Evangelista Régis, Daniel; Artes, Rinaldo
Using multi-state markov models to identify credit card risk
Production, vol. 26, núm. 2, abril-junio, 2016, pp. 330-344
Associação Brasileira de Engenharia de Produção
São Paulo, Brasil

Using multi-state markov models to identify credit card risk

Daniel Evangelista Régis[a]*, Rinaldo Artes[a]**

[a]Insper Instituto de Ensino e Pesquisa, São Paulo, SP, Brasil

*edanielregis@gmail.com, **rinaldoa@insper.edu.br

Abstract

The main interest of this work is to analyze the application of multi-state Markov models to evaluate credit card risk by investigating the characteristics of different state transitions in client-institution relationships over time, thereby generating score models for various purposes. We also used logistic regression models to compare the results with those obtained using multi-state Markov models. The models were applied to an actual database of a Brazilian financial institution. In this application, multi-state Markov models performed better than logistic regression models in predicting default risk, and logistic regression models performed better in predicting cancellation risk.

Keywords

Credit scoring. Survival analysis. Multi-state Markov models. Credit cards. Markov processes.

1. Introduction

According to Wang & Ikeda (2004), the Brazilian credit card market has been in a maturation phase since the demand for credit boom, after the economy stabilized in 1994 with the implementation of the Real Plan. The deployment of the Brazilian Payment System (SPB) in 2002 has also affected the credit card market by leading to a strong migration from checks to electronic payment system.

The economic stability demanded the use of advanced statistical methodologies, such as logistic regressions, discriminant analysis, survey analysis, decision trees, Bayesian inference, and neural networks, to evaluate credit risk.

In logistic regression, one estimates the probability of the transition from state A, non-defaulting, to state B, defaulting, over a certain period of time (Hosmer & Lemeshow, 2000). Other kinds of transitions are not considered. It is known that there are several possible states in the relationship between clients and financial companies, such as *non-defaulting without revolving credit use, non-defaulting with revolving credit use, in delay, voluntary cancellation and default.* Customers pass through these states over time, and this is a characteristic of recurrent events (Hosmer et al., 2008; Paes & Lima, 2004). The existence of several possible states characterizes multi-state events (Hougaard, 1999).

Trench et al. (2003) developed a markovian decision process in order to guide a bank to determine price and credit lines for credit card holders in order to maximize its profits. So & Thomas (2010) used markov chain modeling transition probabilities in logistic models in order to evaluate credit risk of credit card portfolios.

The motivation for this work comes from the interest in taking advantage of these recurrence characteristics to generate transition probability matrices between several possible states over time. This is why a Multi-state Markov model was estimated (Jackson, 2007) and its performance was compared with models developed using logistic regressions.

In section 2, we give a brief introduction to the typology of credit risk forecasting models. In section 3, we describe the Multi-state Markov model that was used and other statistical techniques. The database is presented in section 4, and the obtained results from the models are presented in section 5. Finally, we weave our final considerations in section 6.

Régis, D. E. et al.
Using multi-state markov ... credit card risk. Production, 26(2), 330-344, abr./jun. 2016

PRODUCTION

331

2. Typology of credit risk forecasting models

Due to the expansion of the credit card market, companies have naturally become more and more concerned with the level of default. Application scoring models are widely used to evaluate the credit risk of a new account, and behavior scoring models are used in the credit risk management of clients who have already acquired a product.

Maintaining a long relationship with credit card customers is also of primary importance for these companies once, in general, acquiring a new customer is more expensive than retaining an existing one (Van den Poel & Larivière, 2004). Anti-attrition scoring is used to promote customer retention.

2.1. Credit scoring models

According to Araujo & Carmona (2007) apud Lewis (1992, p. XV),

> Credit scoring models are systems that attach scores to credit decision variables by the application of statistical techniques. These models aim to identify characteristics that can differentiate between good and bad credit.

The development of credit scoring models requires the use of statistical techniques such as logistic regression, discriminant analysis, survey analysis, decision trees, Bayesian inference and neural networks in addition to practical knowledge of the type of customer to be analyzed. Thomas (2009) and Finlay (2010) present a good overview of this subject.

