

Analysis of default risk in credit card use

Análise do risco de inadimplência na utilização de cartões de crédito

DOI:10.34117/bjdv7n6-576

Recebimento dos originais: 07/05/2021

Aceitação para publicação: 24/06/2021

Leonardo Jantsch

Master's Degree in Accounting, UNISINOS

Institution: PortoCred Financeira

Address: Av. Dr. Nilo Peçanha, 2900, 90020-007 Porto Alegre, RS (Brazil)

E-mail: ljantsch@gmail.com

João Luiz Becker

PhD in Management Science, UCLA

Institution: Sao Paulo School of Business Administration, FGV EAESP

Address: Av. 9 de julho, 2029, Bela Vista, 01313-902 São Paulo, SP (Brazil)

E-mail: joao.becker@fgv.br

Pedro Solana-González

PhD in Industrial Engineering, UC

Institution: University of Cantabria (UC)

Address: Avda. de los Castros, 56, 39005 Santander (Spain)

E-mail: pedro.solana@unican.es

Adolfo Alberto Vanti

PhD in Business Administration, UDE

CNPq Researcher / Universal Project,

Address: Dr. Barbosa Gonçalves, 777, Porto Alegre, RS (Brazil)

E-mail: avanti@pq.cnpq.br

ABSTRACT

This paper analyzes the risk of default in the use of credit cards generating probabilities of delay in payment with different variables such as age, gender, credit limit and annual income. The behavior of debtors who use credit cards is studied identifying changes in states of delay of risk levels. A multi-state model of Markov was used to perform the analysis. The study was applied to credit card usage records of individuals in 121 commercial and financial institutions. This research identifies the patterns of use by credit card customers and provides valuable inputs to help financial institutions understand the phenomenon of default risk.

Keywords: Risk analysis, Default risk, Credit card, Markov chains, Financial institutions, Decision-making process.

RESUMO

Este trabalho analisa o risco de inadimplência na utilização de cartões de crédito gerando probabilidades de atraso no pagamento com diferentes variáveis tais como idade, sexo, limite de crédito e rendimento anual. O comportamento dos devedores que utilizam

cartões de crédito é estudado identificando alterações nos estados de atraso dos níveis de risco. Foi utilizado um modelo multiestado de Markov para realizar a análise. O estudo foi aplicado aos registros de utilização de cartões de crédito de indivíduos em 121 instituições comerciais e financeiras. Este estudo identifica os padrões de utilização pelos clientes de cartões de crédito e fornece dados valiosos para ajudar as instituições financeiras a compreender o fenômeno do risco de inadimplência.

Palavras-chave: Análise de risco, Risco de inadimplência, Cartão de crédito, Cadeias de Markov, Instituições financeiras, Processo de tomada de decisão.

1 INTRODUCTION

Consumer credit is a significant part of banking system and credit cards continue to be a dominant and increasingly more popular payment method. A downside for credit card issuers lies, however, in the users' increasing tendency to default on their payments. Aggressive marketing strategies can encourage credit card use beyond payment capacity, thus increasing the bearer's credit risk and resulting in defaults and losses that might have not been properly anticipated (OH; JOHNSTON, 2014).

In Ghodselahi (2011) opinion, the main feature of the banking industry is to properly manage capital and risk for making profits. Success in this industry is directly related to its ability to control and manage related risks. The author states that the accuracy of the credit classification via decision models that discriminate between good and bad borrowers is fundamental for the profitability of the financial institution. This fact heavily influences the cost reduction of credit analysis, the speed of deliberations, the quality of the portfolio and the decrease in default risk.

In addition, consumers' behavioral aspects also affect the levels of demand and composition of the portfolios. In a study of the behavior of US credit card users, Shefrin and Nicols (2014) found that 27% of card holders declare themselves as adherents to the minimum payment of their invoices and yet have a low level of control over their personal finances. Furthermore, only 40% of them attach great importance to being in control of their finances and 44% of that group report low confidence in being able to manage their finances using online technology.

In this way, the rationale behind Getter's (2008) assertion is established in which revolving credit, including the credit card system, entails the greatest risk among all types of bank loans, which consequently increases operating costs, thus generating default risks. Considering the representativeness the credit card system has in the economic context and the related credit risk, the relevance of this study is established by defining below the

research hypothesis and its corresponding objective. Other studies such as Wang, Hu and Li (2017), analyzed the consumption behavior of credit card holders and uncovered an interesting pattern: the features (dimensions of consumption behavior) cluster in different groups and propose a feature selection algorithm. They considered the credit card default feature, selection prediction and detection patterns. Garcia (2017) points out that, especially given the international financial crisis, credit risk has become one of the main challenges of risk management.

The objective of this research is to analyze and predict the behavior of debtors who use credit cards, identifying changes in the states of delay and risk levels according to the characteristics and use patterns of customers, predicting the probability of default through an estimation model of Markov. This research measures the probability of delays in payments of individuals in 121 commercial institutions belonging to a business group with an overall loan portfolio of approximately US \$192 million dollars during the period of interest. Therefore, the risk of non-payment will be analyzed, measuring the probability of delayed payments in a real application of 8,851 customers who have used the credit card on an ongoing basis.

This analysis enhances the decision-making process in analyzing credit risk in credit card loans for individuals. Thus, in order to predict the default risk for clients, it was used a classification model based on multiple explanatory variables. Credit card holders were classified as either defaulters or non-defaulters according to intensity models and transition probabilities.

