

Determinantes Hierárquicos da Inadimplência de Financiamento Imobiliário de Pessoa Física

Renata Barreto Lima¹ , Ricardo Goulart Serra² , Luiz Paulo Lopes Fávero³ 

¹Banco do Brasil. São Paulo-SP. Brasil.

²FECAP. São Paulo-SP. Brasil.

³Universidade de São Paulo. São Paulo-SP. Brasil.



¹renatabalima@gmail.com

²ricardo.serra@fecap.br

³lpfaver@usp.br

Editado por:

Orleans Silva Martins
Paulo Roberto da Cunha

Resumo

Objetivos: Os objetivos do artigo são: (i) identificar os fatores determinantes da inadimplência (*default*) de financiamento imobiliário e (ii) verificar a existência de influência do contexto do indivíduo (agência bancária) na inadimplência.

Método: O artigo estudou 5.113 contratos de financiamento imobiliário de pessoas físicas pertencentes a 1.448 agências de uma instituição financeira de abrangência nacional vigentes em 30 de setembro de 2019. Devido ao caráter multinível da amostra (nível 1: indivíduo e nível 2: agência) e considerando que as distintas agências apresentam diferentes níveis de performance (em diversos indicadores, inclusive no percentual de contratos de financiamento imobiliário inadimplentes), fez-se uso do modelo logístico multinível em detrimento ao modelo logístico tradicional.

Resultados: O modelo logístico multinível se mostrou superior ao modelo logístico tradicional (18,8% da variabilidade da probabilidade de inadimplência se refere ao nível agência). As características do indivíduo negativamente significantes são: idade, tempo de relacionamento, ser proventista e grau de instrução superior e as características positivamente significantes são: grau de instrução até o ensino fundamental e a relação financiamento/renda. As variáveis sexo e estado civil não se mostraram significativas.

Contribuições: As inovações do artigo são: (a) a abrangência nacional da amostra sobre inadimplência de pessoa física, (b) a utilização de duas variáveis inéditas e (c) o emprego do modelo logístico multinível. A identificação de que indivíduos com as mesmas características, porém, provenientes de agências diferentes, têm diferentes probabilidades de inadimplir, leva ao conhecimento dos acadêmicos e dos profissionais da área a recomendação da modelagem hierárquica para análise de crédito imobiliário.

Palavras-chave: financiamento imobiliário; inadimplência; pessoa física; modelo multinível.

Como citar:

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>

Recebido: Maio 09, 2021
Submetido a nova rodada: Outubro 07, 2021
Aceito: Outubro 18, 2021

Introdução

O problema habitacional brasileiro data do início do século XX, com o êxodo da população rural e de pequenas cidades para os grandes centros industriais, em busca dos empregos gerados pelo desenvolvimento manufatureiro-industrial (Pinto, 2015). Inúmeras iniciativas governamentais foram criadas para endereçar o problema: construção de conjuntos habitacionais, na década de 30; Fundação Casa Popular (FCP), na década de 40; Banco Nacional de Habitação (BNH), em 1964; Fundo de Compensação de Variações Salariais (FCVS), em 1967; Coeficiente de Equiparação Salarial (CES), em 1969; Sistema de Financiamento Imobiliário (SFI), em 1997; Programa de Aceleração do Crescimento (PAC), em 2007 e Programa Minha Casa Minha Vida (PMCMV), em 2009. Apesar das inúmeras iniciativas do governo, o déficit habitacional no Brasil foi estimado em 6,355 milhões de domicílios, correspondente a 9,3% do estoque de domicílios particulares (Fundação João Pinheiro, 2018).

Considerando os altos valores dos imóveis e o “ônus excessivo com aluguel urbano” (equivalente a um dispêndio superior a 30% da renda familiar, segundo a Fundação João Pinheiro (2018), a alternativa é a aquisição por meio de financiamento imobiliário. É de interesse das instituições financeiras financiar a aquisição de imóveis desde que os mutuários se mantenham adimplentes, apesar da garantia dada pelo imóvel financiado (não é de interesse das instituições financeiras arcar com os custos relativos à retomada da garantia e sua posterior venda).

O saldo da carteira de crédito com recursos direcionados a pessoas físicas por meio de financiamento imobiliário com taxas reguladas era, em novembro de 2019, de R\$ 569.593 milhões, uma expressiva elevação comparada ao valor de R\$ 79.094 milhões em janeiro de 2010 (Banco Central do Brasil, 2019a). A inadimplência de financiamentos com destinação específica era, em novembro de 2019, de 1,6%, inferior a 2,3% em março de 2011, início da série histórica (Banco Central do Brasil, 2019b).

A identificação dos fatores determinantes da inadimplência dos financiamentos imobiliários permitiria a diminuição do risco destas operações e, conseqüentemente, a possibilidade de operar com juros menores e de ofertar maior volume de crédito, beneficiando: as instituições financeiras, a economia como um todo e inúmeras famílias. Nota-se que a melhor maneira de alocar crédito para os consumidores é uma questão crítica para as empresas de

maneira geral, em específico, as instituições financeiras. Não obstante a abrangência de interesses sobre o tema (financeiro, econômico e social), poucos trabalhos acadêmicos abordam-no, acredita-se, pela dificuldade de obtenção de base de dados.

O presente trabalho analisa contratos de financiamento imobiliário dentro do âmbito do Programa Minha Casa Minha Vida, de um grande banco brasileiro com abrangência nacional. São 5.113 contratos de clientes provenientes de 1.448 agências bancárias, vigentes em 30 de setembro de 2019.

As teorias existentes, propostas dentro do contexto de um modelo com equação única, consideram não haver conexão entre os indivíduos e a sociedade em que vivem, conforme observa Courgeau (2003). Análises que ignoram esta conexão, segundo o autor, descrevem o comportamento dos indivíduos de forma equivocada; sendo que a análise correta dos fenômenos só pode ser realizada a partir do reconhecimento destas conexões. De acordo com Khudnitskaya (2010), a estrutura de dados aninhada ou estrutura hierárquica é típica em ciências sociais.

Nota-se (1) que os contratos são concedidos para indivíduos (primeiro nível de análise) pertencentes a um contexto, sua agência bancária (segundo nível de análise). Considerando que (2) as avaliações de resultados, realizadas por agência, constatam que as agências individualmente apresentam resultados distintos (por exemplo, percentual dos contratos imobiliários inadimplentes), é razoável admitir que o contexto dos indivíduos possa exercer influência sobre os fatores determinantes da inadimplência, de acordo com o apontado por Courgeau (2003).

