NOVA IMS Information Management School

MEGI

Mestrado em Estatística e Gestão de Informação
Master Program in Statistics and Information Management

Multi-State Modeling of Retail Credit Risk

Portuguese Context

Inês Bernardino Nunes Pereira

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Estatística e Gestão de Informação

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

2017

Título: Multi-State Modeling of Retail Credit Risk

Subtítulo: Portuguese Context

Inês Bernardino Nunes Pereira

MEGI

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

MULTI-STATE MODELING OF RETAIL CREDIT RISK: PORTUGUESE CONTEXT

by

Inês Bernardino Nunes Pereira

Dissertação apresentada(o) como requisito parcial para a obtenção do grau de Mestre em Estatística e Gestão de Informação, Especialização em Análise e Gestão de Risco

Orientador/Coorientador: Jorge Bravo, PhD

Novembro 2017

AGRADECIMENTOS

A realização desta dissertação de mestrado contou com os mais diversos apoios e incentivos, aos quais gostaria de prestar aqui o meu sincero agradecimento.

À Instituição Bancária que disponibilizou os dados para esta dissertação, sem os quais teria sido impossível proceder com este estudo.

Ao Professor Jorge Bravo, pelo suporte e orientação que me forneceu ao longo deste projeto.

Aos meus colegas de trabalho, nomeadamente à Patrícia Reis, Manuel Ivo, José Pedro Marques, Dulce Mota e Anabela Mateus.

À minha família por todo o apoio, suporte e estabilidade que me dão sempre e por nunca deixarem de acreditar em mim. Com um especial agradecimento à minha irmã Filipa Pereira que acompanhou esta dissertação desde o princípio e que lutou por ela tanto como eu.

RESUMO

O objetivo deste projeto é analisar a aplicação do modelo Multiestado de Markov por forma a avaliar ao risco de credito do segmento de retalho em Portugal.

Os estados do modelo são definidos de acordo com as características da regulamentação Portuguesa DL n.º 227/2017. Esta regulamentação requere que as Instituições Financeiras Portuguesas tenham procedimentos para regularizar situações de incumprimento. As transições entre os estados são afetadas por variáveis explicativas sobre características do cliente e do seu comportamento em relação ao crédito contratado.

O uso do modelo Multiestado permitirá analisar a dinâmica do comportamento do ciclo de vida dos produtos crédito, devido à estimação de probabilidades de transição entre os diversos estados definidos.

PALAVRAS-CHAVE

Scoring de Crédito; Modelos Multiestado; Processos de Markov; Segmento de Retalho Português

ABSTRACT

The aim of this project is to analyze the application of a multi-state Markov model to evaluate credit risk for the retail Portuguese segment.

The states of the model are defined according to the characteristics of the Portuguese regulation DL no 227/2012, which requests Portuguese financial institutions to have a procedure to regularize default situations. Transitions between the states are affected by explanatory variables about the client's characteristics and his/her credit behavior.

The use of a multi-state model will allow the analysis of the dynamics of the behavior of credit products lifecycle, due to the estimation of transition probabilities between the several defined states.

KEYWORDS

Credit Scoring, Multi-state models, Markov Processes, Retail Portuguese Segment

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

BoP	Bank of Portugal
PARI	Plano de Ação para o Risco de Incumprimento (Action Plan for Default Risk)
PERSI	Procedimento Extrajudicial de Regularização de Situações de Incumprimento (Extra-Judicial Procedure to Regularize Default Situations)

1. INTRODUCTION

The increase in credit defaults during the last financial and economic crisis alerted Bank of Portugal (BoP), the Portuguese Central banking regulation and supervision agency, of the need for financial institutions to monitor their ongoing credits and follow more carefully the default alerts in these products. For this reason, in 2012, BoP released regulation DL no 227/2012, which requests all Portuguese financial institutions to have an Action Plan for Default Risk (PARI)[1] and an Extra-Judicial Procedure to Regularize Default Situations (PERSI).

This study will focus on the PERSI procedure, which consists in a negotiation model that seeks to achieve an agreement between the financial institution and the client to normalize the default situation in credit contracts celebrated with retail clients, without recourse to a court. PERSI presents a strict workflow of communications and analysis that the institutions will have to perform with the client, as soon as he or she fails a payment. This workflow allows to understand all the possible states that a client achieves before entering in default and it will be analysed using a multi-state model.

Monitor and control default risk has been a concern of financial institutions for a long time, especially after the regulation framework Basel II was released. This regulation allowed the institutions to have, with a previous approval of their national bank supervisor, internal rating models to calculate capital requirements, which has promoted researchers to continue to explore different methods for credit scoring modeling.

The aim of this study is to understand if the new Portuguese regulation, DL no 227/2012, (BoP, 2012) can improve the estimation of the probability of default and how the credit products cycles are behaving since the new regulation became effective.

Studying the behaviour of credit product cycles will be based on the use of a Markov multi-state model. While building the multi-state model, it is important to understand precisely the possible states and transitions that credit products life-cycles have. In addition, according with Hougaard (1999), creating a good structure can simplify the calculations and make the assumptions more transparent.

Using Markov chains helps to describe the dynamics of credit risk, since it estimates transition probabilities between different stages, for example different grades (Malik & Thomas, 2012).

This study brings a new perspective of credit scoring modeling for the Portuguese context, since to our knowledge this is the first project to study the effect of using a multi-state model for credit scoring purposes with a Portuguese database. Régis et. al (2015) present a similar study for the Brazilian context, however focusing on credit card risk, while this study addresses all credit products (except leasing contracts) used in the Portuguese market.

[1] PARI consists in an action plan with a set of procedures that financial institutions need to follow as soon as they detect evidences of the possibility of default risk in credit products or when the client informs the institution of the difficulties in complying with his credit obligations. For all clients that show evidences or notify the institution of the possibility of entering in default an evaluation of the financial capacity of the client has to be performed by the financial institution in order to confirm the existence of this risk.

This study has a practical application that can be of great use for financial institutions and banking supervisors if the model studied proves to have a better performance than the most common used model.

Additionally, this study allows to understand the behaviour of the lifecycles of credit products by analysing the probabilities between the different stages that were developed according with the regulation DL no 227/2012 (BoP, 2012).

There are several methodologies used to estimate the client's probability of default. According to Mays (2001) and Thomas (2009), the Logistic Regression Model is the most common method used to build credit scoring models, however this model does not allow studying the client's behaviour between the non-default and default state.

This study found no evidence supporting the use of the multi-state Markov model, designed with different stages of the PERSI program, to study the behaviour of credit products life-cycle. This can be partially explained by some of the limitations found at this stage (e.g. regulation being recent, lack of behavioural variables). It is nevertheless important to note that this is an area full of opportunities for further development particularly in light of the importance and impact that credit default events can have in the financial stability of many countries such as Portugal or Ireland.

The remainder of the document includes in Section 2 a literature review, followed by the Markov methodology and dynamics of the PERSI program in Section 3. Insights on the data and variables used are described Section 4 and Section 5 presents the results obtain in this study. Ultimately, Section 6 provides the conclusions of the study.

2. LITERATURE REVIEW

One of the major activities performed by banks is to concede credit to their clients. This activity has some risks since some of the clients might not be able to face their credit obligations. When a client fails to pay the outstanding debt, he/she is said to be in a state of default. Modelling credit events such as defaults has been a matter of research for the past 30 to 40 years. The following table presents and critically assesses the most common methods used for this purpose:

Table 2.1 - Credit Risk Methodologies

Methodology	Literature	Description	Advantages and Disadvantages
Expert-Judgement	N/A	The first approach to credit risk management was applied with the use of expert-judgment analysis. The Financial Institutions would only rely on subjective analysis based on the characteristics of the debtors.	This type of analysis is easy to implement and examines client by client, however is highly subjective and relies mainly on the debtor's characteristics and the analyst experience.
Discriminant Analysis	Fisher (1936); Altman (1968)	Discriminant analysis is a popular tool to assess the probability of default, and is based on a linear multivariate function. Altman (1968) used financial ratios to predict bankruptcy, and created the Z-score model.	This methodology allows to observe the contribution of each explanatory variables. However, it needs to have normally distributed explanatory variables.
Regression Models	Hosmer & Lemeshow (2013); Martin (1977); West (1985); Datschetzky (2005)	Regression models are also a popular method for credit scoring. The Logit Model assumes that the probability of default follows a logistic distribution, while the Probit Model assumes that it follows a standard normal distribution.	These models can generate simple probabilistic formulas for classification, among other advantages. Yet, they are not able to properly deal with problems of nonlinear and cooperative effects of explanatory variables.

Methodology	Literature	Description	Advantages and Disadvantages
K-nearest Neighbor	Chatterjee and Barcun (1970); Henley and Hand (1996); Brown & Mues, 2012; H. Chen & Chen, 2010; Lahsasna et al 2010	The K-nearest Neighbor is a non-parametric method. This algorithm analyzes patterns of the k nearest observations that are most similar to a new observation.	According to Hooman et al. (2016), the major advantage of this method is that it does not requires a predictive model previous to classification. One of the disadvantages of this method is the fact that it is not possible to produce a simple classification formula and that is highly dependent on the distance measure.
Classification and Regression Trees (CART)	Breiman et al (1984); Frydman, Altman and Kao (1985); Feldman, Gross (2005)	As the name suggests, this method can be described as a decision tree graph that classifies a dataset into a finite number of classes.	This method is very intuitive and is easy to explain, however, it can be computationally heavy when dealing with a large dataset.
Multivariate Adaptive Regression Splines (MARS)	Freidman (1991)	This method uses a non-linear parametric regression that is able to create additive relationships or involve interactions between a small number of variables (Hooman et al., 2016).	One of the biggest advantages of this method is that doesn't require pre-assumptions. On the other hand, it can be computationally heavy to execute.
Neural Networks	Altman (1994); Tam and Kiang (1992)	This method aims to replicate the human brain's way to process information, in order to distinguish the client's characteristics that are more related with the default event.	Neural Networks is a robust method that is able to generalize and deal with large dataset. However, it can lead to bias results when applied to small datasets.
Support Vector Machine	Boser, Guyon and Vapnik 1992; Kim and Sohn, 2010	The Support Vector Machine is an optimization method and a machine learning procedure.	Since this method is non-parametric it does not require data structure assumptions. However, it is a difficult method to interpret.