Among the credit scoring models we may cite the: application scoring, used to decide whether to grant credit to a new customer, and behavior scoring models, that use information about the relationship between the client and the institution to evaluate his risk (e.g. Araujo & Carmona, 2007; Thomas, 2000). Thomas et al. (2002) describe all the necessary stages to develop an application scoring model.

Kuhn (2009) affirms that behavior scoring models help companies manage their relationships with customers who have already acquired a product and are used as an important tool to determine credit limits and which new products to offer. Behavior scoring models are mainly based on the customer's shopping or payment patterns and therefore have a much higher discriminatory power than application scoring models (Hoper & Lewis, 1992). Hoper & Lewis (1992) describe how a behavior scoring model is usually used. Blackwell & Sykes (1992) describe how behavior scoring models can be used to determine the proper credit limit to be assigned to a customer. Thomas et al. (2001) describe how to create behavior scoring models using Markov chains, where the customer is classified in a state according to some variables and then the probability of the customer becoming a debtor is estimated.

Kuhn (2009) says that application and behavior scoring models in Brazil have obtained significant gains in performance by using credit bureau information from firms such as Serasa-Experian. These models use customer behavior information in addition to the company's internal sources of information about a customer.

Credit scoring models are usually used to divide the portfolio into score classes. These classes define groups of customers with similar risk levels, allowing for the creation of specific policies for each customer group.

2.2. Anti-attrition scoring models

Anti-attrition scoring models are developed to predict the risk of a client be in a situation that can put in risk his relationship with the company. The methodologies used to develop anti-attrition scoring models are basically the same as those used to develop credit scoring models.

3. Methodology

In this work, we wish to forecast a credit card owner's future state based on the owner's current state and profile information. We present the logistic regression model and the Multi-state Markov model in this chapter as well as the performance measures used to compare the results

3.1. Logistic regression model

The logistic is one of the most popular models used in risk modeling (e.g. Thomas, 2009, p. 79). Let Y_i, $i = 1,...,n$, be a set of independent dummy variables assuming the value 1 if the costumer i is in default and 0, if not and x_i be a P-dimensional vector – usually the first column is an unitary vector (associated to an intercept parameter) and the remain columns are the values of the observed covariates associated with the customer i, The logistic regression model (Menard, 2010; Hosmer & Lemeshow 2000) is given by

$$E(Y_i \mid x_i) = \pi_i = \pi(x_i) = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} \qquad (1)$$

332

PRODUCTION

Régis, D. E. et al.
Using multi-state markov ... credit card risk. Production, 26(2), 330-344, abr./jun. 2016

where β is a parametric vector. This model can also be written as

$$ln\left(\frac{\pi_i}{1-\pi_i}\right) = x_i^T \beta \tag{2}$$

The model's parameters may be estimated by the maximization of the likelihood function, given by

$$l(\beta) = \prod_{i=1}^{n} \pi_i^{Y_i} \left(1-\pi_i\right)^{1-Y_i} \tag{3}$$

The inference for the parameters is based on the asymptotic properties of the maximum likelihood estimators. Under general regularity conditions, the estimators are consistent and asymptotically normal. In this paper, Wald tests were used to assess the significance of the parameters. Theoretical details may be found in Hosmer & Lemeshow (2000) and Menard (2010).

In credit risk modelling the amount of clients in default is, in general, much lower than the amount of good clients. This may lead to an underestimation of the probability of default. King & Zeng (2001) propose a correction in the intercept estimator in order to eliminate this bias. In credit scoring context, however, this problem is secondary, once, in general, the analyst intention is to sort the clients according to their default risk; what may be obtained from the model without correction.

Thomas (2009) introduces the logistic regression in the context of credit risk modelling and presents examples of its implementation.

3.2. Survival analysis concepts

Survival analysis is a set of statistical procedures for modeling the time of occurrence of events of interest (failure times). For various reasons, there are situations in which the failure time does not occur during the observation period; these situations are called censored observations – censored failure times (Hosmer et al., 2008; David et al., 2008; Andreeva, 2006).