O uso de equações em níveis permite que o pesquisador transcenda as teorias de equações únicas. Conclui-se, portanto, que se faz necessária a utilização de ferramenta capaz de capturar (1) os dois níveis de análise e (2) a influência da agência no inadimplemento, no caso: um modelo logístico multinível. Tal modelagem tem como benefício desagregar a variância da variável dependente nos diversos níveis da análise, de maneira a explicitar a influência dos diversos níveis de análise. Logo, no contexto do presente estudo, tal modelagem permite responder a dois questionamentos: quais características dos indivíduos estão relacionadas à probabilidade de inadimplir e se a conexão entre o indivíduo e o seu grupo (agência

bancária) é relevante na análise de inadimplência?

Portanto, os objetivos deste trabalho são: (i) identificar os fatores determinantes da inadimplência (*default*) de financiamento imobiliário e (ii) verificar a existência de influência do contexto do indivíduo (agência bancária) na inadimplência; ou seja, se indivíduos com mesmas características, porém, de distintas agência, têm (ou não) diferentes probabilidades de inadimplir.

O artigo é pioneiro em três aspectos: (i) no estudo de financiamento imobiliário com abrangência nacional (não se tem conhecimento de artigos nacionais que tenham tido acesso a uma base nacional), (ii) na definição de duas variáveis não utilizadas em outros estudos (financiamento ponderado pela renda, ou seja, financiamento/renda e proventista, ou seja, receber o salário no próprio banco) e (iii) na utilização de ferramenta multinível na análise de crédito (em geral).

Além desta seção introdutória, o artigo tem: a seção 2 com o referencial teórico; a seção 3 com a metodologia; a modelagem e a análise de dados (seção 4); e a última seção, com as considerações finais.

2 Referencial Teórico

2.1 Crédito

Para uma instituição que tem a intermediação financeira como sua principal atividade, o crédito consiste em disponibilizar ao cliente (mutuário, ou tomador de recurso) um determinado valor sob a forma de empréstimo ou financiamento, mediante a promessa de pagamento em data futura (Silva, 2016). Para Securato (2007, p.17), o termo crédito determina uma “relação de confiança entre duas (ou mais) partes numa determinada operação”. Assim, tendo em vista essa relação de confiança, pode-se dizer que uma operação de crédito envolve a expectativa do recebimento de um valor por uma das partes em um determinado período de tempo (Brito & Assaf Neto, 2008).

2.2 Risco de Crédito

O risco de crédito é a probabilidade de que o retorno de capital não seja realizado conforme o esperado (Caouette, Altman, Narayanan, & Nimmo, 2009). A mensuração do risco de crédito é o processo pelo qual a instituição financeira quantifica sua probabilidade de perda, caso os pagamentos das operações de crédito não se confirmem (Brito & Assaf Neto, 2008). Pode ser avaliado individual (por contrato) ou coletivamente (carteira de contratos).

A análise de risco de crédito de pessoas físicas passa pela observação dos chamados Cs do crédito (caráter, capacidade, capital, colateral e condições), podendo ser feita tanto por critério julgamental quanto por processos estatísticos (Silva, 2016). A análise julgamental conta com critérios subjetivos estando subordinada a experiência do analista ou gestor do crédito. O processo estatístico possibilita ampliar a gama de análises, diminuindo seu custo. Para Securato (2007), os Cs do crédito para pessoas físicas podem ser obtidos por – caráter: baseado em informações cadastrais próprias do credor ou de terceiros (como Serasa e SPCPC); capacidade: diretamente relacionada a renda do indivíduo; capital: patrimônio pessoal do solicitante; colateral: garantias que o solicitante coloca à disposição do credor; e condições: fatores macro ou microeconômicos que influenciam na concessão do crédito.

2.3 Credit Scoring

O meio de controle de risco mais utilizado é o sistema de pontuação (Guimarães & Chaves Neto, 2002). O *credit scoring*, um dos principais métodos estatísticos, pode ser definido, portanto, como o processo de atribuição de pontos às variáveis de decisão mediante técnicas estatísticas (Amorim Neto & Carmona, 2004). Sendo assim, atribui-se um valor a cada característica de um cliente, elabora-se um score e se compara com um ponto de corte, deliberando-se pela concessão ou não do crédito. O estabelecimento do ponto de corte se dá a fim de encontrar a quantidade de maus pagadores probabilisticamente aceitos que causariam um prejuízo menor ao lucro gerado pelos bons pagadores (Crespi Júnior, Pereira, & Kerr, 2017). Tais parâmetros utilizados para a concessão de crédito à pessoa física se norteiam nos já mencionados Cs do Crédito.

Os modelos de classificação são geralmente desenvolvidos com as seguintes técnicas estatísticas: regressão linear múltipla, análise discriminante, programação linear, algoritmo genético, análise de sobrevivência, árvore de decisão, redes neurais e regressão logística, com a finalidade de tentar identificar os fatores determinantes da inadimplência (Locatelli, Ramalho, Silvério, & Afonso, 2015).

2.4 Trabalhos Empíricos

A maior parte dos estudos é voltada à análise de empresas, por meio de seus indicadores econômico-financeiros. No exterior, destacam-se os estudos realizados por FitzPatrick (1932), Smith e Winakor (1935), Merwin (1942), Beaver (1966), Tamari (1966), Altman (1968), Backer e Gosman (1978) e Topa (1979), com estudos que se estendem até o presente (Yildirim, 2020). No Brasil, os estudos mais

importantes (Silva, 2016) são: Elizabetsky (1976), Kanitz (1978), Matias (1978) e Silva (1982).

No entanto, existem poucos trabalhos a respeito de análise de risco de crédito de pessoas físicas. Os estudos brasileiros e as variáveis neles utilizadas estão sumarizados na Tabela 1. Deve-se ponderar também a dificuldade no acesso dos dados pelos pesquisadores (p. ex.: Amorim Neto e Carmona (2004) não obtiveram acesso aos dados de renda).

