo orientado. *Iberoamerican Journal of Industrial Engineering*, 7(13), 103–122. doi:10.13084/2175-8018/ijie.v7n13p103-122.

Securato, J. R. (2007). *Crédito: análise e avaliação do risco - pessoas físicas e jurídicas*. São Paulo: Saint Paul Institute of Finance.

Silva, J. P. (1982). *Modelos para classificação de empresas com vistas a concessão de crédito*. (Dissertação de Mestrado). Escola de Administração de Empresas de São Paulo. Fundação Getúlio Vargas, São Paulo, SP, Brasil.

Silva, J. P. (2016). *Gestão e análise do risco de crédito*. (9a ed.). São Paulo: Cengage Learning.
Smith, R. F., & Winakor, A. H. (1935). *Changes in the financial structure of unsuccessful industrial corporations*. Urbana: University of Illinois.

Sousa, Q. H., Petri, S. M., & Anjos, E. A. (2018). Análise dos fatores preditivos de risco para inadimplência dos cooperados em uma cooperativa de crédito. *III Congresso de Contabilidade da UFRGS e III Congresso de Iniciação Científica em Contabilidade da UFRG*, Porto Alegre, RS, Brasil, 3.

Steenbergen, M. R., & Jones, B. S. (2002). Modeling multilevel data structures. *American Journal of Political Science*, 46(1), 218-237. doi: 10.2307/3088424.