This article is structured in sections that elaborate on its contextualization, relevance and objectives. It thus advances the state of the art by proposing a new approach to credit risk analysis, then discusses the methodology and unveils the experimental results. Finally, some concluding considerations are presented.

2 CREDIT RISK AND CREDIT CARD MODEL

In Walker's (2013) conception, the risk and its origin are related to the notion of insurance and the accounting of losses. The author argues that in an investment-oriented business structure, risk is a business reality that is measured to convey the possibility of losses and gains. In this way, he concludes that measuring the changes and understanding when they will happen are the basis for risk management in business which can be treated and analysed with different measurement and risk management models as shown in

Machado, Oliveira and Leite (2021) applying ABNT ISO 31000 to the development process of internal models of credit risk in financial institutions.

Thomas, Ho and Scherer (2001) address the assessment of consumer credit risk through the use of dynamic modelling by means of behavioral, customer and profit scoring. The authors point out that when profitability is considered, it is necessary to use recent consumer behavior in order to estimate subsequent performance over a future time frame. For defaults, these calculations also contribute to estimating how much a creditor needs to establish a provision for doubtful accounts in order to cover these expected losses, and that is also referred to as ‘the debt-provisioning problem’. Credit card frauds are widely reported, but few applications to reduce or prevent such occurrences are known. In this sense, Carneiro, Figueira and Costa (2017) underline the importance of combining automatic and manual use of machine learning in order to reduce this type of risk, as also was studied with five of these techniques to predict credit card defaults, in which decision tree performed the best (TENG; LEE, 2019).

Bellotti and Crook (2013) explored the matter further in terms of macroeconomic variables. In their study they identify that the interest rate is positively related to credit defaults since they increase the value of debt repayment and the level of debt. Similarly, the authors detect that the unemployment rate affects individuals in the same way: debtors who have become or remain unemployed find it more difficult to pay the debt. With regard to behavioral scoring, Caouette, Altman, Narayanan and Nimmo (2008) assume that in these models there is a metric that separates the good credits from the bad ones through different distributions, which are based on the real credit experience of the company. According to the research authors, bad accounts are usually those with at least three repeated missed payments whereas good accounts refer to those that do not present delays at this level.

According to Thomas, Ho and Scherer (2001), consumer behavior models based on Markov chains represent an alternative approach to *behavioral scoring* with extension to *profit scoring*. However, there are few commercial systems that adopt this approach. For Anton and Rorres (2000, p. 286), a Markov chain is a dynamic system whose state vectors across time intervals are represented as probability vectors and are related by an equation of the form $x_{t+1} = Px_t$, where $P = [p_{ij}]$ is a stochastic matrix whose element p_{ij} is the probability that the system will be in state i at time $t + 1$ if it is in state j at time t . The matrix P is called the transition matrix for the system. For the authors the Markov chain model considers

that there is a simple stochastic dynamic model which allows designing the future behavior of each client. Markov chain models may be used to build *behavioral* or *profit scoring* systems and mapping the dynamics of the default state of a population.

For the application of Markov chains, the model to be applied to the sample of clients must differentiate the states of credit card use, including the default state. The transition probabilities are obtained from previous data where $n(i)$ is the total number of months that customers are in state i ($i = 0, 1, 2, 3, 4$) and $n(i, j)$ is the number of times clients go from state i to state j . The maximum likelihood estimate of the transition probability p_{ij} is $n(i, j) / n(i)$ (THOMAS; HO; SCHERER, 2001).

According to Bendle and Horne (2014), in the context of credit card usage, it is possible for the card holder to obtain a loan (with or without interest) for payment of the purchase. The credit card allows purchases to be paid without interest after a period of up to 45 days or installment in parts in the "rotating credit". This outstanding balance is subject to the accrued interest, all fees and interest paid to the card issuer (MUKHERJEE; KAWDE, 2014).

It is also worth mentioning the studies (RÉGIS; ARTES, 2016; LEOW; CROOK, 2014; BELLOTTI; CROOK, 2013; SILVA; VIEIRA; FAIA, 2012; SO; THOMAS, 2011; SANTOS; FAMÁ, 2007) that consider variables related to age, income, employment, residence, credit limit, use of revolving credit, interest rate and unemployment rate. These variables are considered as reference for defining the present study. Moreover, according to Leow and Crook (2014), they are a reference for segmentation by means of the branch of the activity that considers the type of occupation and employment of individuals in the sample. The same authors, structured the forecast period considering the probabilities of transition at the end of the 12th period from the states established in the 6th period. Leow and Crook (2016a) analyzed the exposure at default at the level of the obligor by estimating the outstanding balance of an account, not only at the time of default, but at any time over the entire loan period. The same authors, using a large portfolio of credit card loans, investigated the stability of the parameter estimates of discrete survival models, especially since the start of the credit crisis of 2008 in the UK economy. Two survival models were developed for accounts that were accepted before and since the crisis (LEOW; CROOK, 2016b).

Hon and Bellotti (2016) studied credit card balance in retail finance. Other authors like Xiao, Crook and Andreeva (2017) proposed to incorporate least squares support

vector machine technique into a two-stage modelling framework to predict recovery rates of credit cards from a UK retail bank.

Finally, it should be noted that Saravanan and Babu (2017) carried out a study about the usage and fraud in credit card online transactions through data mining techniques. In parallel, Mankame, Nikam and Gurav (2017) also analyzed this kind of approach that involves data mining detection of fraud in different transactions. In addition, West and Bhattacharya (2016) researched intelligent financial fraud developing a significant literature review in financial fraud detection and anomaly detection with different techniques.