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

Variáveis independentes	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	Contagem
Idade	x	.	x	x	x	x	x	x	x	x	x	x	x	13
Renda	x	x	x		x	x	x	x	x	x	x	x	x	12
Estado civil	x	x			x	x	x	x		x	x	x	x	10
Grau de instrução/Escolar	x	x	x	x	x			x		x	x	x		9
Sexo /Gênero	x	x				x	x	x		x	x	x	x	9
Tempo de relacionamento bancário	x	x	x		x	x				x				6
Valor financiado		x				x	x		x		x		x	6
Endereço residencial /CEP			x	x			x	x	x					5
Nº parcelas / prazo			x			x			x		x		x	5
Valor das parcelas						x	x		x		x		x	5
Tipo de ocupação	x			x			x			x				4
Tipo de empréstimo						x	x				x		x	4
Restrição (interna/externa)	x			x									x	3
Capacidade de pagamento		x								x	x			3
Tempo de residência	x						x	x						3
Tipo de residência								x			x	x		3
Nº dependentes											x	x		2
Tempo na ocupação atual							x	x						2

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

Fonte: elaborada pelos autores.

Pode-se verificar, na Tabela 1, que os cinco parâmetros mais utilizados nas análises de crédito para pessoa física são relacionados a dados socioeconômicos do mutuário (idade, renda, estado civil, grau de instrução/escolaridade e sexo/gênero).

Dos trabalhos pesquisados, apenas o de Locatelli et al. (2015) e o de Cunha (2021) se referem a crédito imobiliário, porém, são estudos regionais (sem a abrangência nacional). Januzzi

(2010) por sua vez, estuda operações destinadas à reforma e ampliação de imóveis (não a sua aquisição). Ritta, Gorla e Hein (2015) analisam contratos para microcrédito produtivo orientado. Guimarães e Chaves Neto (2002) analisam operações de cartão de crédito. Os demais trabalhos estudam empréstimos e financiamentos em geral (CDC).

Alguns estudos não se limitam a pesquisar os fatores determinantes da inadimplência de clientes, testando a mesma base em modelos estatísticos diferentes, como os trabalhos de: Amorim Neto e Carmona (2004), comparando regressão logística a análise discriminante, concluindo pela similaridade dos modelos; Gouvêa, Gonçalves e Mantovani (2013), testando regressão logística e modelos neurais e Guimarães e Chaves Neto (2002), avaliando função discriminante linear de Fisher e regressão logística, sendo que estes dois últimos estudos demonstram equivalência entre os modelos estudados, com ligeira vantagem ao modelo de regressão logística.

No Brasil, Albuquerque, Medina e Silva (2017) analisaram Crédito Direto ao Consumidor (CDC) e Cunha (2021) analisou financiamento imobiliário, ambos por meio de modelo de regressão logística geograficamente ponderada, registrando haver e não haver melhora do modelo comparativamente à regressão logística tradicional, respectivamente. No entanto, os trabalhos não usaram uma técnica multinível. No exterior, Khudnitskaya (2010) desenvolveu com base em operações de cartões de crédito, um *scorecard* de crédito a partir do modelo multinível, sendo nível 1: indivíduo e nível 2: microambiente (baseado em características em comum relativas a condições econômicas e demográficas em suas áreas de residência, mas não necessariamente de ordem geográfica). Os demais estudos trabalham com modelos com um único nível.

Mais recentemente, com o aumento do pedido de crédito *online*, a partir da pandemia, estudos internacionais identificam que as instituições financeiras começaram a usar, para análise de crédito ao consumidor, informações que vão além das fontes tradicionais de informação (p. ex.: informações financeiras). Estas informações são obtidas, entre outras fontes, de dados de aplicativos móveis (Jiang, Liao, Xi, Wang, & Xiang, 2021). Zhou, Wang, Ren, & Chen (2021) utilizaram dados sobre o comportamento dos indivíduos (p. ex.: uso do telefone).

3 Metodologia

3.1 Amostra e Variáveis

O universo é composto de 255.926 contratos de financiamento imobiliário de pessoas físicas sob o programa Minha Casa Minha Vida de uma instituição financeira de abrangência nacional, de operações vigentes em 30 de setembro de 2019. Os autores não tiveram acesso ao universo, mas apenas a uma amostra aleatória de 2% de contratos retirada

do universo por Estado e Distrito Federal, resultando em 5.125 observações distribuídas em 1.448 agências. Os clientes foram identificados por códigos (de 1 a 5.125). Ambos os procedimentos preservaram a confidencialidade dos indivíduos (trata-se de dados confidenciais). A Tabela 2 apresenta as variáveis fornecidas pela instituição financeira, além da agência do cliente (também codificada de 1 a 1.448) e o estado de origem. As duas primeiras variáveis da Tabela 2 (*default* e *atraso*) foram utilizadas em dois modelos distintos (modelo principal – seção 4.2 e análise de sensibilidade – seção 4.5, respectivamente) como variáveis dependentes. Demais variáveis foram utilizadas como variáveis explicativas (de primeiro nível).

Tabela 2. Descrição das variáveis dependentes e independentes

Variável	Especificação
Default	- Sem atraso ou atraso inferior a 180 dias (<i>dummy</i> = 0) - Atraso superior a 180 dias (<i>dummy</i> = 1)
Atrasado	- Sem atraso (<i>dummy</i> = 0) - Atraso a partir de 01 dia (<i>dummy</i> = 1)
Valor Financiado	Valor do financiamento, no momento da contratação, em Reais (R\$)
Renda	Valor da renda mensal, em Reais (R\$)
Idade	Idade do correntista, em anos
Sexo	- Masculino (<i>dummy</i> = 0) - Feminino (<i>dummy</i> = 1)
Estado Civil	- Solteiro (<i>dummy</i> = 0) - Demais (<i>dummy</i> = 1)
Grau de Instrução - Fundamental	- Analfabeto e ensino fundamental (<i>dummy</i> = 1) - Demais (<i>dummy</i> = 0)
Grau de Instrução - Superior	- Superior (<i>dummy</i> = 1) - Demais (<i>dummy</i> = 0)
Tempo de relacionamento	Tempo de relacionamento do cliente com o banco, em anos
Proventista	- Não recebe o salário no banco (<i>dummy</i> = 0) - Recebe o salário no banco (<i>dummy</i> = 1)

Observações: a. casado, separado, divorciado ou viúvo; b. foram analisadas três categorias de grau de instrução, portanto, demandando duas variáveis *dummy* (Fundamental e Superior), sendo o grau de instrução base o ensino médio. c. superior incompleto, superior completo e pós-graduados (*lacto e stricto sensu*).

Fonte: elaborada pelos autores.