Consider T the variable that indicates the failure time; the instantaneous failure rate at time t is conditioned on its survival at this time (failure risk at instant t) as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t} \tag{4}$$

The Cox semi-parametric model (Hosmer et al., 2008; Andreeva, 2006), or proportional hazard model, allows the inclusion of the effects of a customer's probability of failure, whether these characteristics

are used as an explanatory variable, or covariates of the response variable.

The hazard function, assigned to element i of the sample is given by $\lambda(t) = \lambda_0(t) \exp(x_i^T \beta)$, where λ_0 is a non-negative function called the baseline function. Note that $\lambda(t) = \lambda_o(t)$ when $x_i = 0$, where x_i is a fixed vector of covariates. The main idea of Cox model is to separate the effect of the covariates from the effect of time in the hazard function. The risk function model is called semi-parametric because only the covariate effects are treated parametrically. Details of the estimation processes may be found in Hosmer et al. (2008).

3.3. Multi-state markov model

We are assuming a situation where credit card owners can assume different states (for example, non-defaulting, in delay, defaulting) through time. Some of these states are temporary, i.e. the individual can leave this state at some time (for instance, an individual in delay can go back to a non-defaulting situation) and others are absorbent, i.e. once in this state, the individual cannot migrate to another state (for instance, a defaulting individual leaves the individual database and cannot take on another state). The Multi-state Markov model assumes that the transition probabilities among states depend only on the time between transitions and on the covariates (which are eventually time dependent) associated to the individuals. Details about the theoretical development of these models can be found in Kalbfleisch & Lawless (1985), Kay (1986), Jackson et al. (2003) and Jackson (2007).

Figure 1 represents a situation where there are three transitory states. The arrows indicate that the direct transition between the states is possible.

Consider $E_i(t)$, the assumed state of individual i, $i = 1, ..., n$, at the instant $t = 1, ..., \tau$. Admit the existence of K possible states. The probability of an individual i migrating from an r state to an s state in an interval Δt is given by

$$p_{irs}(\Delta t) = P\big(E_i(t + \Delta t) = s \mid E_i(t) = r\big) \tag{5}$$

The transition intensity among states r and s is defined by

$$q_{irs}(t) = \lim_{\Delta t \to 0} \frac{P(E_i(t + \Delta t) = s \mid E_i(t) = r)}{\Delta t} \tag{6}$$

We can observe that the transition intensity, (6), resembles the $\lambda(t)$ given in (4), that is the instantaneous failure rate at time t is conditioned on its survival until time t. It is possible for us to interpret this transition intensity as the instantaneous risk of an individual

Régis, D. E. et al.
Using multi-state markov ... credit card risk. Production, 26(2), 330-344, abr./jun. 2016

PRODUCTION

333

migrating from state r to state s where the failure, in this case, is the migration to state s.

The matrix $Q_i = [q_{irs}]_{K \times K}$ is called the intensity matrix and indicates the possible transitions of state (see Figure 1). We define $q_{irr} = -\sum_{s \neq r} q_{irs}$, except in the case of the absorbent state where $q_{irr} = 0$. In this way it is possible to show that the transition probabilities (5) are obtained through the components of the $P_i(\Delta t) = \exp(Q_i \Delta t)$ matrix (see details in this section's references and in Cox & Miller, 1965).

Let $x_i(t)$ be the covariate vector, observed for individual i in time t. Marshall & Jones (1995) model the transition intensity q_{irs} as

$$q_{irs}(x_i(t)) = q_{rs}^{(0)} \exp(\beta_{rs}^T x_i(t)) \tag{7}$$

where $x_i(t)$ is the p-dimensional vector with the values of the observed explanatory variables at instant t for individual i, $q_{rs}^{(0)}$ is the baseline for the transition r-s, and β_{rs}^T is the estimator associated with the variable $x_i(t)$ for the transition r-s. Analogous to the baseline function of the Cox's model, $q_{rs}^{(0)}$ is the expected value of the intensity function when the covariate vector is zero.