3 METHODOLOGY

The structuring of the analysis model involved several significantly operational steps that were grouped into some broader ones related to business understanding, Markov chains and analysis model. In this way, it was possible to analyze and generate results about the default risk on credit card use of 8,851 users.

According to Smith (2015), comparing and observing values with the purpose of measuring the significance of the sample elements and the degree to which they relate to the studied event can be adapted to this work. The goal here was to discover factors related to customer characteristics and their credit card usage patterns that contribute to the phenomenon of credit card default risk.

Thus, the present work revolves around the business understanding by structuring Markov chains and generating an analysis model considering 18 main variables. An integrated view of the technique and the application field is unveiled, thus amplifying the context developed mainly by Leow and Crook (2014) that estimate the probability of delinquency and default for a sample of credit card loans using intensity models. The aim is to develop analyses that meet the methodological requirements and produce useful information for management purposes, more specifically to enhance the decision-making process.

3.1 RESEARCH CONTEXT

The understanding of the business that defined the research context corresponds to a set of 121 financial institutions that operate under the same brand in 21 Brazilian states. The performance model considers the geographic segregation with the

regionalized performance among the 121 financial institutions, thus avoiding competition among the group companies.

These financial institutions, belonging to the same business group, are card issuers with their own flag, together with the MasterCard and Visa flags. The number of cards issued was 2.1 million, of which 24.7% were active cards. The total loan portfolio in that period was approximately US \$192 million.

The initial sample included 119,839 customers having the largest balance in the card portfolio among the other institutions in the group. For bank secrecy purposes, the extracted data was treated in Excel® and the card numbers and customer identification data were totally omitted. This was done in consideration of the Complementary Law 105/2001 (LCP-105, 2001), by which it is the financial institutions' responsibility to keep their clients' data safe.

In order to have a consistent sample based on customers using credit cards as a regular means of payment, the initial sample was reduced to 8,851 customers considering the following refinement parameters and excluding the following customers:

1. Those who did not have a consistent portfolio over the 24 months of the sample.
2. Those who had an average accounting portfolio of less than R\$ 100.00 over the entire sample period.
3. Those who started the sampling period either using revolving credit or in the default state.

Based on the indicators obtained in the literature discussed, considering the variables in Table 1 and by accessing the databases of the companies supporting this research, it was possible to identify 18 variables that correspond to the data collection phase.

3.2 DATA COLLECTION

For data collection it was defined and used the variables presented below, together with their respective attributes.

i. Client: The client identifier, i.e., a sequential code assigned to each individual in the sample. The total sample is made up of 119,839 customers, which is then reduced to 8,851 customers after applying some filtering parameters.

ii. Date of birth: The client's date of birth. For the analysis, the sample was segmented into 7 groups, considering 10-year periods to define the client age ranges: [18 to 27 years], [28 to 37], [38 to 47], [48 to 57], [58 to 67], [68 to 77] and over 77 years.

iii. Date of entry into the National Financial System: Start date of the client's participation in the national financial system. This date was considered to calculate the time of participation in the financial market. For the analysis, the sample was segmented into 7 groups, considering the following distribution among clients: [0 to 3 years of participation], [4 to 6 years], [7 to 9 years], [10 to 12 years], [13 to 15 years], [16 to 18 years] and over 18 years.

iv. Active branch: Branch of the client's economic activity. Its values are: agropastoral, autonomous, commercial, unemployed, industrial, and other services or unidentified ones.

v. Type of residence: Type of residence of the client. Its values are: own, rented, relative and others or not disclosed.

vi. Start date: Start date of the client's professional activities.

vii. Marital status: Marital status of the client. Its values are: single, married, common-law union, widowed, divorced and others.

viii. Dependents: Number of customer dependents. For this attribute, the sample was segregated into with or without dependents.

ix. Gender: Segregated into male or female.

x. Source of income: Defines the source of the customer's income. Its values are: salary, agricultural business, pro-labore, commissions, spouse and others.

xi. Total credit limit: Total credit limit granted for the cards issued to the customer. For the analysis, the sample was segmented into 7 groups, considering the following cumulative frequency of the clients' number in each range: [from 0% to 14%], [14.01% to 29%], [29.01% to 42%], [42.01% to 57%], [57.01% to 74%], [74.01% to 89%] and [from 89.01% to 100%].

xii. Annual income: The client's annual income. For the analysis, the sample was also segmented into 7 groups, with the following cumulative frequency of the clients' number in each range: [from 0% to 14%], [14.01% to 29%], [29.01% to 44%], [44.01% to 59%], [59.01% to 74%], [74.01% to 89%] and [from 89.01% to 100%].

xiii. Inhibition date: Being the 61st day after expiration of the unpaid invoice.

xiv. Monthly billing balance: Cash purchases, installment purchases, withdrawals, fees and charges that will be billed or already included in an invoice to the bearer and have an expected short-term settlement date, usually up to 30 days.

xv. Monthly balance of reserved stockholders: Balance of installments for future posting, originated by purchases financed by the establishment, not updated by the

issuer's interest and that have a scheduled settlement date greater than 30 days. In this modality, the receipt by the issuer and the payment to the establishment is done in installments.

xvi. Monthly revolving and withdrawal balance: Represents the short-term interest debt, stipulated with average maximum maturity of up to 30 days. It is important to emphasize that the existence of a rotary is the first step in the characterization of default.

xvii. Monthly balance of divided interest: Balance of installments for future posting originated by purchases financed by the issuer. In this modality, the receipt by the issuer is paid in installments and the payment to the store or establishment is paid in full 30 days on average after the transaction date.

xviii. Monthly loss balance: Balance of the portfolio with losses, considered as a delay of more than 360 days.