A amostra foi analisada para verificação de possíveis ausências e/ou divergências. Foram excluídas: nove observações por ausência de tempo de relacionamento; duas observações com renda mensal de “R\$ 0,01” e uma observação com informação divergente aos códigos fornecidos para os campos “Estado Civil” e “Grau Instrução”, não sendo possível auferir o código correto. Assim, a amostra analisada contém 5.113 observações válidas.

Partindo do pressuposto que o valor financiado e a renda individualmente não são variáveis absolutas, criou-se a variável “financiamento/renda” (calculado como o logaritmo natural do valor financiado dividido pela renda mensal),

com o intuito de relativizar o valor financiado pela renda do indivíduo. Ressalta-se que o valor financiado representa um dado do passado (no momento da contratação) e a renda do indivíduo representa um dado atualizado (considerando-se a necessidade de atualização do cadastro dos clientes). Entende-se que a variável criada (como inovação do presente estudo) se constitui em um melhor indicador da capacidade de pagamento do indivíduo do que as variáveis absolutas: valor financiado e renda, quando individualmente utilizadas (prática comum nos estudos da área).

3.2 Hipóteses de Pesquisa

Com os objetivos de: (i) identificar os fatores determinantes da inadimplência (*default*) de financiamento imobiliário e (ii) verificar a existência de influência do contexto do indivíduo (agência bancária) na inadimplência, as hipóteses de pesquisa são:

Hipótese I: há características dos indivíduos que explicam a variabilidade da probabilidade de inadimplência de indivíduos de uma mesma agência (H-I, características dos indivíduos explicam diferenças entre indivíduos).

Hipótese II: há variabilidade significativa da probabilidade de inadimplência dos indivíduos de diferentes agências (H-II, variabilidade entre agências).

As hipóteses alternativas às H-I (H-Ia) e H-II (H-IIa) são as suas negativas.

Neste sentido, além de identificar as características dos indivíduos relevantes para a análise de concessão de crédito, tem-se o interesse em analisar se indivíduos com as mesmas características, porém, de distintas agências, têm (ou não) diferentes probabilidades de inadimplir. Considerando as hipóteses de pesquisa descritas, o modelo a ser utilizado é o modelo logístico binário multinível.

3.3 O Modelo Logístico Binário Multinível

A regressão logística é muito utilizada para modelagem de análise de risco de crédito, considerando a característica de estimação de probabilidade de classificação prévia dos clientes como adimplentes e inadimplentes (Ritta, Gorla, & Hein, 2015). É a técnica que melhor auferir a previsibilidade de acerto em comparação com outras técnicas estatísticas, segundo análises realizadas por Guimarães e Chaves Neto (2002).

Os modelos logísticos binários diferem dos tradicionais modelos estimados por mínimos quadrados ordinários (MQO)

uma vez que se concentram na análise de probabilidades de determinado evento Y que se apresenta de maneira discreta e qualitativa, de modo que pode ser definida uma variável *dummy* para o caracterizar ($Y = 1$ para caracterizar a ocorrência do evento, e 0 , caso contrário).

Desta maneira, no modelo logístico, a probabilidade de ocorrência de um evento é dada, segundo Fávero e Belfiore (2017), por:

$$p_i = 1 / (1 + e^{Z_i}) \quad (1)$$

em que p_i é a probabilidade de ocorrência do evento e Z é o logito, definido por:

$$Z_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} \quad (2)$$

em que α é a constante, β_j ($j = 1, 2, \dots, k$) são os parâmetros estimados para as variáveis explicativas, X_{ij} são as j variáveis explicativas, para cada observação i da amostra. Os parâmetros do logito são estimados por máxima verossimilhança, a partir da maximização do logaritmo natural da função de verossimilhança, definida por:

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

De maneira geral, pode-se definir um modelo logístico binário com dois níveis, em que o primeiro nível oferece as variáveis explicativas X_1, \dots, X_Q referentes a cada indivíduo i ($i = 1, \dots, n$), e o segundo nível, as variáveis explicativas W_1, \dots, W_S referentes a cada grupo ou contexto j ($j = 1, \dots, J$), invariantes para as observações pertencentes a um mesmo grupo:

$$P_{ij} = \frac{\text{Nível 1}}{1 + e^{-(b_{0j} + b_{1j} \cdot X_{1ij} + b_{2j} \cdot X_{2ij} + \dots + b_{Qj} \cdot X_{Qij})}} \quad (4)$$

em que p_{ij} representa a probabilidade de ocorrência do evento de interesse para cada observação i pertencente a determinado grupo j e b_{qj} ($q = 0, 1, \dots, Q$) se referem aos coeficientes de nível 1 (Raudenbush & Bryk, 2002; Rabe-Hesketh & Skrondal, 2012).

$$\text{Nível 2} \quad b_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} \cdot W_{sj} + u_{qj} \quad (5)$$

em que γ_{qs} ($s = 0, 1, \dots, S_q$) se referem aos coeficientes de nível 2 e u_{qj} são os efeitos aleatórios de nível 2, normalmente distribuídos, com média igual a zero e variância τ_{qj} . Além

disso, eventuais termos de erro independentes de u_{qj} apresentam média igual a zero e variância $\pi^2/3$.

De acordo com Fávero e Belfiore (2017), a principal vantagem dos modelos multinível sobre os modelos tradicionais de regressão (GLM - Generalized Linear Models) é a possibilidade de considerar o aninhamento natural dos dados, permitindo, assim, que sejam identificadas e analisadas heterogeneidades individuais e entre grupos a que pertencem estes indivíduos, tornando possível a especificação de componentes aleatórios em cada nível de análise.

Em outras palavras, os modelos multinível corrigem o fato de que as observações no mesmo grupo não sejam independentes e, portanto, comparadas às modelagens tradicionais, levam a estimativas não viesadas de erros-padrão dos parâmetros. Conforme enfatizam Steenbergen e Jones (2002), Arceneaux e Nickerson (2009) e Hair e Fávero (2019), se pesquisadores estiverem interessados em testar se covariáveis em nível de grupo moderam os efeitos em nível individual, os modelos multinível parecem ser a escolha mais apropriada. E um teste de razão de verossimilhança pode ser elaborado para que se verifique a adequação da estimação multinível à estrutura dos dados, bem como propicia uma comparação com estimações oriundas, por exemplo, de modelos tradicionais GLM.