The likelihood function for the parameters, including $q_{rs}^{(0)}$, under the assumption of independence among the individuals, is given by

$$L = \prod_{i=1}^{n} \prod_{k=1}^{\tau-1} P_{E_i(t_k)E_i(t_{k+1})} \tag{8}$$

under general regularity conditions, the asymptotic distribution of the maximum likelihood estimators are asymptotically normal. The estimates of the probabilities $P_{E_i(t_k)E_i(t_{k+1})}$ are obtained from Q_i matrix. Wald tests for evaluating the parameters significance were developed by Marshall & Jones (1995). These tests are based on the asymptotic distribution of the maximum likelihood estimators.
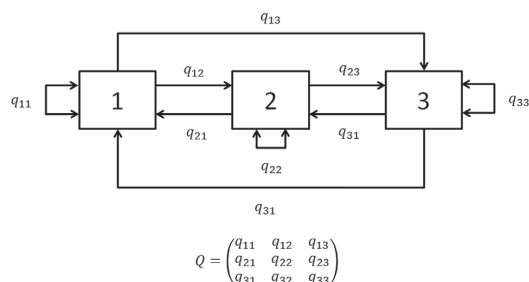
Transitions among the following states over time are considered in this work:

States with recurrence characteristics:

1 – In compliance (IC): the customer paid the total amount owed.

2 - Revolving (RE): the customer paid part of the amount owed, i.e. some value between the minimum payment and the total amount owed. In this case, the customer has a normal relationship with the credit card company.

3 – In delay (ID): the customer did not even pay the minimum portion of the amount owed.

Absorbent states:

4 - Voluntary cancellation (VC): credit card cancellation initiated by the customer.

5 - Default (DE): credit card cancellation initiated by the credit card company due to default. In this study, the occurrence of three consecutive delays characterizes default.

In Figure 2 the possible transitions and the intensity matrix structure are illustrated. The possible transitions for customers that are in the *in compliance* state are: *in compliance* to *revolving*, *in compliance* to *in delay* and *in compliance* to *voluntary cancellation*. The transition from the *in compliance* state directly to the *default* state is not possible, since it is only possible to go to a *default* state from an *in delay* state in this study.

The possible transitions from the *revolving* state are: *revolving* to *in compliance*, *revolving* to *in delay* and *revolving* to *voluntary cancellation*. The *in delay* state is represented by the letter A. The possible transitions for customers that are *in delay* are: *in delay* to *in compliance*, *in delay* to *revolving*, *in delay* to *voluntary cancellation* and *in delay* to

3.4. Transitions in credit card risk

$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix}$$

Figure 1. A multi-state model and transition intensity matrix. Source: the authors.
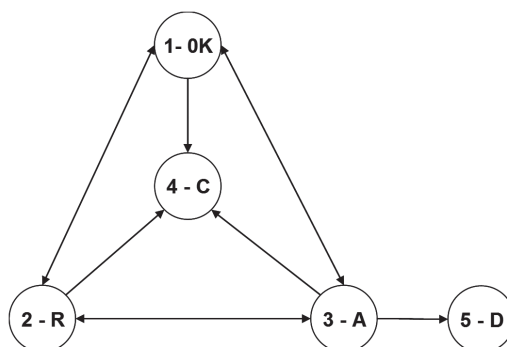
Figure 2. Multi-state model for credit card risk and transition intensity matrix. Source: the authors.

334 | **PRODUCTION**

Régis, D. E. et al.
Using multi-state markov … credit card risk. Production, 26(2), 330-344, abr./jun. 2016

default. An important observation is that the customer cannot stay in the *in delay* state for more than two consecutive months because this is considered to be *default* in this study. For customers in the *default* state, the only possibility is to remain in this state for all subsequent months, given that this is an absorbent state. The same is true for customers who are in the *voluntary cancellation* state.

3.5. Performance measures

There are several methods to measure and compare the performance of models of credit scoring (e.g. Thomas, 2009). In this section, we present the statistics of Kolmogorov-Smirnov, the Gini coefficient and graphs for visual inspection. In Thomas et al. (2005), Thomas (2009), Abdou & Pointon (2011), Chi & Hsu (2012) and Gupta et al. (2014) one may find an explanation about these techniques and applications in a credit risk modelling context.