The balance information was used in this study to assign the following transition states:

1 - With portfolio: In this state, the customer uses the credit card and is up to date with his/her payments, not presenting use of the rotary;

2 - With rotary: In this state, the customer uses the credit card and keeps his/her payments up to date; however, he/she has not made full payment of the invoice and resorts to using the revolving credit. Since the accounting system treats residual values of payments as a rotary, the cases in which the value used was less than R\$ 50.00 were disregarded as a use of rotary. In this way, 22,281 occurrences of actual use of the rotary were considered.

3 - Overdue for more than 31 days: In this state, the customer has used the credit card and has his payments overdue for more than 31 days, he has already two bills in arrears.

4 - Arrears exceeding 60 days: In this state, the customer is considered to be in default and his card is canceled by the financial institution. After 60 days, the third invoice in arrears is characterized, which for the credit card context means default. This state is also considered as the absorbing state since reaching at this state causes the client to remain in it indefinitely.

By means of the four aforesaid states, the transition probability calculations and the ensuing analyses are carried out, considering the transition probabilities of the sample in the general context and, in turn, the segregation of this by customer profile. Thus, the variables used in the analysis are summarized in Table 1 below.

Table 1: Variables used in the study

Client	Marital Status	Inhibition date
Date of Birth	Dependents	Monthly billing balance
Date of Entry into the National Financial System	Gender	Monthly balance of reserved stockholders
Active Branch	Source of Income	Monthly revolving and withdrawal balance
Type of residence	Total Credit Limit	Monthly balance of interest parceled
Start Date	Annual Income	Monthly loss balance

Source: Research data

4 RESEARCH ANALYSIS AND RESULTS

This section aims to present the analyses and results of the previously described scenario using the proposed methodology. Considering the state transition probabilities and the number of accounts in the interim periods, we have the distribution presented in Table 2.

Table 2: Accounts and states for periods 1, 6, 12 and 24

Situation/Period	With portfolio	With rotary	Overdue for more than 31 days	In default	Total number of observations
Period 1	8,851	-	-	-	8,851
Period 6	8,103	611	41	96	8,851
Period 12	7,785	760	45	261	8,851
Period 24	7,252	898	68	633	8,851

Source: Research data

Table 3 shows the percentage values of the scenarios. It can be observed that the distribution gradually becomes more significant for states 2 (use of the rotary) and 4 (default).

Table 3: Account distribution by state for periods 1, 6, 12 and 24

Situation/Period	With portfolio	With rotary	Overdue for more than 31 days	In default
Period 1	100.0%	0.0%	0.0%	0.0%
Period 6	91.5%	6.9%	0.5%	1.1%
Period 12	88.0%	8.6%	0.5%	2.9%
Period 24	81.9%	10.1%	0.8%	7.2%

Source: Research data

As shown in Table 3, after the 24th period, 81.9% of the portfolio was in state 1 (up-to-date payments), 10.1% in state 2 (use of the rotary), 0.8% in state 3 (delayed payment over 31 days), and 7.2% of the portfolio was in the default state.

Regarding the sample scenario, the present work develops a model that individually considers the transition probability matrices among the account states. In a second attempt, we used segmentation by covariables, thus obtaining a model by customer profile.

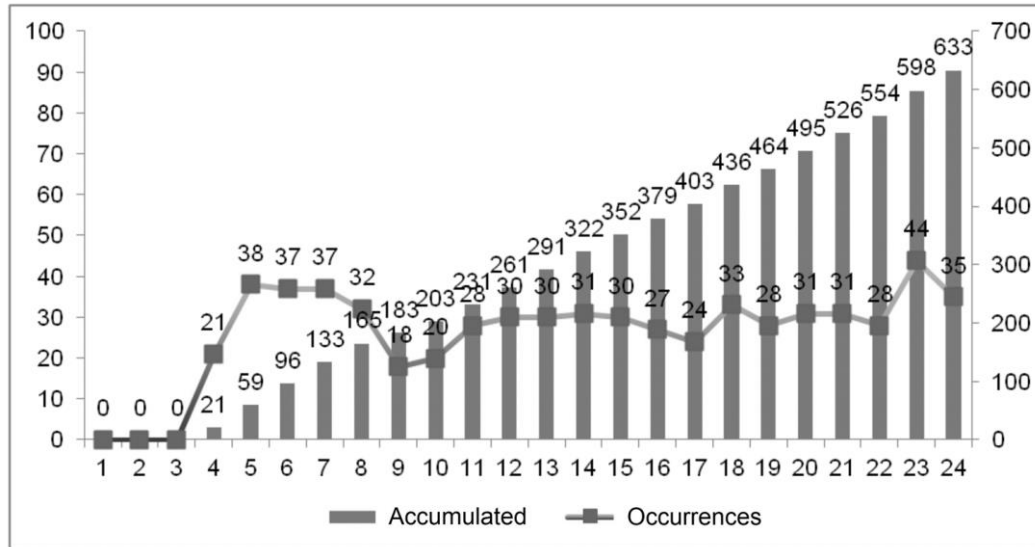
The goal of this analysis was to evaluate whether different characteristics of the individuals can be used to generate predictions that consider both the own business knowledge and the use of Markov chains. Transition matrices were obtained by individually considering, for each account, the transition occurrences' number of each state in the 24 sample periods.