Segundo Courgeau (2003), dentro de uma estrutura de modelo com uma única equação, parece não haver conexão entre os indivíduos e a sociedade em que vivem. Nesse sentido, o uso de equações de nível permite ao pesquisador “pular” de uma ciência para outra: estudantes e escolas, famílias e bairros, empresas e países. Ignorar esse relacionamento significa elaborar análises incorretas sobre o comportamento dos indivíduos e, igualmente, sobre o comportamento dos grupos. Somente o reconhecimento dessas influências recíprocas permite a correta análise dos fenômenos.

Isso está de acordo com o chamado por Mathieu e Chen (2011) de paradigma multinível, que se refere a uma maneira de pensar: considerar os fenômenos de gestão em contexto e procurar variáveis direcionadoras não apenas da unidade focal de análise, mas também dos níveis acima e abaixo. Essa abordagem geralmente implica no desenvolvimento de teorias e investigações multidisciplinares, qual é o espírito articulado por Hitt, Beamish, Jackson e Mathieu (2007) ao discutir a construção de pontes teóricas e empíricas por meio de contextos na modelagem multinível. E este é o objetivo do presente trabalho, que procura relacionar níveis distintos, como atributos individuais e condições contextuais de agências bancárias sobre a probabilidade de inadimplência de financiamento imobiliário.

3 Modelagem e Análise dos Dados

4.1 Estatísticas Descritivas

A análise considera a agência como segundo nível. No entanto, por questões de confidencialidade do indivíduo e conforme descrito em 3.1, os autores tiveram acesso a uma amostra aleatória de 2% de contratos, retirada do universo por Estado e Distrito Federal, com a identificação dos indivíduos e das agências por um índice (de 1 a 5.125 para os indivíduos e de 1 a 1.448 para as agências). Portanto, pode-se apenas identificar os contratos analisados por Estado e Distrito Federal, conforme a Tabela 3.

Tabela 3. Distribuição dos Contratos por Estado e Distrito Federal

Estado/DF	%	Estado/DF	%	Estado/DF	%
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

Fonte: elaborada pelos autores.

A Tabela 4 apresenta a estatística descritiva das variáveis discretas (todas de primeiro nível, uma vez que ligadas aos indivíduos).

Tabela 4. Estatísticas descritivas das variáveis discretas

Variável	Especificação	Frequência	(%)
Default	- sem atraso, ou atraso inferior a 180 dias	4.954	96,89
	- atraso superior 180 dias	159	3,11
Atrasado	- sem atraso	3.747	73,28
	- em atraso (a partir de 01 dia)	1.366	26,72
Sexo	- masculino	2.724	53,28
	- feminino	2.389	46,72
Estado civil ^a	- solteiro(a)	3.625	70,90
	- casado(a)	1.020	19,95
	- viúvo(a)	44	0,86
	- separado(a)	52	1,02
	- divorciado(a)	372	7,28
Grau de instrução ^b	- analfabeto	2	0,04
	- ensino fundamental	483	9,45
	- ensino médio	3.396	66,42
	- superior incompleto	374	7,31
	- superior completo	830	16,23
	- pós-graduado	23	0,45
	- mestrado	2	0,04
- doutorado	3	0,06	
Proventista	- não	4.195	82,05
	- sim	918	17,95

Observações: a. representado por uma variável dummy em que 0 = solteiro e 1 = casado, viúvo, separado e divorciado; b. representado por duas variáveis dummies (por haver três categorias, tendo como grau de instrução base o ensino médio), sendo uma dummy para analfabeto e ensino fundamental e outra dummy para superior incompleto, superior completo e pós-graduados (lacto e stricto sensu).

Fonte: elaborada pelos autores.

Verifica-se, na Tabela 4, que 3,11% dos contratos estão em *default* (atrasos superiores a 180 dias), enquanto 26,72% estão atrasados (atrasos superiores a 01 dia). A amostra é composta de homens (53,28%) e mulheres (46,72%) em percentual similar, sendo a grande maioria solteira (70,90%), com ensino médio (66,42%) e não proventista (82,05%).

A Tabela 5 apresenta as estatísticas descritivas das variáveis contínuas.

Tabela 5. Estatísticas descritivas das variáveis contínuas

Variável	Média	Mediana	Desvio Padrão	Mínimo	Máximo
Idade	35,10	33,00	8,98	19	79
Tempo de relacionamento	7,64	6,00	4,92	0	33
Financiamento	91.052,32	91.040,00	19.927,84	18.665,84	198.842,70
Renda	2.257,79	1.800,00	1.552,95	400,00	22.616,33
Financ./renda (em ln)	3,82	3,88	0,51	0,00	5,53

Fonte: elaborada pelos autores.

A Tabela 5 indica que o indivíduo tem, em média, 35,1 anos, 7,6 anos de relacionamento com o banco, salário médio mensal de R\$ 2,3 mil e financiamento médio de R\$ 91,0 mil. A média do logaritmo natural do financiamento dividido pela renda mensal é 3,82 (equivalente a um financiamento 45,77 vezes maior do que a renda mensal).

Os indivíduos estão distribuídos em 1.448 agências, algumas das quais com grande concentração (as três agências com maior número de observações representam 3,93%; 3,03% e 2,97% do total, ante o valor esperado abaixo de 0,1%, equivalente a 1/1.448). Tal concentração reforça a necessidade da utilização de modelo multinível, pois, do contrário, pode-se inflacionar eventuais influências da agência (e suas características) na análise, ao atribuí-las aos indivíduos.

4.1 Estatísticas Descritivas

O objeto de estudo principal é a inadimplência (variável dependente *default*). Nesta seção, as análises são feitas considerando *default* como variável dependente. Na seção 4.5 se faz as análises de sensibilidade com a variável dependente atrasado.

Considerando as variáveis definidas na Tabela 2 (sendo que renda e financiamento foram combinados na variável financ./renda, que é o logaritmo natural da divisão de financiamento por renda) e a Equação 4, o modelo de primeiro nível da probabilidade de inadimplir é (Equação 6):

$$\text{Nível 1: } p_{ij} = \frac{1}{1 + e^{-Z_{ij}}} \quad (6)$$

Em que Z_{ij} é o logito do indivíduo i pertencente à agência j e é calculado pela Equação 7:

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

Em que b_{qj} ($q = 0, 1, \dots, Q$) se referem aos coeficientes de nível 1, posteriormente definidos pela Equação 8:

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

Em que γ_{q0} se refere ao coeficiente de nível 2 e u_{q0} é o efeito aleatório de nível 2, normalmente distribuídos, com média igual a zero e variância τ_{qq} . Além disso, eventual termo de erro independente de u_{q0} apresenta média igual a zero e variância $\pi^2/3$.