Methodology	Literature	Description	Advantages and Disadvantages
Hybrid Models	Zhang et al. (2008a); Harris, (2015); Oreski, (2014)	Hybrid Models are credit score models build by combining one or more methods. One example can be provided by Chen et al. (2009) with the conjugation of the CART and MARS methods.	By combining two or more methods these models can minimize weaknesses and assumptions, though are hard to implement and execute.
Survival Analysis	Beran and Djadja (2007); Chamboko & Bravo (2016)	This methodology was recently studied for credit scoring purposes. As opposed to other methodologies this method is able to predict the time of the event's incidence.	A clear advantage of this model is the ability to combine the probability of default with the time of the event. This model requires an optimization process that can be hard to execute when dealing with a large dataset.
Multi-State Models	Hougaard (1999)	These types of models are used to describe the history of a client. Multi-State models are stochastic processes that predict at which time the individual transits into specific state, from a set of states.	One advantage of these models over the other is that it can estimate the time of occurrence of more than one event. These models can be hard to implement and execute due to the complex optimization process and the number of explanatory variables.

Most methodologies used to evaluate credit risk, namely to estimate the client's probability of default, only consider the client's transition from a non-defaulting state to a defaulting state, for example the Logistic Regression Model (Régis et al., 2015). However, the default definition by Basel Committee on Banking Supervision (2005) allows the existence of several possible states between the client and the financial institution.

The Basel Committee on Banking Supervision (2005) defined the default state as the following:

> "…A default is considered to have occurred with regard to a particular obligor when either or both of the two following events have taken place:
>
> - The bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without recourse by the bank to actions such as realizing security (if held).

- The obligor is past due more than 90 days on any material credit obligation to the banking group. Overdrafts will be considered as being past due once the client has breached an advised limit or been advised of a limit smaller than current outstandings (p. 100).”

Using Markov chains helps to describe the dynamics of credit risk, since it estimates transition probabilities between different states, for example transitions between risk grades (Malik &Thomas, 2012). Leow & Crook (2014) mention two advantages of these models: they can estimate predictions of transition probability matrix in any future time period and are able to elaborate more complex predictions of all states of delinquency until the state of default, which, aligned with explanatory variables about the debtors’ characteristics and behaviour towards the credit, allows the financial institution to understand the factors that have an impact on the movements towards the default or the recovery.

The particularity of Markov models is that they account for the Markov assumption, which states that the probability of the next transitions only depends on the current time, that is, it’s independent from the historical background (Hougaard, 1999). This assumption might not always be appropriate for certain studies. Nevertheless, according to Hougaard (1999) choosing the structure of the model is of extreme importance, since it can simplify the calculations, alter the assumptions of the model and, ultimately, turn a non-Markov model into Markov.

Several Studies have been carried out to test the suitability and accuracy of multi-state models for credit risk management purposes. Table 2 summarizes the literature on these models, including chosen transitions states and explanatory variables.

Table 2.2 - Multi-State Model Studies

Using multi-state Markov models to identify credit card risk Régis, D. E. & Artes, R. - 2015
Study Objectives This study aims to analyse the application of a multi-state Markov model to predict credit card default and product cancelation. Additionally, this study compares the Markov model with a Logistic Regression model in order evaluate which one performs better.
States/Transitions This study uses a Markov model with 5 states: 1. Compliance (transits to revolving, delay and voluntary cancelation); 2. Revolving (transits to compliance, delay and voluntary cancelation); 3. Delay (transits to compliance, revolving, voluntary cancelation and default); 4. Voluntary Cancelation (absorbent state); 5. Default (absorbent state).
Variables This study applies 7 variables, 1 regards to a clients’ characteristic and the other 6 regard to the clients’ credit behaviour.

1. Categorical variable with a range of credit limit according with the client's income;
2. Use of revolving credit over 12 months;
3. Inactivity of the client over 12 months;
4. Intensity of delay problems over 12 months;
5. Intensity of product use over 6 months (assigning a bigger weight to recent months);
6. Usage of credit limit over 6 months;
7. Maximum client debt over 6 months.

Conclusions

Regarding the variables used in this study all were significant at least for two kinds of transitions.
The multi-state Markov model showed a better performance to predict default and worse to predict the cancellation of the credit card.

Intensity models and transition probabilities for credit card loan delinquencies

Leow, M. & Crook, J. - 2014

Study Objectives

The goal of this study is to estimate the probability of delinquency and default for the product credit cards, with an intensity model, via semi-parametric hazard models with time-varying covariates.

States/Transitions

This study is modelling the clients' behaviour over 4 states:
1. up-to-date (transits to one month in arrears, two months in arrears or default)
2. one month in arrears (transits to up-to-date or two months in arrears)
3. two months in arrears (transits to up-to-date and default)
4. default (absorbent state)

Variables

This study applies 15 variables, 10 are characteristic variables (one of them is not explained due to reasons of confidentiality) and 5 are behavioural variables.
The characteristic variables are: age, employment status, number of cards at application, time at address in years, indicator for presence of landline, time in bank in months, indicator for missing time with the Bank, Income and Indicator for missing income.
The behavioural variables used in this study were: credit limit over 3 months, payment amount over 3 months, proportion of credit over 3 months, rate of total jumped over 3 months, indicator of improvement in state from the 3 previous months.

Conclusions

In this study the authors observed that most application variables affect the risk of delinquency similarly to what was expected. Also, it was concluded that some group of people are better in keeping themselves in delinquency without moving to default.
When validating the model with a validation sample, the authors observed that on an overall level the model made good predict, however on an account level the model didn't perform so well.

Using Markov chains to estimate losses from a portfolio of mortgages

Betancourt, L. - 1999

Study Objectives

The objective of this study is to assess the suitability and accuracy forecast of three Markov chain models of mortgages loan losses, and assess whether the data complies or not with the assumption of credit payment behaviour being homogenous and the transition probabilities being stationary.
The three Markov models used in this study are the following: 1. Base model that assumes homogeneity and stationarity; 2- LTV Model, which is in all similar to the Base model but introduces the use of an explanatory

categorical variable "Loan-to-Value"; 3- KS model, which uses the space definition of the Base model and an adjusted transition matrix, in order to cover the problem of non-stationary transition probabilities.

States/Transitions

The states defined in this study are the states of a mortgage loan:

1. Active; 2. Thirty days of delinquent; 3. Sixty days of delinquent; 4. Ninety plus days of delinquent; 5. Foreclosure; 6. Real Estate Owned (REO), and (7) Paid-off.

Variables

Only in the second model the author introduces one explanatory variable "Loan-to-value".

Conclusions

Regarding the suitability of the models, the author studied the assumptions of stationarity and homogeneity. Both assumptions were rejected in all tests performed. For the forecast accuracy of the models the author used four tests: Mean absolute Percentage Error, Root Mean Square Error, Mean Forecast Error and Quarterly Forecast Error. The author concluded that all models unpredicted the level of loan losses, however the KS model generated the most accurate forecasts over each time horizon. The results of the error metrics showed that the predictive accuracy of the Markov chain approach is highly influenced by the time period used to estimate the transition matrix.

Markov chain for delinquency: Transition matrix estimation and forecasting

Grimshaw, S. D. & Alexander, W. P. - 2010

Study Objectives

This paper proposes two estimation methods for the transition matrix of subprime loans.

In this study the segmentation is performed by pooling data from loans in the same segment and borrowing strength from data in other segments. Additionally, this study uses loan-level models for key transitions to allow the use of covariates.

States/Transitions

The example presented in this study uses the following states:

1. Current; 2. 30 days past due; 3. 60 days past due; 4. 90 days past due; 5. 120 or more days past due; 6. Loss; 7. Paid

Variables

The variables used in the example presented in this study are:

1. Number of delinquent months; 2. Loan-to-value percentage; 3. Proprietary credit score; 4. Number of months since loan origination; 5. Interest Rate

Conclusions

Using a simulation study the authors concluded that the estimated transition probability matrix produces reasonable predictions.

The multi-latent factor intensity model for credit rating transitions

Koopman, S. J. & Lucas, A. & Monteiro, A. – 2007

Study Objectives

This study introduces a new model for credit rating transitions. The model introduced is a parametric intensity-based duration model with multiple states, exogenous variables and latent dynamic factors.

States/Transitions

The simulation carried out in this study uses 8 states, which are the familiar rating grades:

1. AAA; 2. AA; 3. A; 4. BBB; 5. BB; 6. B; 7. CCC; 8. D

Variables

This study doesn't use explanatory variables to perform the developed model.

Conclusions

The authors found a significant common risk factor in the credit rating migrations, which have a higher impact for downgrades than for upgrades.

The structure of the presented model can easily incorporate general specification, such as, observed firm-specific and economic variables.

Additionally, the model specification proposed in this study allows the estimation and testing of the number of latent factors driving to default.

Credit Scoring with Macroeconomic Variables Using Survival Analysis

Bellotti, T. & Crook, J. – 2007

Study Objectives

The aim of this paper is to study an application of survival analysis to model the probability of default in a database of credit card accounts.

Additionally, this paper tests the hypothesis that the probability of default is affected by general economic conditions.