Two test sets were considered in the validation model. The first set was generated from segregating the portfolio into two groups with different numbers of individuals, termed in this work as *transversal validation* with data extraction in a specific period. The second set was generated from the segregation of the portfolio considering the observations of twelve consecutive months, referred to in this work as *longitudinal validation*, thus obtaining two data sets for the 24 months of the sample.

The general model includes the 8,851 individuals, considering the state which the account is in for each time period. For cross-validation purposes, the sample was segregated considering an 80% - 20% ratio, thus obtaining groups with 7,080 and 1,771 individuals, respectively. In the longitudinal validation, in which segregation occurs for the periods, the sample considered 8,851 individuals and their performance during the first 12 months of the sample and it was compared with the performance in the subsequent 12 months.

Bearing in mind that the main objective of this study was the evaluation and characterization of credit risk, we have in state 4 (default) the highest risk to be ascribed. During the data analysis phase, it was confirmed that 633 accounts ended the period in this state, representing 7.15% of the portfolio. The evolution of defaulted accounts over time and the cumulative amounts are depicted in Figure 1 below.

Figure 1: Evolution of defaulted accounts by sample period



Source: Research data

As shown in Figure 1, the portfolio does not exhibit defaulted accounts in periods 1, 2, and 3. This is not surprising as this was one of the criteria for the selection of individuals in the sample. From period 4 onwards there is a varying number of defaulted accounts, as shown in the chart. The main default risk probabilities are given in the following subsections.

4.1 ANALYSIS WITH RESPECT TO THE TOTAL SAMPLE

For the analyses, the 8,851 individuals in the sample and the transition probabilities between the identified states for the periods 6, 12, 18 and 24 months were considered.

The obtained probabilities show that, after 24 months, 83.05% of the individuals who are in state 1 (up-to-date payments) will remain in this state. This shows a strong trend of non-use of revolving credit. For state 2 of use of the rotary after 24 months, 8.66% of the accounts will be in this state. State 3 (delayed payment over 31 days) is not significant, but default status 4 indicates that 7.78% of individuals who started at state 1 will be at state 4 after 24 months.

Regarding state 2 (use of the rotary), the probability of migration to state 1 (up-to-date payments) is 75.64%, which shows a trend of recovery of the individual and abandoning use of the rotary. On the other hand, 7.88% of the individuals will continue in this state and use the rotary after 24 months. It is interesting to emphasize this group of individuals since they are subject to high-interest rates on the rotary instead of seeking

a credit solution that better suits their profile. Furthermore the probability of becoming a defaulter after 24 months is 16% for this group, which is a high estimate for default of this type of credit.

With respect to state 3 (delayed payment over 31 days), the last step before default, the results show that 23.83% of the individuals will be able to recover and will become deferred again. However, 73.54% will move to state 4 of final default.

4.2 AGE-BASED ANALYSIS

In the calculation of the transition probability matrices for period 6 by grouping the portfolio according to age, the probability of remaining in state 1 (up-to-date payments) increases as the customer age increases. It ranges from 86.80% for group 1 to 94.45% and 93.80% for groups 5 and 6, respectively. It is evident that up to group 6 the probability steadily increases, and in the last group, corresponding to the highest age range in the sample scenario, there is a small reduction in the percentage.

A further observation is that younger age groups have a higher probability of using revolving credit, as can be seen from the transition probability from state 1 (up-to-date payments) to state 2 (use of the rotary). The probability of default risk starts at 10.78% and steadily decreases until it reaches 5.70% in the last group.

4.3 ANALYSIS ON THE TIME OF PARTICIPATION IN THE FINANCIAL SYSTEM

In the analysis from the perspective of participation time in the financial system for period 6, it was observed that in relation to permanence in state 1 (up-to-date payments), the probability increases significantly, varying from 84.11% to 95.64% as the time of experience in this market goes up.

Once again, the younger, less experienced groups have a higher probability of using revolving credit, as can be observed in the probability of transition from state 1 (up-to-date payments) to state 2 (use of the rotary). The probability of default risk starts at 12.59% and steadily decreases until it reaches 4.09% in the last group.

4.4 ANALYSIS BASED ON ACTIVITY BRANCH

Considering the rushed calculations for the accounts that were originally in state 1 (up-to-date payments), the highest probability of remaining in this state corresponds to the agropastoral industry, quantified as a probability of 94.27%. The lowest probability of remaining in state 1 corresponds to the unemployed group in the order of 86.92%.

For the analysis of the highest probability of default, for the accounts that migrated to state 4, the highest probability corresponds to the Others and Commercial branches, at 13.4% and 13.5%, respectively. On the other hand, the lowest probability of migration to state 4 corresponds to the agropastoral group at about 0.56%.

4.5 ANALYSIS ON THE TYPE OF RESIDENCE

Analyzing the state transition probabilities in relation to the type of residence, it can be observed that the probability of maintaining payments to date is higher when the individual owns a residence.

Considering state 1 (up-to-date payments), the probability of remaining in this state is 91.78% for individuals who own their residence. Therefore, it can be concluded that the ownership of assets constitutes a guarantee of payment compliance. In the same way, it translates into a better use of the credit card in terms of day-to-day payments and maintenance of credit card statements.

4.6 ANALYSIS BASED ON MARITAL STATUS

Considering marital status, the singles group is perceived as less likely (85.81%) to remain in state 1 of up-to-date payments. On the other hand, the group of accounts composed of widowers/widows presents the highest probability of remaining in state 1, about 93.12%. Additionally, with 11.27% the singles translate into the group most likely to use the rotary.