4.2.1 Primeiro Passo: Modelo Multinível Nulo

O modelo nulo é o primeiro passo da análise (Tabela 6). Com ele, objetiva-se: (i) verificar a pertinência da modelagem multinível e (ii) identificar a origem da variabilidade da variável em estudo (*default*), nos níveis de análise (indivíduo e agência).

Tabela 6. Modelo multinível nulo (variável dependente *default*)

Coefficiente	Estimativa	Erro Padrão (p-valor)
Constante	-3,9069	0,1835 (0,000)
Efeito Aleatório	0,7632	0,3224
Teste LR	16,02	(0,0000)

Fonte: elaborada pelos autores.

A partir do efeito aleatório apresentado na Tabela 6 e da variância do termo de erro de $\pi^2/3$ (Fávero & Belfiore, 2017, p. 928), calcula-se a correlação intraclasse (segundo nomenclatura de Raudenbush e Bryk, 2002) em 18,8%, por meio da Equação 9. Portanto, pode-se dizer que 18,8% da variância da probabilidade de inadimplência é devido a influência das agências às quais os indivíduos pertencem (sendo o complemento referente a parcela da variância devido a influência dos indivíduos).

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

Tais resultados indicam haver variabilidade entre indivíduos e variabilidade entre agências (H1), tendo em vista que a variância pode ser decomposta em duas parcelas significativas: a parcela oriunda do nível do indivíduo e a parcela oriunda do nível da agência. Logo, indivíduos com características iguais oriundos de diferentes agências têm diferentes probabilidades de inadimplir. O teste LR (*likelihood-ratio*) corrobora a superioridade do modelo logístico multinível comparativamente ao modelo logístico tradicional, reforçando o que já havia sido constatado sobre H1 (significância de 0,000).

4.2.2 Segundo Passo: Modelo Multinível Final

O modelo apresentado na Tabela 7 contempla todas as variáveis disponibilizadas no nível do indivíduo (segundo e último passo da análise do presente artigo), incluindo seu contrato de financiamento, com exceção das variáveis valor do financiamento e renda, substituídas pela variável *financ/renda* (calculada como valor do financiamento/renda) pelos motivos já explicitados no item 3.1. Dado que as características das agências não foram disponibilizadas pela instituição financeira (2º nível da análise) e que os testes indicaram não haver aleatoriedade significativa dos coeficientes angulares, o modelo da Tabela 7 é o modelo final.

Tabela 7. Modelo multinível final (variável dependente *default*)

Variável	Todas as variáveis		Apenas variáveis significativas	
	Estimativa	Erro padrão (p-valor)	Estimativa	Erro padrão (p-valor)
Idade	-0,0312	0,0115 (0,007)	-0,0316	0,0110 (0,004)
Tempo de relacionamento	-0,0555	0,0228 (0,015)	-0,0553	0,0228 (0,015)
Fundamental	0,4756	0,2558 (0,063)	0,4908	0,2545 (0,054)
Superior	-0,8816	0,4684 (0,060)	-0,8915	0,4681 (0,057)
Financ./renda	0,5963	0,2026 (0,003)	0,5784	0,2004 (0,004)
Proventista	-2,4367	0,7183 (0,001)	-2,4438	0,7182 (0,001)
Sexo	-0,1027	0,1698 (0,545)		
Estado civil	-0,0172	0,2087 (0,934)		
Constante	-4,44626	0,9092 (0,000)	-4,4324	0,9077 (0,000)
Efeito Aleatório	0,4815	0,2763	0,4825	0,2759
Teste LR	7,96	(0,0024)	8,03	(0,0023)

Fonte: elaborada pelos autores.

As variáveis dos indivíduos positivamente significativas, ou seja, que aumentam a probabilidade de *default*, *ceteris paribus*, são: o grau de instrução fundamental (baixa instrução indica piores decisões financeiras) e relação *financ/renda* (alto valor financiado em relação à renda indica maior chance de inadimplência). Demais autores não usaram esta última variável (proposta no presente estudo), tendo optado por usar o valor do financiamento e a renda isoladamente.

As variáveis do indivíduo negativamente significativas, ou seja, que diminuem a probabilidade de *default*, *ceteris paribus*, são: idade (mais velhos podem ter maior responsabilidade ou menor propensão ao risco), em linha com Ferreira, Celso e Barbosa Neto (2012) e Gouvêa, Gonçalves e Mantovani (2013), contrário a Locatelli et al. (2015) e Maciel e Maciel (2017) e parcialmente em linha com Jannuzzi (2010); tempo de relacionamento (maior o comprometimento com o relacionamento bancário), compartilhado por Ferreira et al. (2011), Locatelli et al. (2015) e Ferreira, Celso e Barbosa Neto (2012); o grau de instrução superior (alta instrução indica melhores decisões financeiras), em linha com Ferreira et al. (2011), Ferreira, Celso e Barbosa Neto (2012), Maciel e Maciel (2017) e Sousa, Petri e Anjo (2018) e contrário

a Locatelli et al. (2015) e proventista (salário cai na conta gerida pelo banco). A última variável não foi considerada pelos demais autores.

As variáveis referentes a sexo e estado civil não foram significantes, sendo excluídas do modelo final. Enquanto a primeira também não se mostrou significativa em Ferreira, Celso e Barbosa Neto (2012), Lopes et al. (2017), Maciel e Maciel (2017) e Ritta, Gorla e Hein (2015), o estado civil não apresentou significância estatística para Ferreira et al. (2011) e Sousa, Petri e Anjo (2018).

Por outro lado, a variável referente a sexo foi significativa para Ferreira et al. (2011), Locatelli et al. (2015), Gouvea, Gonçalves e Mantovani (2013), Jannuzzi (2010) e Sousa, Petri e Anjo (2018), que encontraram maior probabilidade de *default* para indivíduos do sexo masculino. Estado civil se mostrou estatisticamente significativo para Gouvêa, Gonçalves e Mantovani (2013) – em que ser solteiro eleva a probabilidade de inadimplência, e para Locatelli et al. (2015), Ferreira, Celso e Barbosa Neto (2012), Lopes et al. (2017), Jannuzzi (2010), Maciel e Maciel (2017) e Ritta, Gorla e Hein (2015) – em que ser casado (ou em alguns casos, diferente da situação de solteiro) eleva a probabilidade de inadimplência.