States/Transitions

Since this paper uses survival analysis, only two states are considered in this study:

1. Solvent; 2. Default

Variables

This study focus on the use of macroeconomic variables, such as:

interest rate, ratio of earning, FTSE, unemployment rate, index of production, house price index, consumer confidence index.

Besides these variables, this study also used application variables, which for reasons of confidentiality were not reported in the paper.

However, an automated selection model described in the paper, selected some interaction between the previously mentioned macroeconomic variables and some characteristic variables, such as: client's income, home owner (y/n), private tenant (y/n), home council (y/n), employed (y/n), self-employed (y/n), unemployed (y/n) and bureau score.

Conclusions

One of the main conclusions obtained in this study is that the inclusion of macroeconomic variables in the Cox PH model improves model fit.

Regarding the significance of the variables, the most important macroeconomic variable observed to estimate the risk of default is the interest rate.

The interactions that proved to be significant were interest rate with income and the index of production with the credit bureau score.

Affine Markov chain models of multifirm credit migration

Hurd, T. & Kuznetsov, A. - 2006

Study Objectives

The purpose of this paper is to present an extension of the Chen-Filipovic affine models with a Markov chain for the "credit ratings" of each firm.

States/Transitions

The states considered in this study are the traditional credit ratings:

1. AAA; 2. AA; 3. A; 4. BBB; 5. BB; 6. B; 7. CCC; 8. D

Variables

This study doesn't mention the use of explanatory variables in the example performed.
Conclusions One of the conclusions of this study was that the introduced model is flexible and computationally efficient. This model reflects the dynamics of the market conditions.

Transition matrix models of consumer credit ratings Malik, M. & Thomas, L. C. – 2012
Study Objectives The aim of this study is to develop a model for the credit risk of portfolios of consumer loans. The authors build a Markov chain credit risk model based on behavioural scores.
States/Transitions The states used in this study were selected from the behavioural scores defined by the data provider. In this case the higher the score is the least risky the client is. The range between the scores, used to create the states, were obtained with the analysis of a chi-square statistic. The chosen scores are the following: 1. [13;680]; 2. [681;700]; 3. [701;715]; 4. [716;725]; 5. [726;++)
Variables Since this model is based on a behavioural score model, which is usually constructed using characteristic variables, the variables explored to be used in this study where macroeconomic variables and variables related with the product: 1. % change in the consumer price index; 2. Monthly average sterling inter-bank lending; 3. Annual return on FTSE 100; 4. % change in the GDP; 5. Unemployment rate; 6. % change in net lending over 12 months; 7. Product months on the books
Conclusions Regarding the transition probabilities estimated, the authors observed that the least risky and the most risky states have the highest probability of staying in the same state, while others tend to move more. By comparing the transition probabilities between two different periods, the authors were able to prove that the change in economic conditions has an impact in the transition matrix. Finally, the authors showed that the second-order model with economic variables present a better prediction of occurred defaults than the second-order model without these variables.

The Portuguese regulation, DL no 227/2012, (BoP, 2012) requires financial institutions to have a workflow process, the PERSI, of communications and analysis with the clients as soon as they stop paying, in order to facilitate an agreement between the institution and the client to regularize the default situation.

The PERSI was designed for retail clients that have celebrated a credit contract with a financial institution. A detailed scheme of this program is presented in the Methodology Section.[2]

This regulation requires financial institutions to notify their clients when they fail a payment within 15 days. If the client continues to not comply with his credit obligations, between the 31st and 60th day of not paying his debts, this client must integrate the PERSI program and is notified, within 5 days, of this integration.

[2] Leasing contracts of durable goods that establish the rights and duties of purchasing the good are not covered by the PERSI program.

When a client enters the PERSI program, the institution has to evaluate his financial capacity and present to the client, within 30 days from the start of the PERSI program, a proposal to regularize the default situation. After the proposal is presented, both the client and the institution have until the 90th day since the PERSI program has started to negotiate the conditions of the proposal and reach an agreement. When both parts agree on a proposal, the client returns to a solvent state and is no longer considered to be in a default situation.

The PERSI is automatically closed if one of the following situations happens:

- The client pays all his debts;
- The client and the institution reach to an agreement to regularize the default situation;
- In the 91st day since the program has started, the client and the institution did not agree on any of the presented proposals and there was no delay of the deadline established between the two parts;
- The institution declares bankruptcy.

Meanwhile, under certain situations, the financial institution can close the PERSI program, for example in the case of a mortgage loan if the pledge is executed.

This study will use the PERSI procedures to determine the states and transitions between the solvent and the default states. For this reason, an alternative definition of default will be used: a client enters a state of default after the PERSI program has finished and no solution was found to solve the client’s debts.

3. METHODOLOGY

A multi-state model is described as a model for stochastic process, where individuals move through several states (Hougaard, 1999). A change of state is denoted as a transition, which can be reversible or irreversible. The states can be absorbing, where no transition happens from this state, or transient, that is, transitions go in and out of this state.

The Markov multi-state model assumes that future evolution only depends on the current state and on the covariates (Jackson, 2011) and that the transition probability is only dependent on the difference between the times of the two states. Considering $E_i(t)$ the state of the client i at instant t, the probability of a client to transfer form an l state into an m state is given by the following formula:

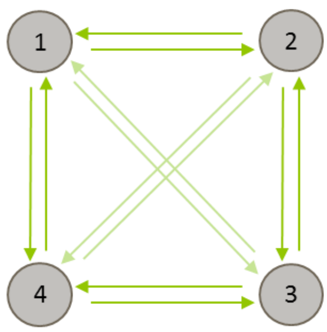
$$p_{lm}(\Delta t) = P(E_i(t + \Delta t) = m \,|E_i(t) = l) \tag{1}$$

Figure 1 illustrates an example of a multi-state model with 4 states. The arrows represent the possible transitions between the states. The individual's movements between states are ruled through the transition intensities, $q_{lm}(t)$, of each combination of states l and m.

The transition intensity can be defined as the instantaneous risk of an individual migrating from the state l to m, and it is given by the following:

$$q_{lm}(t) = \lim_{\Delta t \to 0} \frac{P(E_i(t + \Delta t) = m \,|E_i(t) = l)}{\Delta t} \tag{2}$$

The intensities form a square matrix Q of order n, where n is the number of states in question and the sum of the rows is equal to 0 since the $q_{rr} = \sum_{i \neq r} q_{ri}$.



$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{12} & q_{12} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{pmatrix}$$

Figura 3.1 - General multi-state model and respective transition intensity matrix

Working with time-homogeneous continuous-time Markov model allows to assume that the period of occupancy in a state l is given by an exponential distribution with a rate of $-q_{ll}$. Due to this assumption, it is possible to demonstrate that the transition probability matrix at time t can be calculated through the exponential of the multiplication of t by the transition intensities matrix Q.

$$P(\Delta t) = \exp(\Delta t \times Q) \tag{3}$$

Having $x_i(t)$ a vector of explanatory variables of the debtor i at time t, Marshall & Jones (1995) developed a form to determine the transition intensity matrix:

$$q_{rs}\big(x_i(t)\big) = q_{rs}^{(0)} \exp(\beta_{rs}^T x_i(t)) \tag{4}$$

where $q_{rs}^{(0)}$ is the baseline transition from the state r to the states s and β_{rs}^T is a vector of coefficients associated with the explanatory variables for the transition from the state r to the states s. The coefficients β_{rs}^T are estimated through the optimization of the maximum likelihood:

$$L = \prod_{i=1}^{n} \prod_{k=0}^{t-1} p_{S(t_k)S(t_{k+1})} \tag{5}$$

Hougaard (1999) and Jackson (2016) present a more detailed explanation of the theoretical concepts behind these models.

Constructing and structuring a multi-state model is a fundamental part of this project, as it will not only determine certain variables that need to be collected, but it will also determine the potential of using a Markov and a non-Markov model (Hougaard, 1999).

The PERSI model, developed in the regulation DL no 227/2012 (BoP, 2012), served as a guide to understand all the states and transitions that clients go through until they reach the default state.

The following scheme presents the process flow since the client stops complying with his credit obligations until the PERSI program ends:

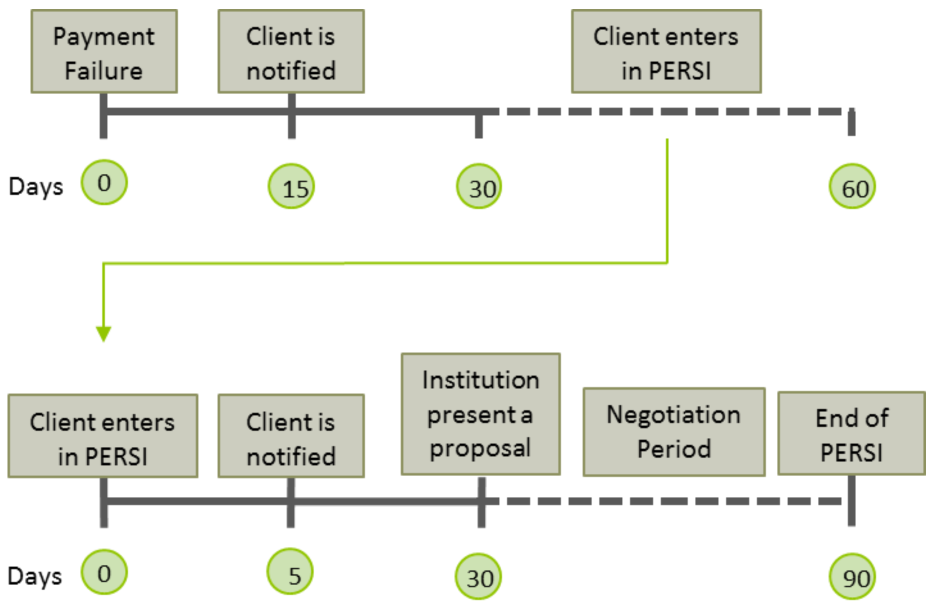


Figure 3.2 - PERSI Scheme

By analysing and understanding the process flow of the PERSI program and the full life-cycle of credit product, it is possible to identify the main states that clients go through. The states considered in the study are the following:

- State 1 – Solvent: In this state all the credit obligations of the client with his credit product have been complied.
- State 2 – Payment failure: the customer failed his credit obligation of the credit product in question.
- State 3 – PERSI: the client has integrated the PERSI program, since he has not paid his credit charges for a maximum of 60 days.
- State 4 – Restructured product: the client has restructured is product in order to be able to comply with his credit obligations.
- State 5 – Default: the client has left the PERSI program without a negotiation or without paying his obligation, he entered a state of default.