For the analysis of the highest probability of default, for the accounts that migrated to state 4, the highest probability is found in the group of singles, 2.04%. In contrast, the lowest probability of moving to state 4 is for the group of widowed individuals, at about 0.42%.

4.7 ANALYSIS BASED ON THE EXISTENCE OF DEPENDENTS

In the analysis of the probabilities considering the existence of dependents, it can observe that in state of payment 1, the highest probability of remaining in this state is

associated with the profile of individuals with dependents, 93.37%. The probability of moving to state 2 (use of rotary) and of remaining in this state, is higher for the group of individuals without dependents, considering the percentages of 9.84% and 12.75%, respectively.

For the analysis of the highest probabilities of default, considering the accounts that migrated to state 4, the highest probability is related to the group of individuals with no dependents, 1.44%. The lowest probability of migration to state 4, ascribed to the group with dependents, it is measured in the order of 0.32%.

4.8 GENDER-BASED ANALYSIS

With regard to transition probabilities by gender, it is observed that there are no significant differences between the transition probabilities in the two groups. The transition probability values are very similar in the two groups, so it can be inferred that this segmentation is not significant for a credit risk analysis process.

In terms of percentages, the results show that in the state of payment 1 the probability of remaining in this state is 89.65% for men and 89.04% for women. From state 3 of default, the probability of migration to state 4 is 72.27% for men and 70.10% for women.

4.9 ANALYSIS S ON THE APPROVED CREDIT LIMIT

For the present study, considering the analysis of the probability of permanence in state 1 (up-to-date payments), group 1 with the lowest approved credit limit shows the lowest probability, in the order of 79.59%, this probability reaching 93.97% for the highest credit limit group. Therefore, it is observed that the probability of remaining in state 1 increases as the limit granted also increases.

Regarding migration to state 2 (use of the rotary), the same effect occurs since for group 1 a probability of 14.44% was identified and, for group 7, a probability of 5.77%. Therefore, the probability tends to decrease as the approved credit limit increases.

4.10 ANALYSIS BASED ON ANNUAL INCOME

Considering the earned annual income as an analysis variable, it is identified that the greater the individual's income, the greater the probability of remaining in state 1 (up-to-date payments). This increasing probability from the lowest income group to the highest income group goes from 85.92% to 92.50%.

Conversely, we have the transition probability between state 1 (up-to-date payments) and state 2 of the rotary users. The probability fluctuates between 10.84% and 6.84%, with the highest probability of use being identified in the lower income group.

4.11 ANALYSIS ON INCOME SOURCE

From the income source perspective, the salary group is the group with the highest probability of migrating to state 2 (use of the rotary) from state 1 (up-to-date payments). This probability is estimated at 9.75%. Likewise, this is the group with the lowest probability, in the order of 88.22%, of remaining in state 1 of up-to-date payments.

On the other hand, the group of individuals involved in the agricultural industry is the one with the highest probability of remaining in state 1 of up-to-date payments, which is estimated at 93.93%.

4.12 CREDIT CARD DEFAULT PROFILE

In this study, it has sought to characterize credit risk in the use of credit cards by outlining the profile of the high-risk individuals. To this end, starting from the probabilities of state 1 (up-to-date payments), the lowest probabilities of remaining in this state were obtained in each category. For state 2 (use of the rotary) coming from state 1, the highest probabilities of using this type of credit in each category were obtained.

The highest credit risk profile from state 1 corresponds to the young individuals with little experience in the financial market and the unemployed. Non-resident homeowners, i.e., those who rent/lease, unmarried, without dependents, with a low granted credit limit and low annual income when they earned income, the declared source was their salary.

By analyzing the transition from state 2 to the state 3 (delayed payment over 31 days), the same profile trends emerge, changing only the source of the pro-labore wage income.

However, in the analysis of state 3 (delayed payment over 31 days) the biggest changes are observed. At this point, the highest probabilities of migration to state 4 (default) correspond to individuals in age groups 6 and 7, the two oldest age groups in the sample. With regard to participation in the financial system, the trend of credit risk default pertained to the least experienced individuals.

Regarding the business sector, the individuals involved in the industrial activity branch are associated with the highest probabilities of default risk, which in the previous

analysis was ascribed to the unemployed individuals. Considering the type of residence, the individuals of the sample characterized in the *others* group are those exhibiting a higher probability of migration to the payment default state. However, renters continue to have a very significant probability of default risk.

With regard to marital status, the highest probabilities are related to widowed individuals. In addition, those who declare to have no dependents as well as a salary-based income source, have a high approved credit limit and low income. These are the most significant attributes of individuals who progress towards a state of credit default.

It should be noted that most of the findings in this study are in line with conclusions obtained in previous studies. In the analysis of transition probabilities of the general model, a probability of 89% was identified for an account to start in state 1 of up-to-date payments and to remain in this state until the sixth projection period. The Leow and Crook (2014) study, for accounts that were originally in state 0 of non-delayed payments, also pointed to the existence of a very high probability, above 80%, of accounts remaining in the initial state after 6 months.

However, the probability of remaining in the original state is not repeated for the analysis based on state 2 transitions. Leow and Crook (2014) identified that the probability of being in state 2 after 60 days of delayed payment is less than the probability of being in the credit default state 3. For the authors of the manuscript, this seems to suggest that an overdue account is more likely to migrate to default state 3 than to remain in state 2. In relation to this study, this is a recurring trend considering that, once reaching state 3, there is a 69% probability that the account will transition to state 4 of credit default.