Portanto, considerando que existem variáveis dos indivíduos que explicam a probabilidade de inadimplência (com significância estatística), a hipótese H1 (características dos indivíduos explicam diferenças entre indivíduos) também foi corroborada.

4.3 Modelo Logístico Tradicional

A Tabela 8 apresenta os coeficientes do modelo logístico tradicional (apenas com as variáveis significantes, obtida por meio do procedimento *stepwise*), que desconsidera o aspecto aninhado dos dados e para a variável dependente *default*. Observa-se que são as mesmas variáveis e com os mesmos sinais que aqueles apresentados na Tabela 7, para o modelo logístico multinível. As variáveis estado civil e sexo foram descartadas, por não terem significância estatística na explicação da probabilidade de *default* dos indivíduos.

Tabela 8. Modelo logístico tradicional (variável dependente *default*)

Variável	Estimativa	Erro padrão (p-valor)
Idade	-0,0310	0,0108 (0,004)
Tempo de relacionamento	-0,0552	0,0223 (0,013)
Fundamental	0,4949	0,2474 (0,045)
Superior	-0,9540	0,4621 (0,039)
Financ/renda	0,6051	0,1938 (0,002)
Proventista	-2,4762	0,7164 (0,001)
Constante	-4,2571	0,8733 (0,000)

Fonte: elaborada pelos autores.

Para confirmar o resultado do teste LR apresentado na

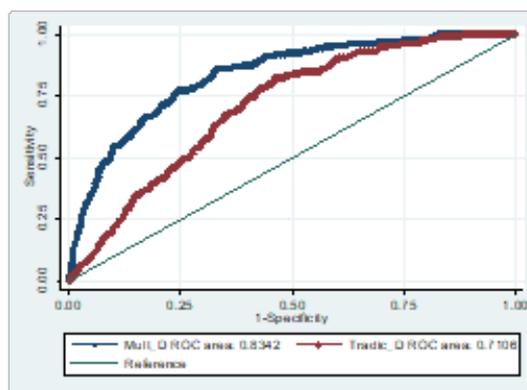
Tabela 7, que indica a superioridade do modelo multinível logístico sobre o modelo logístico tradicional, observa-se que o *log likelihood* (LL) do modelo multinível é -658,2227, comparativamente a -662,4228 do modelo tradicional.

4.4 Curva ROC (Receiver Operating Characteristic)

Segundo Fávero e Belfiore (2017), a curva ROC é amplamente utilizada em modelos de gestão de risco de crédito e de probabilidade de *default*. Ela mede o comportamento do *trade-off* entre especificidade e sensibilidade. A especificidade diz respeito ao percentual de acerto de um determinado evento, considerando as observações que não são evento, dado um determinado *cutoff*. Já a sensibilidade corresponde a taxa de acerto de determinado evento, a partir de um *cutoff*, considerando apenas as observações que realmente são evento. O *cutoff*, por sua vez, é um ponto de corte definido pelo pesquisador (ou a critério de uma organização) a fim de que sejam classificadas as observações em função de suas probabilidades calculadas (Fávero & Belfiore, 2017). De acordo com Khudnitskaya (2010), nas aplicações para instituições financeiras de varejo, a curva ROC mostra o *trade-off* entre os benefícios obtidos pelo credor ao classificar corretamente os inadimplentes e os custos ao classificar incorretamente os não-inadimplentes.

Seguindo o realizado por Khudnitskaya (2010), neste trabalho também foi utilizada a curva ROC para confirmar a precisão preditiva dos modelos de pontuação estimados a partir da área delimitada, não tendo a pretensão de determinar valores de *cutoff*. A área abaixo da curva ROC do modelo logístico tradicional (Tradíc_D ROC) é de 0,7106, enquanto a área para o modelo multinível (Mult_D ROC) é de 0,8342 (Figura 1).

Figura 1. Curva ROC entre os modelos multinível e o modelo logístico tradicional.



Fonte: Stata.

O teste para igualdade de áreas sob as curvas ROC foi elaborado a partir de código em Stata desenvolvido por Cleves (2002), a partir de algoritmo sugerido por DeLong, DeLong e Clarke-Pearson (1988). O teste qui-quadrado indica que as áreas são diferentes, ao nível de significância

de 0,000, o que sugere que há diferença significativa entre as áreas e permite que se conclua, novamente, que existe superioridade do modelo logístico multinível sobre o modelo logístico tradicional.

4.5 Análise de Sensibilidade

Como análise de sensibilidade, trocou-se a variável dependente de (a) *default*, que considera atrasos a partir de 180 dias (159 indivíduos, correspondente a 3,11% da amostra) para (b) atrasados, que considera atrasos a partir de 1 dia (1.366 indivíduos, correspondente a 26,72% da amostra).

4.5.1 Modelo Multinível

O efeito aleatório do modelo nulo (1º passo da análise, cuja tabela foi suprimida do presente estudo) para a variável dependente atrasado é 0,3350, indicando que 9,24% da variância provém do nível agência. O modelo multinível se mostrou preferível ao modelo tradicional (significância 0,0000 do teste LR).

A Tabela 9 apresenta o modelo multinível final (segundo e último passo da análise do presente artigo) para a variável dependente atrasado:

Tabela 9. Modelo multinível final (variável dependente atrasado)

Variável	Todas as variáveis		Apenas variáveis significativas	
	Estimativa	Erro padrão (p-valor)	Estimativa	Erro padrão (p-valor)
Idade	-0,0084	0,0043 (0,052)	-0,0095	0,0041 (0,020)
Tempo de relacionamento	-0,0406	0,0080 (0,000)	-0,0407	0,0080 (0,000)
Fundamental	0,3392	0,1136 (0,003)	0,3390	0,1129 (0,003)
Superior	0,0161	0,1308 (0,902)		
Financ./renda	0,5233	0,0778 (0,000)	0,5258	0,0770 (0,000)
Proventista	-1,0721	0,1222 (0,000)	-1,0728	0,1221 (0,000)
Sexo	0,0046	0,0687 (0,947)		
Estado civil	-0,0622	0,0810 (0,442)		
Constante	-2,4387	0,3434 (0,000)	-2,4251	0,3411 (0,000)
Efeito Aleatório	0,1944	0,0617	0,1960	0,0618
Teste LR	29,19	(0,0000)	29,87	(0,0000)

Fonte: elaborada pelos autores.