The states 4 and 5 are considered as absorbent states. The figure 3 represents the process of this model with all the states and the possible transitions between them.

From the state 1 is possible to transit to and from the state 2, besides it is also possible to transit from the state 3 to the state 1, when a client pays all his debts while he is in the PERSI program. From the state 2 to the state 3 exists an irreversible transition. Since the states 4 and 5 are absorbent states, there are no transitions going out from these states, there are only two irreversible transitions from the state 3 to each one of these states.

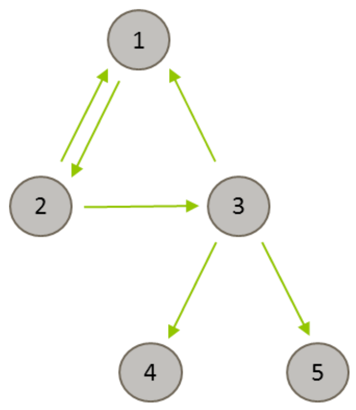


Figura 3.3 - Multi-state model of a credit life cycle

The definition of default used in this study is completely linked with the end date of the PERSI program. As previously mentioned the Basel Committee considers that a client enters in a state of default when he is past due 90 days on his credit obligation. The end date of the PERSI program has an interval range that can vary from 60 days to 150 days after the client's credit obligation failure.

There are several analyses to measure the goodness of fit of these models such as Gini Coefficient and Kolmogorov Test (Régis et al., 2015), accuracy ratios (Leow & Crook, 2014) and Pearson Chi-square test (Jackson, 2016).

Leow and Crook (2014) mention an accuracy ratio to assess the performance of the estimated model, the ratio is calculated as the following:

$$AR_{hj} = \frac{number\ of\ expected\ hj\ transitions}{number\ of\ observed\ hj\ transitions} \tag{6}$$

The closer the ratio is to 1, the better the model performs.

The Pearson's Chi-square test, when used with the purpose of assessing the goodness of fit of a model, compares the distribution of the observed transitions to the theoretical distribution. The Pearson Chi-square statistic is given by the following formula:

$$T = \sum_{rs} \frac{(Obs_{rs} - Exp_{rs})^2}{Exp_{rs}} \tag{7}$$

$$T \sim \chi^2_{n-p}$$

where Obs_{rs} represent the observed transitions form the state r to s, Exp_{rs} the expected transitions between these states, n the number transitions and p the degrees of freedom (Jackson, 2016).

4. DATA

The data used in this project has been provided by one of the largest Portuguese financial institutions, quoted in Euronext Lisbon and part of the PSI 20[3], which operates mainly in retail banking. Due to reasons of confidentiality, the company's name will not be revealed and the sample used in this project will not reflect the company's credit risk portfolio characteristics.

The multi-state model has some constrains regarding the construction of the model inputs. The data must be aggregated in a series of observations, grouped by individual. Moreover, there are certain variables that are fundamental when applying a multi-state model, such as the time of the observation, the observed state and the client's identification number.

The dataset used for this study was collected by account and included 135.000 unique accounts of housing credit and 20.000 accounts of consumer credit. Both datasets were divided into two samples on an 80%/20% proportion of unique accounts. The smaller samples were stored separately, in order to have independent samples to validate the model.

In the beginning of 2013, the Portuguese financial institutions were required to have the PERSI program installed. The storage database created for this purpose by the data provider entity suffered some changes. For this reason, the observation period chosen for this project goes from June 2014 to July 2016. During this period, a history of the accounts state was collected within every 15 days.

The criteria used to select the sample was the following:

- The accounts are open during the full observation period;
- All accounts are, at the reference date (1 of June 2014, beginning of the observation period), in the state one;
- None of the accounts had had any payment failure before the reference date.

The following tables present the number of observations per transitions:

Table 4.1 - Transitions Observations

Consumer Credit					
from	**1**	**2**	**3**	**4**	**5**
1	965327	2865	0	0	0
2	2259	9615	231	0	0
3	193	0	629	4	21

Housing Credit					
from\to:	**1**	**2**	**3**	**4**	**5**
1	1488956	2398	0	0	0
2	2007	5344	194	2	0
3	159	0	507	12	12

Source: author's preparation based on data.

Depending on the number of variables the multi-state Markov model can become highly computational heavy. The higher the number of variables added to the model the more complex it becomes the optimization process.

For this reason, the model was developed with only 7 variables, 4 of which are behaviour variables that capture the credit performance of the customer. The other variables are customer characteristics that can have an impact on the credit behaviour. The chosen variables are the following:

[3] Euronext Lisbon is the Portuguese stock exchange and PSI 20 (Portuguese Stock Index) is comprised of the 20 largest companies.

Variable 1: Age group of the client;
Variable 2: Marital status of the client;
Variable 3: Employment status of the client;
Variable 4: Default of the client previous to the reference date;
Variable 5: Time interval group of how long the account has been open until the reference date;
Variable 6: Percentage of months with excess balance;
Variable 7: Number of months since the last debit.

The choice of the behavioural variables was restricted to the database provided to develop this study; it was therefore not possible in the context of this study to evaluate some of the most impactful variables for the default and recovery probabilities estimation. On the other hand, the characteristic variables were chosen according with the study of Leow & Crook (2014), which demonstrated good predictability using mainly characteristic variables.

The variables mentioned above are all categorical; the output of these variables is expressed in Table 4.

Table 4.2 - Covariates

Variable	**Type**	**Description**	**Output**
Variable 1	Characteristic	Age	A1 - Between [18-25] years old A2 - Between [26-30] years old A3 - Between [31-40] years old A4 - Between [41-55] years old A5 - Between [56-65] years old A6 - Between [66-999] years old
Variable 2	Characteristic	Marital Status	MS1 - Married MS2 - Divorced MS3 - Single MS4 - Widower
Variable 3	Characteristic	Employment Status	ES1 - Employed ES2 - Self Employed ES3 - Unemployed ES4 - Not Employed
Variable 4	Behavioural	Default	D1 - Client never default D2 - Client with previous default
Variable 5	Behavioural	Time on the Books	TB1 – Between [0,49] days TB2 – Between [50,69] days TB3 – Between [70,89] days TB4 – Between [90,109] days TB5 - Between [110,129] days TB6 - Between [130,164] days TB7 - Between [165,199] days TB8 – More than 199 days
Variable 6	Behavioural	% Months with excess balance	EB1 – Between [0%,9%[EB2 – Between [9%,27%[EB3 – Between [27%,60%[EB4 – Between [60%,100%]
Variable 7	Behavioural	Months Since Last Debit	LD1 – 0 months LD2 – 1 months LD3 – Between [2,3] months LD4 – Between [4,9] months LD5 – More than 9 months

Source: author's preparation based on data.

Based on the samples stored for the model development test an analysis on the correlation between the variables was performed. The following tables present the variables correlation matrix for the each product type.

Table 4.3 - Variable Correlation Matrix for Consumer Accounts

Consumer Credit							
Var	**A**	**MS**	**ES**	**D**	**TB**	**EB**	**LD**
A	1.00	-0.39	0.31	-0.01	0.44	-0.06	0.07
MS	-0.39	1.00	-0.03	0.02	-0.26	0.05	-0.04
ES	0.31	-0.03	1.00	0.02	0.05	0.00	0.02
D	-0.01	0.02	0.02	1.00	-0.01	0.15	-0.13
TB	0.44	-0.26	0.05	-0.01	1.00	-0.06	0.07
EB	-0.06	0.05	0.00	0.15	-0.06	1.00	-0.82
LD	0.07	-0.04	0.02	-0.13	0.07	-0.82	1.00

Source: author's preparation.

Table 4.4 - Variable Correlation Matrix for Housing Credit Accounts

Housing Credit							
Var	**A**	**MS**	**ES**	**D**	**TB**	**EB**	**LD**
A	1.00	-0.30	0.20	0.02	0.35	-0.03	0.04
MS	-0.30	1.00	-0.03	0.01	-0.16	0.01	0.00
ES	0.20	-0.03	1.00	0.03	0.08	0.03	-0.02
D	0.02	0.01	0.03	1.00	-0.01	0.21	-0.18
TB	0.35	-0.16	0.08	-0.01	1.00	-0.03	0.04
EB	-0.03	0.01	0.03	0.21	-0.03	1.00	-0.83
LD	0.04	0.00	-0.02	-0.18	0.04	-0.83	1.00

Source: author's preparation.

Apart from the variables 6 and 7, none of the variables present high correlation ratios, which indicates that there is no overlapping in the variables. Although the variables 6 and 7 present a high correlation coefficient, both provide a good insight on the client credit behaviour and for this reason they were both kept for the development of the model.

5. RESULTS

To develop this study, the SAS and R software solutions were used, the first one to treat and arrange the data and the second one to perform and estimate the model using the package “msm” developed for multi-state Markov modelling (Jackson, 2016).

Due to the differences in the characteristics of the two credit products studied in this thesis, consumer and housing credit, two models were estimated using the same variables, one for each credit product.