In the study by Leow and Crook (2014), for accounts that were originally in state 2, the authors identified different probabilities among different occupation types. For employees, there is a greater probability of going into default (48%) than into recovery (39%). For the self-employed or unemployed, there is a greater probability of going into recovery (from 42% to 54%) than into default (from 32% to 44%). For the research authors, these figures seem to suggest that employed individuals are more susceptible to becoming defaulters.

Although not in reverse order, the results of the study identify both the unemployed and the self-employed as having a lower probability of default, 61% and 60%, respectively. Likewise, they are more likely to recover, with probabilities of 30%

and 36%, confirming the study by Leow and Crook (2014) that employed persons are more susceptible to default.

For analyses based on the approved credit limit, So and Thomas (2011) use segregation by score for segmentation between higher and lower level of risk. Similarly, they assign credit limit bands. As a result, they obtain a matrix where they identify that the volatility of the scoring transitions decreases as the credit limit is increased. In terms of probability, they identified that 75.2% of the accounts with the highest score but with the lower limit are in the same state range after one month, whereas for the accounts with the highest score and the highest limit, 88.6% remain in the same state range after one month.

For the present study, considering the analysis of the probability of permanence in state 1, group 1 with the lowest approved credit limit is the one with the lowest probability, in the order of 79%. It is worthwhile mentioning that this probability increases as the approved credit limit also goes up, which corroborates So and Thomas (2011)'s findings.

Regarding migration to state 2 (use of the rotary), the same effect occurs, where for group 1 a 14% probability was identified. It can be observed that there is a decreasing tendency of this probability as the approved credit limit granted increases, reaching 5% for group 7.

According to Bellotti and Crook (2013), an increase in the credit limit reduces risk. Initially, this statement may be surprising, since one could argue that a high credit limit would encourage its use and, therefore, it would entail a greater default risk. However, in the first place, at least in the short term, the authors state that a high credit limit allows the borrower to make room for debt before reaching default. Therefore, in the granting of credit, the granted limit is part of an evaluation process in which the debtor's behavior is considered, with greater limits being assigned to the borrowers who are less susceptible to default.

In order to finalize the tests and analyses, the findings of the three validation processes confirm the conclusions of Thomas, Ho and Scherer (2001), for which the Markov chain model considers that the dynamics of its subsequent behavior follow the pattern mapped by the Markov chain, and that there is a simple stochastic dynamic model that allows designing the future behavior of each client.

5 CONCLUSIONS

The credit card product through the incentive to consumption and high interest rates generates for the financial institution a context of high profitability but also of exposure to risks with impact on results, due to the risk of non-payment and the consequent loss forecast.

This work studied the behavior of debtors who use credit cards throughout the period of use, identifying changes in states of delay of risk levels. With this, it was possible to predict the probability of default considering several variables and profiles by structuring a multi-state model of Markov analyses.

In this paper, we chose to analyze the risk of default in credit card use, measuring this risk through the probability of delayed payment and credit, a characterization of the "rotary". The work was applied in credit card loans to individuals in financial institutions belonging to a same business group.

This analysis for the prediction of default risk in credit card use was developed by applying a multi-state intensity model with the use of Markov chains, thus positioning this research concerning entrepreneurial reality. It was considered the generation of intensity matrix through application of the Markov model with the multivariate data structure to obtain the parameter estimates. The results were used in the construction of simulations of the clients' characterization over multiple time horizons.

This study's main contribution is that, through Markov chain application, the profile characterization of individuals with a higher risk of default emerges, which improves the decision-making process both for the business approach and the card holder approach. In this study, the following variables were used: customer, age, time of participation in the financial system, activity branch, type of residence, marital status, dependents, gender, approved credit limit, annual income and income source. With regard to macroeconomic variables, we considered the unemployment rate and the average interest rate of credit operations. As portfolio variables, it was obtained total portfolio, rotary and the inhibition date.

In this way, an extended view of the studies by Leow and Crook (2014) and Régis and Artes (2016) was obtained. For the variable element, we opted for an expanded scenario considering profile, portfolio and macroeconomic items, and for the analyses, it was decided to separately highlight the probabilities of each profile and its stages of evolution among the possible states regarding the use of the granted credit.

For future studies, use of the Markov model for default analysis is recommended, taking into account the general credit portfolio as well as other products used by the client. In this way, producing performance estimates for a product can affect or justify performance for other products. It is possible to apply for studies with sequential fraud detection for prepaid cards, using hidden Markov model divergence like the one analyzed by Robinson and Aria (2018). Other future studies may also address the interactions between institutions in different countries, as well as the influence of new laws on the refinancing of credit card debtors.

REFERENCES

ANTON, H.; RORRES, C. Elementary linear algebra: Applications version. New York, NY: John Wiley & Sons Inc., 2000.

BELLOTTI, T.; CROOK, J. Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, v. 29, n. 44, p. 563-574, 2013. doi: 10.1016/j.ijforecast.2013.04.003

BENDLE, N.; HORNE, D. Visa Inc. and the global payments industry. Ontario: Ivey Publishing, 2014.

CAOUCETTE, J.; ALTMAN, E.; NARAYANAN, P.; NIMMO, R. Management credit risk: The great challenge for global financial markets. New Jersey, NJ: John Wiley and Sons, Inc., 2008.

CARNEIRO, N.; FIGUEIRA, G.; COSTA, M. A data mining based system for credit-card fraud detection in e-tail. *Decision Support Systems*, v. 95, p. 91-101, 2017. doi: 10.1016/j.dss.2017.01.002

GARCÍA F.J.P. Credit risk: Measurement. In: *Financial risk management*. Palgrave Macmillan, Cham, 2017. doi: 10.1007/978-3-319-41366-2_9

GETTER, D. The credit card market: recent trends, funding, cost issues, and repricing practices. Washington D.C.: Congressional Research Service, Library of Congress, Order Code RL34393.1-14, 2008.