O modelo multinível final para a variável dependente atrasado exclui as mesmas variáveis do modelo para variável dependente *default* (Tabela 7), porém exclui adicionalmente a variável superior, por não apresentar significância. Todos os resultados apresentados nessa seção 4.5.1 corroboram os resultados da seção 4.2.

4.5.2 Modelo Logístico Tradicional

A Tabela 10 apresenta os coeficientes do modelo logístico tradicional para a variável dependente atrasado, como complemento à análise de sensibilidade.

Tabela 10. Modelo logístico tradicional (variável dependente atrasado)

Variável	Estimativa	Erro padrão (p-valor)
Idade	-0,0086	0,0040 (0,029)
Tempo de relacionamento	-0,0404	0,0077 (0,000)
Fundamental	0,3404	0,1086 (0,002)
Financ./renda	0,5539	0,0735 (0,000)
Proventista	-1,0715	0,1193 (0,000)
Constante	-2,4590	0,3277 (0,000)

Fonte: elaborada pelos autores.

O *log likelihood* do modelo logístico tradicional (-2.809,7434) é menor do modelo multinível final (-2.794,8075). A curva ROC do modelo logístico tradicional tem área (0,6991) inferior à do modelo multinível (0,7664), sendo estatisticamente diferentes (0,000). Portanto, também para a variável dependente atrasado, conclui-se pela superioridade do modelo logístico multinível.

5 Considerações Finais

Estudou-se uma amostra de 5.113 contratos de financiamento imobiliário de pessoas físicas sob o Programa Minha Casa Minha Vida, de clientes de uma instituição financeira de abrangência nacional, em 30 de setembro de 2019.

Entende-se que seja um dos primeiros estudos brasileiros sobre o tema. Outros autores estudam majoritariamente crédito de pessoa jurídica, com poucos trabalhos de crédito de pessoa física, principalmente quando se trata de financiamento imobiliário. Neste último grupo, não se encontrou trabalho no Brasil, com abrangência nacional apenas o trabalho de Locatelli et al. (2015), com operações em Minas Gerais, Espírito Santo e Rio de Janeiro. Para relativizar a variável financiamento, utilizou-se, e ao que tudo indica pela primeira vez, a variável financiamento dividido pela renda. A variável proventista (clientes que recebem salário na mesma instituição credora) também foi utilizada pela primeira vez.

Outra inovação do presente trabalho é a utilização da técnica multinível, em linha com as características da base de dados, que apresenta indivíduos oriundos de 1.448 agências espalhadas pelo Brasil, configurando o caráter hierárquico da amostra. A não homogeneidade da distribuição dos indivíduos pelas agências faz com que, em existindo alguma influência da agência na variabilidade da probabilidade de *default*, o não tratamento correto do aninhamento pode interferir no resultado por inflacionar a influência do nível agência pelo número de clientes das respectivas agências.

Neste sentido, o objetivo do trabalho é identificar (i) os fatores

determinantes da inadimplência (*default*) de operações de financiamento imobiliário e (ii) a influência do contexto do indivíduo (agência bancária).

Os testes indicam a superioridade do modelo logístico multinível sobre o modelo logístico tradicional (significância do teste LR de 0,000). Sendo assim, pode-se concluir que indivíduos com as mesmas características, porém, de agências distintas têm probabilidades diferentes de inadimplir.

O correto tratamento dos dados, por meio da modelagem hierárquica, produziu um modelo cuja área da curva ROC, amplamente utilizada em modelos de gestão de risco de crédito e de probabilidade de *default* (Fávero & Belfiore, 2017), mostrou-se superior em relação ao modelo logístico tradicional (área de 0,8356 para a modelagem hierárquica em comparação a área de 0,7052 para o modelo tradicional, com diferença estatística significativa de 0,000).

O modelo hierárquico logístico final identificou as seguintes variáveis com influência positiva na inadimplência: (a) relação financiamento / renda e (b) instrução do indivíduo ser até o ensino fundamental. Por outro lado, as variáveis que influenciam negativamente a inadimplência são: (a) idade, (b) tempo de relacionamento com o banco, (c) ser proventista e (d) instrução do indivíduo ser superior (incompleto, completo ou pós-graduação, mestrado e doutorado; menor probabilidade de inadimplência em relação aos indivíduos com ensino médio). As variáveis sexo (masculino e feminino) e estado civil (solteiro e demais) não se revelaram significativas na amostra analisada.

Como análise de sensibilidade, ao invés de analisar a inadimplência (caracterizada pelo atraso superior a 180 dias e que contava com 3,11% da amostra), utilizou-se como variável dependente o atraso (caracterizado pelo atraso superior a 1 dia, representando 26,72% da amostra). Os resultados foram mantidos, tanto em termos da superioridade do modelo hierárquico sobre o modelo tradicional, quanto pelas variáveis significativas (exceto pelo fato de a variável superior não ser significativa para atrasado).

Como limitação do trabalho, não se teve acesso às características das agências não sendo, portanto, possível identificar quais delas podem ter relação com a inadimplência e até quais delas moderam o efeito das características dos indivíduos na probabilidade de inadimplir. Futuros trabalhos podem explorar essas relações bem como expandir a análise para um terceiro nível, com o agrupamento das agências (por critério geográfico – capital ou interior, estado etc. – ou por alguma característica tal como tamanho).

Considerando os resultados obtidos, recomenda-se que as análises de crédito para concessão de financiamento imobiliários sejam conduzidas por meio de modelos multiníveis que considerem o aninhamento dos indivíduos nas suas respectivas agências bancárias.

Referências

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 Finanças*, 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 binomial 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/DE290R257.
- 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) (3^o ed.). College Station: Stata Press.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: applications and data analysis methods* (2^o 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.

Tamari, M. (1966). Financial ratios as a means of forecasting bankruptcy. *Management International Review*, 6(4), 15-21.

Topa, L. E. (1979). *La gestión de créditos*. Barcelona: Ediciones de Palma.

Yıldırım, A. (2020). The effect of relationship banking on firm efficiency and default risk. *Journal of Corporate Finance*, 65, 101500. doi: 10.2139/ssrn.2849466.

Zhou, J., Wang, C., Ren, F., & Chen, G. (2021) Inferring multi-stage risk for online consumer credit services: na integrated scheme using data augmentation and model enhancement. *Decision Support Systems*, Article 113611. In press, corrected proof. Disponível online: 7 Jun 2021. doi: 10.1016/j.dss.2021.113611.