As shown in the previous section, the dataset used in this study is highly concentrated in the state 1 and there are not many cases in the states 4 and 5. The characteristics of the dataset affect the results obtained and do not allow to have strong conclusions regarding the impact of the variables in the transitions between states.

5.1. MODEL ESTIMATION

The obtained estimators are expressed in Table 7 and represent the relationship between the variables and the transitions. The asterisk indicates the estimators that are statistical not significant at 5%.

By observing the table 7 it is possible to verify that all variables are statistically insignificant for transitions between the state 1 and 2. This fact can be explained by the characteristic of the dataset used for this study, such as the amount of transitions in the state 1, followed by transitions between the states 1 and 2. This feature does not allow capturing the impact of the variables between the solvent state and the payment failure.

Tables 8 and 9, present the transitions probabilities estimated for consumer credit accounts and housing credit accounts, respectively, one year after the reference date.

The probability of remaining in state 1 after one year is quite high for both models and for all variables. This result can be explained by the assumptions that were taken into consideration when the dataset was collected, namely, this study only considers, at the reference date, accounts that never had a default or a payment failure.

Table 5.1 - Transitions Intensities

Variable - Description	Transitions											
	1-2		2-3		3-4		3-5		2-1		3-1	
	CC	HC	CC	HC	CC	HC	CC	HC	CC	HC	CC	HC
A1 - Between [18-25] years old	-3.5079*	-3.3091*	-1.226	-0.760	-2.266	-0.515	-0.814	-1.605	2.3652*	2.8223*	1.512	1.378
A2 - Between [26-30] years old	-3.5254*	-3.2381*	-1.231	-0.716	-2.268	-0.500	-0.808	-1.535	2.3638*	2.8031*	1.546	1.308
A3 - Between [31-40] years old	-3.5425*	-3.1675*	-1.236	-0.673	-2.271	-0.484	-0.801	-1.465	2.3624*	2.7839*	1.580	1.238
A4 - Between [41-55] years old	-3.56*	-3.0967*	-1.241	-0.629	-2.273	-0.469	-0.794	-1.395	2.361*	2.7646*	1.614	1.167
A5 - Between [56-65] years old	-3.5773*	-3.026*	-1.245	-0.585	-2.276	-0.453	-0.787	-1.325	2.3597*	2.7454*	1.648	1.097
A6 - Between [66-999] years old	-3.5947*	-2.9552*	-1.250	-0.541	-2.278	-0.438	-0.781	-1.255	2.3583*	2.7261*	1.683	1.027
MS1 - Married	-3.4122*	-3.3385*	-1.181	-0.549	-2.257	-0.513	-0.838	-1.678	2.3517*	2.8536*	1.418	1.569
MS2 - Divorced	-3.334*	-3.2976*	-1.140	-0.295	-2.251	-0.495	-0.854	-1.682	2.3368*	2.8655*	1.359	1.690
MS3 - Single	-3.2555*	-3.2563*	-1.100	-0.040	-2.245	-0.478	-0.870	-1.685	2.3219*	2.8775*	1.300	1.810
MS4 - Widower	-3.177*	-3.2154*	-1.060	0.215	-2.239	-0.460	-0.887	-1.689	2.307*	2.8895*	1.240	1.930
ES1 - Employed	-3.4213*	-3.2416*	-1.139	-0.702	-2.245	-0.520	-0.832	-1.647	2.3839*	2.9999*	1.487	1.480
ES2 - Self Employed	-3.3518*	-3.1035*	-1.057	-0.600	-2.227	-0.509	-0.844	-1.619	2.4013*	3.1583*	1.496	1.511
ES3 - Unemployed	-3.2826*	-2.9654*	-0.975	-0.498	-2.209	-0.498	-0.855	-1.592	2.4187*	3.3167*	1.505	1.542
ES4 - Not Employed	-3.2131*	-2.8273*	-0.893	-0.396	-2.190	-0.487	-0.866	-1.564	2.4361*	3.4751*	1.514	1.573
D1 - Client previous default	-3.4907*	-3.3796*	-1.221	-0.804	-2.263	-0.531	-0.821	-1.675	2.3666*	2.8416*	1.477	1.449
D2 - Client never default	-2.5132*	-2.352*	-1.204	-0.613	-2.265	-0.533	-0.832	-1.682	1.9226*	2.8515*	1.445	1.293
TB1 – Between [0,49] days	-3.5947*	-3.4754*	-1.130	-0.818	-2.249	-0.608	-0.780	-1.599	2.4027*	2.7919*	1.523	1.518
TB2 – Between [50,69] days	-3.6985*	-3.5713*	-1.038	-0.831	-2.234	-0.686	-0.738	-1.522	2.4389*	2.7421*	1.569	1.588
TB3 – Between [70,89] days	-3.8023*	-3.6671*	-0.946	-0.845	-2.220	-0.763	-0.696	-1.446	2.475*	2.6924*	1.615	1.657
TB4 – Between [90,109] days	-3.9065*	-3.7627*	-0.855	-0.858	-2.206	-0.840	-0.654	-1.369	2.5112*	2.6427*	1.661	1.727
TB5 - Between [110,129] days	-4.0102*	-3.8585*	-0.763	-0.872	-2.191	-0.918	-0.613	-1.292	2.5473*	2.593*	1.707	1.796
TB6 - Between [130,164] days	-4.1141*	-3.9544*	-0.672	-0.885	-2.177	-0.995	-0.571	-1.216	2.5835*	2.5433*	1.753	1.866
TB7 - Between [165,199] days	-4.2179*	-4.0501*	-0.580	-0.898	-2.162	-1.073	-0.529	-1.139	2.6196*	2.4936*	1.799	1.935
TB8 – More than 199 days	-4.3222*	-4.1458*	-0.488	-0.912	-2.148	-1.150	-0.487	-1.063	2.6558*	2.4439*	1.845	2.005
EB1 – Between [0%,9%[	-2.6397*	-2.6468*	-1.234	-0.809	-2.280	-0.496	-0.812	-1.675	2.2487*	2.7833*	1.499	1.477
EB2 – Between [9%,27%[	-1.7886*	-1.9139*	-1.247	-0.814	-2.297	-0.462	-0.802	-1.676	2.1309*	2.7251*	1.5215*	1.504
EB3 – Between [27%,60%[	-0.9377*	-1.1809*	-1.259	-0.818	-2.313	-0.428	-0.793	-1.676	2.0131*	2.6669*	1.5436*	1.532*
EB4 – Between [60%,100%]	-0.087	-0.448	-1.272	-0.823	-2.330	-0.394	-0.783	-1.676	1.8953*	2.6086*	1.5657*	1.5597*
LD1 – 0 months	-4.2666*	-4.3685*	-1.334	-0.936	-2.265	-0.526	-0.811	-1.676	2.0547*	2.5005*	1.435	1.397
LD2 – 1 months	-5.0427*	-5.3579*	-1.446	-1.067	-2.266	-0.522	-0.800	-1.676	1.7428*	2.1594*	1.392	1.344
LD3 – Between [2,3] months	-5.8189*	-6.347*	-1.559	-1.199	-2.267	-0.518	-0.789	-1.677	1.4309*	1.8184*	1.349	1.292
LD4 – Between [4,9] months	-6.5951*	-7.3362*	-1.671	-1.331	-2.268	-0.514	-0.778	-1.677	1.119*	1.4773*	1.306	1.240
LD5 –More than 9 months	-7.3707*	-8.3253*	-1.784	-1.462	-2.270	-0.510	-0.767	-1.678	0.8071*	1.1362*	1.264	1.188

Source: author's preparation.

The probability of default in the variable age (variable 1) slightly decreases with the increase of the age in the consumer credit. The same is not observed in housing credit accounts, which have an increase in the probability of default as the age increases.

Common sense tells us that a married couple has a lower probability of defaulting and a higher probability of recovering. In this study, this variable (variable 2) does not have any impact when transiting from state 1, but, when transiting from state 3, clients that are widowers are more likely to default in consumer credit products and married clients are slightly more likely to default in housing credit products.