GHODSELAHI, A. A hybrid support vector machine ensemble model for credit scoring. *International Journal of Computer Applications*, v. 17, n. 5, p. 1-5, 2011. doi: 10.5120/2220-2829

HON, P.; BELLOTTI, T. Models and forecasts of credit card balance. *European Journal of Operational Research*, v. 249, n. 2, p. 498-505, 2016. doi: 10.1016/j.ejor.2014.12.014

LCP-105. Lei complementar nº 105, de 10 de Janeiro de 2001. Dispõe sobre o sigilo das operações de instituições financeiras e dá outras providências, 2001.

LEOW, M.; CROOK, J. A new mixture model for the estimation of credit card exposure at default. *European Journal of Operational Research*, v. 249, n. 2, p. 487-497, 2016a. doi: 10.1016/j.ejor.2015.10.001

LEOW, M.; CROOK, J. Intensity models and transition probabilities for credit card loan delinquencies. *European Journal of Operational Research*, v. 236, n. 2, p. 685-694, 2014. doi: 10.1016/j.ejor.2013.12.026

LEOW, M.; CROOK, J. The stability of survival model parameter estimates for predicting the probability of default: empirical evidence over the credit crisis. *European Journal of Operational Research*, v. 249, n. 2, p. 457-464, 2016b. doi: 10.1016/j.ejor.2014.09.005

MACHADO, W.; OLIVEIRA, E.; LEITE, L. Gestão de riscos no processo de modelagem do risco de crédito em instituições financeiras. *Brazilian Journal of Development*, v. 7, n. 4, p. 40174-40193, 2021. doi:10.34117/bjdv7n4-468

MANKAME, R.; NIKAM, S.; GURAV, A. Using data mining detection of fraud in transaction. *International Journal of Engineering research Online*, v. 5, p. 2, p. 152-156, 2017.

MUKHERJEE, J.; KAWDE, S. Citibank India credit cards: Strategy for profitable growth. Ontario: Ivey Publishing, 2014.

OH, J.; JOHNSTON, W. Credit lender–borrower relationship in the credit card market–implications for credit risk management strategy and relationship marketing. *International Business Review*, v. 23, n. 6, p. 1086-1095, 2014. doi: 10.1016/j.ibusrev.2014.06.010

RÉGIS, D.; ARTES, R. Using multi-state markov models to identify credit card risk. *Production*, v. 26, n. 2, p. 330-344, 2016. doi: 10.1016/j.eswa.2017.08.043

ROBINSON, W.; ARIA, A. Sequential fraud detection for prepaid cards using hidden Markov model divergence. *Expert Systems With Applications*, v. 91, p. 235-251, 2018.

SANTOS, J.; FAMÁ, R. Avaliação da aplicabilidade de um modelo de credit scoring com variáveis sistêmicas e não-sistêmicas em carteiras de crédito bancário rotativo de pessoas físicas. *Revista Contabilidade & Finanças*, v. 18, n. 44, p. 105-117, 2007. doi: 10.1590/S1519-70772007000200009

SARAVANAN, S.; BABU, S. Literature study – data mining techniques on detecting fraudulent activities in credit card. *International Journal of Emerging Research in Management & Technology*, v. 6, n. 10, p. 60-69, 2017.

SHEFRIN, H.; NICOLS, C. Credit card behavior, financial styles, and heuristics. *Journal of Business Research*, v. 67, n. 8, p. 1679-1687, 2014. doi: 10.1016/j.jbusres.2014.02.014

SILVA, L.F.; VIEIRA, V.A.; FAIA, V.S. Determinants of debt and default propensity associated with the bankruptcy of individuals. *Análise*, v. 23, n. 3, p. 207-221, 2012.

SMITH, M. *Research methods in accounting*. London: Sage, 2015.

SO, M.; THOMAS, L. Modelling the profitability of credit cards by Markov decision processes. *European Journal of Operational Research*, v. 212, n. 1, p. 123-130, 2011. doi: 10.1016/j.ejor.2011.01.023

TENG, H.; LEE, M. Estimation procedures of using five alternative machine learning methods for predicting credit card default. *Review of Pacific Basin Financial Markets and Policies*. v. 22, n. 3, 1950021, 2019. doi: 10.1142/S0219091519500218

THOMAS, L.; HO, J.; SCHERER, W. Time will tell: Behavioural scoring and the dynamics of consumer credit assessment. *IMA Journal of Management Mathematics*, v. 12, n. 1, p. 89-103, 2001. doi: 10.1093/imaman/12.1.89

WALKER, R. *Winning with risk management*. Chicago: World Scientific, 2013.

WANG, Q.; HU, Y.; LI, J. Community-based credit card default feature selection prediction. *International conference on complex networks and their applications, complex networks & their applications VI*, p. 153-165, 2017. doi: 10.1007/978-3-319-72150-7_13

WEST, J.; BHATTACHARYA, M. Intelligent financial fraud detection: A comprehensive review. *Computers & Security*, v. 57, p. 47-66, 2016. doi: 10.1016/j.cose.2015.09.005

XIAO, Y.; CROOK, J.; ANDREEVA, G. Enhancing two-stage modelling methodology for loss given default with support vector machines. *European Journal of Operational Research*, v. 263, n. 2, p. 679-689, 2017. doi: 10.1016/j.ejor.2017.05.017