3.5.1. Kolmogorov-Smirnov statistics

The Kolmogorov-Smirnov (KS) statistics (Thomas, 2009) is used to compare the score distribution of good and bad clients. A good model will predominantly produce high scores for good clients and low scores for bad clients. Let s be the value of a score; define $F_g(s) = \dfrac{\#\,good\ clientes\ with\ score \leq s}{n_g}$ and $F_b(s) = \dfrac{\#\,bad\ clientes\ with\ score \leq s}{n_b}$, where n_g and n_b are, respectively the number of good and bad clients. In a good model is expected that $F_b(s)$ will approach 1 faster than $F_g(s)$. The faster the growth of $F_b(s)$ and the slower the growth of $F_g(s)$, the better is the model. The KS statistics is defined as the greatest distance between $Fb(s)$ and $Fg(s)$ (Figure 3), ranging from 0 to 1 and the closer to 1, the better the performance of the model is.
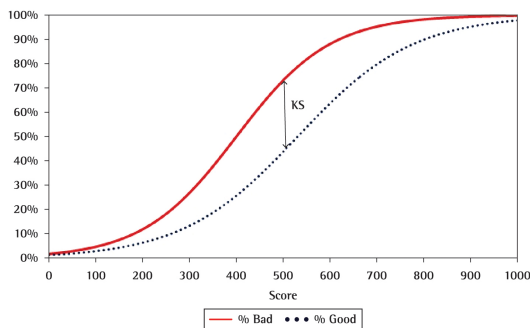
$$KS = \max_{0 \leq x \leq 1}\left\{ F_b(s) - F_g(s) \right\} \tag{9}$$

3.5.2. Gini coefficient

Assume that a client is classified as bad if its score is lower or equal to c, and as good, otherwise. In this case $F_b(c)$ is the proportion of bad clients that are correctly identified and $1 - F_g(c)$ is the proportion of good clients that are correctly classified – $F_g(c)$ is the proportion of "false positive" detected by the model using the value c as cutpoint. The ROC curve (Receiver Operating Curve) is a graph with $F_g(c)$ in abscisses and $F_b(c)$ in the y-axis , for ordered values of c (Thomas, 2009). Figure 4 is an example of this graph. If the model produces a good classification rule, it would be expected that $F_b(c)$ would be close to one and $F_g(c)$ close to zero. In this case the ROC curve would be close to the line $y = 1$. When $F_b(c) \cong F_g(c), \forall c$ the model produces a classification that is almost random, in other words, it is a bad model.

An indicator of the quality of the model is the area under the ROC curve: $AUROC$ $(0 \leq AUROC \leq 1)$, the closer to one the better is the model. The Gini coefficient is defined as

$$Gini = 2\,AUROC - 1 \tag{10}$$

Good models will produce higher values of *Gini*.

3.5.3. Back Test Graphs

A simple but highly efficient technique to evaluate a model performance is to graphically verify the ordering of scores in relation to the response variable. The population is divided into percentiles of the scores predicted by the model and the percentage of bad clients in each percentile is verified. It is expected that a good model will present a higher percentage of bad clients in the percentiles of lower score and a small percentage of bad clients in percentiles of higher score, with the percentage of defaulters decreasing monotonically from the lowest range to the highest range, as shown in Figure 5.
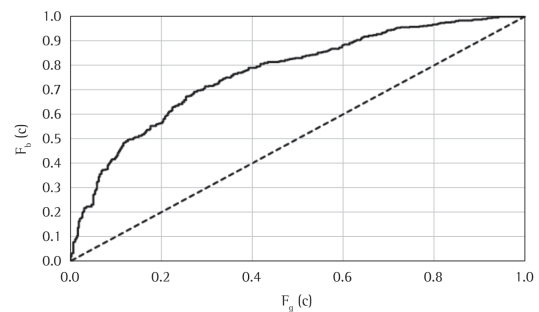

Figure 3. Example of a KS Graph. Source: the authors.

Figure 4. Example of a ROC Curve. Source: the authors.