Table 5.2 - Probabilities Consumer Credit

Variable	Transitions														
	1-1	1-2	1-3	1-4	1-5	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	3-5
A1	0.997	**0.003**	0.000	0.000	0.000	**0.994**	0.003	**0.000**	0.001	0.002	**0.884**	0.002	0.006	**0.020**	**0.087**
A2	0.997	**0.003**	0.000	0.000	0.000	**0.994**	0.003	**0.000**	0.001	0.002	**0.888**	0.002	0.005	**0.020**	**0.085**
A3	0.997	**0.003**	0.000	0.000	0.000	**0.994**	0.003	**0.000**	0.001	0.002	**0.891**	0.002	0.005	**0.019**	**0.083**
A4	0.997	**0.003**	0.000	0.000	0.000	**0.994**	0.003	**0.000**	0.000	0.002	**0.895**	0.002	0.004	**0.018**	**0.081**
A5	0.997	**0.003**	0.000	0.000	0.000	**0.995**	0.003	**0.000**	0.000	0.002	**0.898**	0.002	0.003	**0.018**	**0.079**
A6	0.997	**0.003**	0.000	0.000	0.000	**0.995**	0.003	**0.000**	0.000	0.002	**0.901**	0.002	0.003	**0.017**	**0.077**
MS1	0.997	**0.003**	0.000	0.000	0.000	**0.993**	0.003	**0.001**	0.001	0.003	**0.874**	0.003	0.010	**0.022**	**0.092**
MS2	0.996	**0.003**	0.000	0.000	0.000	**0.992**	0.003	**0.001**	0.001	0.003	**0.866**	0.003	0.012	**0.024**	**0.095**
MS3	0.996	**0.004**	0.000	0.000	0.000	**0.991**	0.004	**0.001**	0.001	0.003	**0.858**	0.003	0.015	**0.025**	**0.098**
MS4	0.996	**0.004**	0.000	0.000	0.000	**0.990**	0.004	**0.001**	0.001	0.003	**0.850**	0.003	0.019	**0.026**	**0.102**
ES1	0.997	**0.003**	0.000	0.000	0.000	**0.993**	0.003	**0.001**	0.001	0.003	**0.882**	0.003	0.007	**0.021**	**0.087**
ES2	0.997	**0.003**	0.000	0.000	0.000	**0.993**	0.003	**0.001**	0.001	0.003	**0.883**	0.003	0.007	**0.021**	**0.085**
ES3	0.996	**0.003**	0.000	0.000	0.000	**0.993**	0.003	**0.001**	0.001	0.003	**0.885**	0.003	0.007	**0.022**	**0.084**
ES4	0.996	**0.003**	0.000	0.000	0.000	**0.992**	0.003	**0.001**	0.001	0.003	**0.886**	0.003	0.006	**0.022**	**0.082**
D1	0.997	**0.003**	0.000	0.000	0.000	**0.994**	0.003	**0.001**	0.001	0.002	**0.880**	0.002	0.007	**0.021**	**0.089**
D2	0.988	**0.011**	0.001	0.000	0.000	**0.982**	0.012	**0.002**	0.001	0.004	**0.869**	0.010	0.009	**0.022**	**0.090**
TB1	0.997	**0.002**	0.000	0.000	0.000	**0.994**	0.002	**0.000**	0.001	0.003	**0.883**	0.002	0.006	**0.020**	**0.089**
TB2	0.998	**0.002**	0.000	0.000	0.000	**0.994**	0.002	**0.000**	0.001	0.003	**0.885**	0.002	0.005	**0.020**	**0.088**
TB3	0.998	**0.002**	0.000	0.000	0.000	**0.994**	0.002	**0.000**	0.001	0.003	**0.887**	0.002	0.004	**0.019**	**0.088**
TB4	0.998	**0.002**	0.000	0.000	0.000	**0.995**	0.002	**0.000**	0.001	0.003	**0.889**	0.001	0.003	**0.019**	**0.088**
TB5	0.998	**0.001**	0.000	0.000	0.000	**0.995**	0.001	**0.000**	0.001	0.003	**0.891**	0.001	0.002	**0.018**	**0.088**
TB6	0.999	**0.001**	0.000	0.000	0.000	**0.995**	0.001	**0.000**	0.001	0.003	**0.892**	0.001	0.002	**0.018**	**0.087**
TB7	0.999	**0.001**	0.000	0.000	0.000	**0.995**	0.001	**0.000**	0.001	0.003	**0.894**	0.001	0.001	**0.017**	**0.087**
TB8	0.999	**0.001**	0.000	0.000	0.000	**0.995**	0.001	**0.000**	0.001	0.004	**0.895**	0.001	0.001	**0.017**	**0.087**
EB1	0.992	**0.007**	0.000	0.000	0.000	**0.989**	0.007	**0.001**	0.001	0.003	**0.879**	0.006	0.007	**0.020**	**0.088**
EB2	0.980	**0.019**	0.001	0.000	0.000	**0.976**	0.019	**0.001**	0.001	0.003	**0.870**	0.017	0.007	**0.020**	**0.087**
EB3	0.949	**0.048**	0.003	0.000	0.001	**0.945**	0.048	**0.003**	0.001	0.004	**0.845**	0.042	0.008	**0.019**	**0.087**
EB4	0.876	**0.116**	0.006	0.000	0.002	**0.872**	0.116	**0.006**	0.001	0.005	**0.783**	0.102	0.010	**0.018**	**0.086**
LD1	0.998	**0.002**	0.000	0.000	0.000	**0.993**	0.002	**0.001**	0.001	0.003	**0.875**	0.002	0.009	**0.022**	**0.093**
LD2	0.999	**0.001**	0.000	0.000	0.000	**0.990**	0.004	**0.001**	0.001	0.004	**0.869**	0.001	0.010	**0.022**	**0.097**
LD3	0.999	**0.001**	0.000	0.000	0.000	**0.979**	0.013	**0.003**	0.001	0.005	**0.862**	0.001	0.012	**0.023**	**0.102**
LD4	1.000	**0.000**	0.000	0.000	0.000	**0.950**	0.039	**0.005**	0.001	0.006	**0.855**	0.000	0.014	**0.024**	**0.106**
LD5	1.000	**0.000**	0.000	0.000	0.000	**0.895**	0.090	**0.007**	0.001	0.006	**0.847**	0.000	0.016	**0.025**	**0.111**

Source: author's preparation.

In both credit types, the results obtained from the variable employment state (variable 3) show that this variable is irrelevant for this study. The probabilities of transition do not vary between the four categories of this variable.

The variable 4 (client with previous defaults) behaves has expected. For both products, clients that have never defaulted have a smaller probability of failing with their credit obligations and a slightly higher probability of recovering without entering in the PERSI program than clients with previous defaults. However, it is observed that in consumer credit, for the transition between the states 3 and 4 and the states 3 and 5 this variable has no impact. Testing this variable with a variable that considers the amount of time that a client is in the Bank could be a good complement to this observation, since the lack of defaults can be associated to the short period of time that a client has been registered in the Bank.

Finally, variable 5 (how long the account has been open until the reference date, client seniority) was important to include in the model since the dataset used in the study considers accounts open at different periods, differently from previous literatures (Leow & Crook, 2014) and (Régis et al., 2015).

In consumer credit accounts this variable does not seem to have any impact, and a reason can be due to the fact that these accounts have smaller maturities than housing credit products. In housing credit, recent accounts are less likely to recover from the PERSI program, however, are more likely to restructure the product.

Table 5.3 - Probabilities Housing Credit

Variable	Transitions														
	1-1	1-2	1-3	1-4	1-5	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	3-5
A1	0.998	**0.002**	0.000	0.000	0.000	**0.993**	0.002	**0.001**	0.003	0.001	**0.823**	0.002	0.009	**0.124**	**0.042**
A2	0.997	**0.002**	0.000	0.000	0.000	**0.992**	0.002	**0.001**	0.004	0.001	**0.807**	0.002	0.011	**0.133**	**0.047**
A3	0.997	**0.003**	0.000	0.000	0.000	**0.991**	0.002	**0.001**	0.004	0.002	**0.789**	0.002	0.014	**0.142**	**0.053**
A4	0.997	**0.003**	0.000	0.000	0.000	**0.989**	0.003	**0.001**	0.005	0.002	**0.771**	0.002	0.017	**0.151**	**0.060**
A5	0.996	**0.003**	0.000	0.000	0.000	**0.988**	0.003	**0.001**	0.006	0.002	**0.750**	0.002	0.021	**0.160**	**0.067**
A6	0.996	**0.003**	0.000	0.000	0.000	**0.986**	0.003	**0.002**	0.006	0.003	**0.729**	0.002	0.025	**0.169**	**0.075**
MS1	0.998	**0.002**	0.000	0.000	0.000	**0.993**	0.002	**0.000**	0.004	0.001	**0.854**	0.002	0.004	**0.107**	**0.033**
MS2	0.998	**0.002**	0.000	0.000	0.000	**0.992**	0.002	**0.000**	0.004	0.001	**0.868**	0.002	0.002	**0.098**	**0.030**
MS3	0.997	**0.002**	0.000	0.000	0.000	**0.991**	0.002	**0.000**	0.005	0.001	**0.880**	0.002	0.001	**0.090**	**0.027**
MS4	0.997	**0.002**	0.000	0.000	0.000	**0.991**	0.002	**0.000**	0.005	0.002	**0.891**	0.002	0.001	**0.082**	**0.024**
ES1	0.998	**0.002**	0.000	0.000	0.000	**0.994**	0.002	**0.000**	0.003	0.001	**0.841**	0.002	0.006	**0.114**	**0.037**
ES2	0.998	**0.002**	0.000	0.000	0.000	**0.994**	0.002	**0.000**	0.003	0.001	**0.844**	0.002	0.005	**0.112**	**0.037**
ES3	0.998	**0.002**	0.000	0.000	0.000	**0.995**	0.002	**0.000**	0.002	0.001	**0.847**	0.002	0.004	**0.110**	**0.037**
ES4	0.998	**0.002**	0.000	0.000	0.000	**0.995**	0.002	**0.000**	0.002	0.001	**0.849**	0.002	0.004	**0.109**	**0.037**
D1	0.998	**0.002**	0.000	0.000	0.000	**0.994**	0.002	**0.000**	0.003	0.001	**0.839**	0.002	0.007	**0.116**	**0.037**
D2	0.994	**0.005**	0.001	0.000	0.000	**0.988**	0.005	**0.001**	0.004	0.001	**0.810**	0.004	0.013	**0.131**	**0.042**
TB1	0.998	**0.002**	0.000	0.000	0.000	**0.994**	0.002	**0.000**	0.003	0.001	**0.853**	0.002	0.005	**0.102**	**0.038**
TB2	0.998	**0.002**	0.000	0.000	0.000	**0.994**	0.002	**0.000**	0.002	0.001	**0.867**	0.002	0.004	**0.089**	**0.039**
TB3	0.998	**0.002**	0.000	0.000	0.000	**0.995**	0.002	**0.000**	0.002	0.001	**0.878**	0.001	0.003	**0.078**	**0.040**
TB4	0.998	**0.002**	0.000	0.000	0.000	**0.995**	0.002	**0.000**	0.002	0.001	**0.888**	0.001	0.002	**0.068**	**0.040**
TB5	0.998	**0.002**	0.000	0.000	0.000	**0.995**	0.002	**0.000**	0.002	0.001	**0.897**	0.001	0.001	**0.060**	**0.041**
TB6	0.998	**0.001**	0.000	0.000	0.000	**0.995**	0.001	**0.000**	0.002	0.001	**0.904**	0.001	0.001	**0.052**	**0.042**
TB7	0.998	**0.001**	0.000	0.000	0.000	**0.996**	0.001	**0.000**	0.001	0.001	**0.911**	0.001	0.001	**0.045**	**0.042**
TB8	0.999	**0.001**	0.000	0.000	0.000	**0.996**	0.001	**0.000**	0.001	0.001	**0.917**	0.001	0.000	**0.039**	**0.043**
EB1	0.995	**0.004**	0.000	0.000	0.000	**0.991**	0.004	**0.001**	0.003	0.001	**0.837**	0.004	0.006	**0.117**	**0.036**
EB2	0.989	**0.009**	0.001	0.000	0.000	**0.985**	0.009	**0.001**	0.004	0.001	**0.833**	0.008	0.006	**0.118**	**0.035**
EB3	0.977	**0.020**	0.002	0.001	0.000	**0.973**	0.020	**0.002**	0.004	0.001	**0.824**	0.017	0.006	**0.119**	**0.034**
EB4	0.951	**0.043**	0.003	0.002	0.000	**0.947**	0.043	**0.004**	0.005	0.001	**0.803**	0.037	0.006	**0.121**	**0.033**
LD1	0.999	**0.001**	0.000	0.000	0.000	**0.994**	0.001	**0.000**	0.004	0.001	**0.831**	0.001	0.008	**0.122**	**0.039**
LD2	0.999	**0.001**	0.000	0.000	0.000	**0.992**	0.001	**0.001**	0.005	0.002	**0.822**	0.000	0.010	**0.127**	**0.040**
LD3	1.000	**0.000**	0.000	0.000	0.000	**0.989**	0.002	**0.002**	0.006	0.002	**0.813**	0.000	0.012	**0.133**	**0.042**
LD4	1.000	**0.000**	0.000	0.000	0.000	**0.977**	0.010	**0.003**	0.008	0.002	**0.803**	0.000	0.014	**0.139**	**0.043**
LD5	1.000	**0.000**	0.000	0.000	0.000	**0.947**	0.035	**0.006**	0.009	0.003	**0.793**	0.000	0.017	**0.145**	**0.045**

Source: author's preparation.

Variable 6 (percentage of months with excess balance) is the one that presents a higher impact on the probability of transiting from the state 1 to the state 2. As expected, the higher the percentage of months with excess balance is, the higher is the probability of the individual failing with his/her payments. On the other hand, this variable seems to have no impact in the transitions between states 3 and 4 and states 3 and 5.

Ultimately, the variable 7 (months since the last debit) only presents a slight impact in the transitions from the state 3. This variable behaves as expected, clients that had recent debits in their accounts are more likely to recover from the PERSI program and less likely to default.

5.2. MODEL VALIDATION

The validation was performed using independent samples of 15.000 unique accounts of each credit product type, and using transition probability matrix from 6 months (after the reference date) to one year. These samples were separately stored in the beginning of the study in order to allow the performance of an independent validation analysis.

The probability of an account moving from one state to another or remaining in the same is related with the characteristics and credit behaviour of the client that are established in the covariates. Since this study is only using categorical variables it is possible to obtain a finite number of sets with the outputs of the 7 variables, this will produce also a finite number of transition probability matrixes which will be used to predict the transitions that can occur one year after the reference date.

By observing the accuracy ratios that consider all the predictions without acknowledging the transitions that occurred, expressed in Table 10, it is possible to conclude that almost all accounts where correctly predicted in both models, consumer and housing credit. However, when looking at the ratios obtained in each transition is possible to observe that in all transitions the ratios are quite different from 1, except when the account remains in the state 1, in this case the ratio is very close to 1. Since the amount of accounts in the first transition (from state 1 to state 1) is truly high, around 99.5% of the housing credit accounts and 98.8% of the consumer credit accounts, the difference between observed and predicted accounts will not cause much impact, meanwhile, the other transitions have much less accounts consequently, any difference between observed and predicted will cause a big impact.

Table 5.4 - Accuracy Ratios

Housing Credit					
from\to:	**1**	**2**	**3**	**4**	**5**
1	1.00	1.83	0.74	2.59	2.27
2	2.82	0.04	0.06	0.36	0.00
3	1.98	0.00	0.00	0.48	0.15
4	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00

Consumer Credit					
from\to:	**1**	**2**	**3**	**4**	**5**
1	0.99	2.19	0.71	0.00	4.72
2	2.61	0.11	0.08	0.00	0.51
3	1.64	0.00	0.06	0.14	0.36
4	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00

Accuracy Ratio:	0.995

Accuracy Ratio:	0.988

Source: author's preparation.

Taking into consideration the conclusion obtained from the accuracy ratios, and observing the results in the Table 10, Pearson Chi-square test results, is viable to conclude that the model doesn't fit well. Not only the accounts are all concentrated in the transition from state 1 to state 1, but also the Pearson statistic T is quite higher than the p-value $\chi^2_{\mathrm{n-p}}$.

Table 5.5 - Pearson Test

Housing Credit					
from\to:	**1**	**2**	**3**	**4**	**5**
1	0.07	14.32	0.93	0.97	0.71
2	21.18	649.42	43.55	1.14	0.30
3	0.97	0.22	0.04	0.55	9.83
4	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00

Consumer Credit					
from\to:	**1**	**2**	**3**	**4**	**5**
1	0.58	46.93	2.03	0.91	2.93
2	40.72	468.90	85.58	0.20	0.96
3	1.00	0.52	13.85	5.16	2.30
4	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00

T:	744.21
$\chi^2_{0.05:\,20}$:	10.85

T:	672.57
$\chi^2_{0.05:20}$:	10.85

Source: author's preparation.

6. CONCLUSIONS

By using a large dataset of consumer and housing credits, this work studied the behaviour of accounts of these credit products with a multi-state Markov model in continuous time. This model has the advantage of allowing the use of covariates and providing probabilities of all possible transitions depending on the time.

Five states were defined to study the behaviour of these credit products: solvent, payment failure, PERSI, restructured product and default. From the interpretation of the Portuguese regulation DL no 227/2012 (BoP, 2012), when a client's account enters in the PERSI program only one of three possible transitions can occur: (i) the client solves is debts and returns to the solvent state, (ii) the credit product is restructured in order for the client to be able to comply with his credit obligations or (iii) the client and the financial institution don't reach to an agreement and the clients enters into a state of default.

With the estimated intensity transitions for both models, consumer and housing credit models, it was possible to observe and analyse the transitions probabilities that each covariate output would origin, assuming the mean of all other covariates. Due to the fact that the observations were highly concentrated in the state 1, the results obtained were not conclusive. This means that no statistical evidence was found on the use of the designed multi-state Markov model to predict transitions between the remaining states.

Nevertheless, all variables behaved as expected apart from the variable employment state, which did not present any impact on the transition probabilities.

Regarding the validation of the models, the analyses on the observed and predicted transitions of two random independent samples, one for each type of account, were performed using accuracy ratios and a Pearson Chi-square test. The validation allowed to conclude, from both analysis, that the overall accounts are being correctly predicted due to the amount of accounts remaining in the state 1 after six months. The other transitions have considerably less occurrences and, as a result, any difference causes a significant impact in both the accuracy ratios and the Pearson Chi-square test. The model failed this test with an estimated statistic much higher than the p-value, meaning that statiscally there was no evidence to support the predictability skills of the designed multi-state Markov model.

Certain limitations to this study may have had an impact in the results and to some extent even compromised the applicability of the multi-state Markov model to study the behaviour of consumer and housing credit. For instance, it is expected for the number of observations in the transitions between the worst states (PERSI, restructured product and default) to be much lower than the transition between state 1 and 2 (solvent and payment failure). However, the design of the model states and definition of default, aggravated this fact resulting in abnormally scarce observations in the transitions between the worst states. The design of the model states was performed according to the Portuguese regulation DL no 227/2012 (BoP, 2012), as mentioned above, which defines a program to regularize default situations and a process to prevent payment failure. The use of this regulation facilitated the identification of concrete states where a client can be before entering the state of default/restructured product. In contrast, the use of the regulation created rigid criteria for the clients to enter in these absorbent states. This is due to the fact that the purpose of the processes created by the regulation was to prevent and regularize credit payment failures as promptly as possible.

Additionally, the regulation became effective in 2013, constraining the study to a short period of observation.

This study was additionally limited to the data kindly provided by the financial institution, which made available a restricted number of variables. Potentially relevant is the fact that it was not possible to develop an analysis based on different explanatory variables to guarantee that the most relevant variables were selected to be part of the model.

Finally, the multi-state Markov model applied, by use of the software R, presents a technical limitation that can lead to the loss of observations. This limitation relies on the necessity of the model to be applied to periodic observations instead of a dataset containing only event (state transition) observations. The periodic observations may not capture the real date of the state transition and creates a time gap between each observation. Moreover, the addition of variables to this model significantly increases the optimization complexity, which subsequently leads to extremely long periods of time to compute the model.

Regardless of the inconclusive results of this study, the use of multi-state models to analyse the behaviour of credit products' lifecycle has much potential for further development, particularly in light of the importance and impact that credit default events can have in the financial stability of countries.

Suggestions for future research include (i) a profound analysis of the assumptions of the multi-state Markov model and identification of the best multi-state model to analyse the behaviour of these products; (ii) the analysis of different explanatory variables that present the model with the best performance results and, (iii) the incorporation of more years of data history with the PERSI program.

Based on the data limitation handled in this study, it is also important to note that future research should guarantee that every state contains reasonable amount of observations, without disrupting the repetitiveness of the population, in order to produce conclusive and accurate results.

Finally, an interesting study to access the effectiveness of the regulation implemented by BoP would be the comparison between the behaviour of clients before and after the PERSI program was implemented.

In conclusion, this study can be a starting point to the study of credit account's behaviour using multi-state models in the Portuguese context.

7. BIBLIOGRAPHY

Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance, 23(4), p.589. doi:10.2307/2978933

Altman, E., Marco, G. and Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). Journal of Banking & Finance, 18(3), pp.505-529. doi:10.1016/0378-4266(94)90007-8

Altman, E. and Saunders, A. (1997). Credit risk measurement: Developments over the last 20 years. *Journal of Banking & Finance*, 21(11-12), pp.1721-1742. doi:10.1016/s0378-4266(97)00036-8

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C., & Schuermann, T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance, 26*(2-3), 445-474. doi:10.1016/s0378-4266(01)00229-1

Bank of International Settlements (2005). International Convergence of Capital Measurements and Capital Standards: *A revised framework*. doi:bis.org/publ/bcbs118.pdf

Bellotti, T. and Crook, J. (2007). Credit scoring with macroeconomic variables using survival analysis. Edinburgh: University of Edinburgh Management School.

Beran, J. and Djaïdja, A. (2007). Credit risk modeling based on survival analysis with immunes. Statistical Methodology, 4(3), pp.251-276. doi:10.1016/j.stamet.2006.09.001

Betancourt, Luis. 1999. Using Markov Chains to Estimate Losses from a Portfolio of Mortgages. Review of Quantitative Finance and Accounting, Vol.12, No. 3, 303-317.

Boser, B. E., Guyon, I. M., Vapnik, V. N. (1992). A Training Algorithm for Optimal Margin Classifiers. Paper presented at 5th Annual ACM Workshop on Computational Learning Theory. New York, NY: ACM Press

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.I. (1984). Classification and regression trees. Belmont, Calif.: Wadsworth.

Brown, I. and Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 39(3), pp.3446-3453. doi:10.1016/j.eswa.2011.09.033

Chamboko, R., Bravo, J. M. (2016). On the modelling of prognosis from delinquency to normal performance on retail consumer loans. Risk Management, December 2016, Volume 18, Issue 4, pp 264–287.

Chatterjee, S. and Barcun, S. (1970). A Nonparametric Approach to Credit Screening. Journal of the American Statistical Association, 65(329), pp.150-154. doi:10.1080/01621459.1970.10481068

Chen, H., & Chen, Y. (2010). A comparative study of discrimination methods for credit scoring. Paper presented at the Computers and Industrial Engineering (CIE), 2010 40th International Conference on.

Datschetzky, D., Kuo, Y. D., Tscherteu, A., Hudetz, T., Hauser-Rethaller, U. (2005). Rating Models and Validation, Guidelines on Credit Risk Management. Vienna: Oesterreichische Nationalbank (OeNB), Austrian Financial Market Authority (FMA)

Decreto Lei no 227/12 de 25 de outubro do Ministério da Economia e do Emprego. Diário da República: I série, No 207 (2012). doi:dre.pt/application/file/192484

Einarsson, A. I. (2008). Credit Risk Modeling (Doctoral dissertation). Retrieved from http://etd.dtu.dk/thesis/224338/ep08_100.pdf

Feldman, D. and Gross, S. (2004). Mortgage Default: Classification Trees Analysis. *SSRN Electronic Journal.* doi:10.2139/ssrn.659881

Ferguson, N., Datta, S., & Brock, G. (2012). msSurv: An R Package for Nonparametric Estimation of Multistate Models. *Journal of Statistical Software, 50*(14), 1-24.

Fisher, R. (1936). THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. Annals of Eugenics, 7(2), pp.179-188. doi:10.1111/j.1469-1809.1936.tb02137.x

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19: 1. doi:10.1214/aos/1176347963

Frydman, H., Altman, E. and Kao, D. (1985). Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. *The Journal of Finance*, 40(1), pp.269-291.

Frydman, H., & Schuermann, T. (2008). Credit rating dynamics and Markov mixture models. *Journal of* Banking & Finance, 32(6), 1062-1075. doi:10.1016/j.jbankfin.2007.09.013

Grimshaw, S. and Alexander, W. (2010). Markov chain models for delinquency: Transition matrix estimation and forecasting. Applied Stochastic Models in Business and Industry, 27(3), pp.267-279. doi:10.1002/asmb.827

Halim, S., & Humira, Y. V. (2014). Credit Scoring Modeling. *Jurnal Teknik Industri, 16*(1), 17-23. doi:10.9744/jti.16.1.17-24

Hand, D. and Henley, W. (1997). Statistical Classification Methods in Consumer Credit Scoring: a Review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), pp.523-541. doi:10.1111/j.1467-985x.1997.00078.x

Harris, T. (2015). Credit scoring using the clustered support vector machine. Expert Systems with Applications, 42(2), pp.741-750. doi:10.1016/j.eswa.2014.08.029

Henley, W. and Hand, D. (1996). A k-Nearest-Neighbour Classifier for Assessing Consumer Credit Risk. *The Statistician*, 45(1), p.77. doi:10.2307/2348414

Hooman, A., Marthandan, G. and Karamizadeh, S. (2013). Statistical and Data Mining Methods in Credit Scoring. *SSRN Electronic Journal*. doi:10.2139/ssrn.2312067

Hosmer, D. W., Lemeshow, S. (2013). Applied logistic regression (3rd ed.). Hoboken, New Jersey: *John Wiley & Sons*.

Hougaard, P. (1999). Multi-state models: A review. *Lifetime Data Analysis, 5*(3), 239-264. doi:10.1023/a:1009672031531

Hurd, T. and Kuznetsov, A. (2007). Affine Markov chain model of multifirm credit migration. The Journal of Credit Risk, 3(1), pp.3-29. doi:10.21314/jcr.2007.058

Leong, C. (2015). Credit Risk Scoring with Bayesian Network Models. Computational Economics, 47(3), pp.423-446. doi:10.1007/s10614-015-9505-8

Leow, M. and Crook, J. (2014). Intensity models and transition probabilities for credit card loan delinquencies. European Journal of Operational Research, 236(2), pp.685-694. doi:10.1016/j.ejor.2013.12.026

Jackson, C. H. (2011). Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software, 38*(8), 1-28.

Jackson, C. H., Sharples, L. D., Thompson, S. G., Duffy, S. W., & Couto, E. (2003). Multistate Markov models for disease progression with classification error. *Journal of the Royal Statistical Society Series D-the Statistician, 52*, 193-209. doi:10.1111/1467-9884.00351

Jarrow, R., Lando, D. and Turnbull, S. (1995). A Markov model for the term structure of credit risk spreads. [Copenhagen]: Institute of Mathematical Statistics, University of Copenhagen.

Kalbfleisch, J. D., & Lawless, J. F. (1985). THE ANALYSIS OF PANEL DATA UNDER A MARKOV ASSUMPTION. *Journal of the American Statistical Association, 80*(392), 863-871. doi:10.2307/2288545

Kim, H. and Sohn, S. (2010). Support vector machines for default prediction of SMEs based on technology credit. European Journal of Operational Research, 201(3), pp.838-846. doi:10.1016/j.ejor.2009.03.036

Koopman, S. J., Lucas, A., & Monteiro, A. (2008). The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics, 142*(1), 399-424. doi:10.1016/j.jeconom.2007.07.001

Lahsasna, A., Ainon, R. N., & Teh, Y. W. (2010). Credit Scoring Models Using Soft Computing Methods: A Survey. The International Arab Journal of Information Technology, 7(2), 115-123.

Malik, M., & Thomas, L. C. (2012). Transition matrix models of consumer credit ratings. *International Journal of Forecasting, 28*(1), 261-272. doi:10.1016/j.ijforecast.2011.01.007

Martin, D. (1977). Early warning of bank failure. Journal of Banking & Finance, 1(3), pp.249-276. doi:10.1016/0378-4266(77)90022-x

Mays, E. (2001). Handbook of credit scoring (pp. 71-104). Chicago: *Glenlake Publishing Company*

Meira-Machado, L. (2011). INFERENCE FOR NON-MARKOV MULTI-STATE MODELS: AN OVERVIEW.*Revstat-Statistical Journal, 9*(1), 83-+.

Oreski, S. (2014). Hybrid Techniques of Combinatorial Optimization with Application to Retail Credit Risk Assessment. Artificial Intelligence and Applications, 2014(1), pp.21-43. doi:10.15764/aia.2014.01002

Régis, D. E., & Artes, R. (2016). Using multi-state markov models to identify credit card risk. *Production, 26*(2), 330-344. doi:10.1590/0103-6513.160814

Sabzevari, H. , Soleymani, M. , & Noorbakhsh, A. (2007). A comparison between statistical and data mining methods for credit scoring in case of limited available data. In Proceedings of the 3rd CRC Credit-scoring Conference, Edinburgh, UK

So, M. M. C., & Thomas, L. C. (2011). Modelling the profitability of credit cards by Markov decision processes. *European Journal of Operational Research, 212*(1), 123-130. doi:10.1016/j.ejor.2011.01.023

Tam, K. and Kiang, M. (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. Management Science, 38(7), pp.926-947.doi:10.1287/mnsc.38.7.926

Thomas, L. C. (2009). Consumer credit models: pricing, profit and portfolio (pp. 79-84). New York, NY: *Oxford University Press*

West, R. (1985). A factor-analytic approach to bank condition. *Journal of Banking & Finance,* 9(2), pp.253-266. doi:10.1016/0378-4266(85)90021-4

Wilson, T. C. (1998). Measuring and managing credit portfolio risk. *Risk Management, Econometrics and Neural Networks*, 259-306.

Zhang, D., Hifi, M., Chen, Q., & Ye, W. (2008). *A hybrid credit scoring model based on genetic programming and support vector machines*. Paper presented at the Natural Computation, 2008. ICNC'08. Fourth International Conference on. Jinan, China. doi:10.1109/ICNC.2008.205