Régis, D. E. et al.
Using multi-state markov ... credit card risk. Production, 26(2), 330-344, abr./jun. 2016

PRODUCTION

335

4. Database

The data analyzed in this study come from a huge Brazilian financial company that operates in the credit card market. For reasons of confidentiality, a transformed portfolio was generated. This transformed portfolio does not reflect this company's indices of revolving credit, delays, cancellations and defaults.

To apply the Multi-state Markov model, a sample was generated containing the history of nineteen thousand clients, ten thousand of which were used to develop the model and nine thousand were used to validate the model, and they were selected based on the following criteria:

- Were offered credit initially between January 2003 and June 2004

- Had active credit cards in January 2005, having had them activated for at least six months

- As of January 2005 (the observation time) were in one of the following states: *in compliance, revolving,* or *in delay*

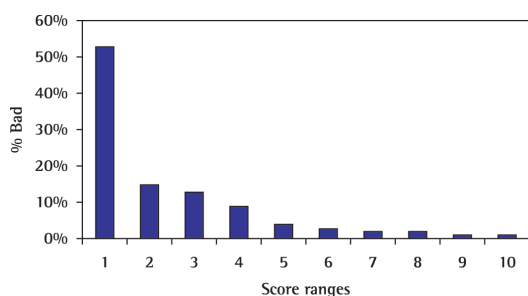The objective of this selection was to create a database with a population of clients who were

actively using the company's products and had a six to eighteen-month relationship with the company. Older clients could have been selected, but without cutoff criteria for the length of the relationship the transition study could have been compromised. Very old clients, for instance, can be "loyal clients", i.e., they have a smaller probability of cancelling credit cards on their own initiative or for default reasons. We could develop models for other portions of the population to get the predictive variable effects for clients with relationships of various durations.

To develop our model we decided to use a reduced number of behavior variables which indicated a strong relationship with the tendency of a client to present *cancellation, delay or default problems* and to use revolving credit.

The historical variables used are mainly based on: an arithmetic mean of twelve months, an exponential mean of six months (weighted mean with greater weights applied to newer observations, number of months using the credit card, number of months in a determined state, number of consecutive months, maximum usage or maximum amount of some metric that characterizes buying behavior, the percentage of limit used and the delay profile or the use of revolving credit.

To create historical variables, we used the twelve months before January 2005 and the observation of future transitions was carried out for the twelve months following January 2005, as is shown in Figure 6.

Seven explanatory variables were selected (six behavioral variables that were treated in continuous form and one cadastral variable was treated in a categorized form) using judgmental criteria and the CHAID algorithm (*Chi-squared Automatic Interaction Detection*), using, as a response variable, the state of
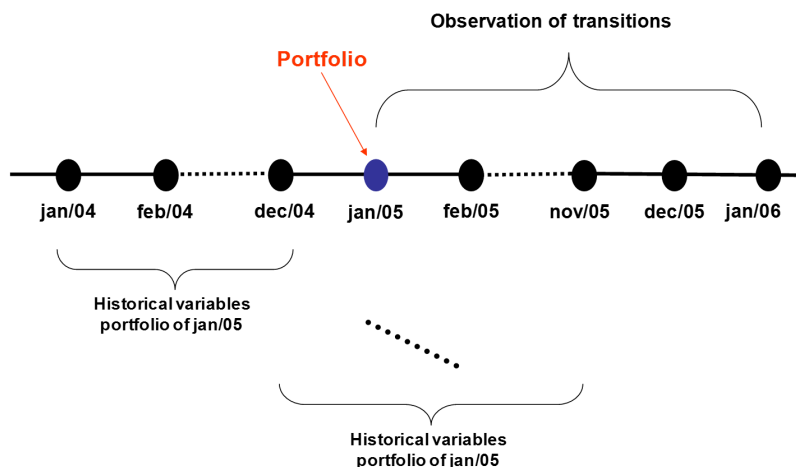

Figure 5. Example of a Back Test Graph. Source: the authors.

Figure 6. Database structure for the Multi-state Markov model. Source: the authors.

336 | PRODUCTION

Régis, D. E. et al.
Using multi-state markov ... credit card risk. Production, 26(2), 330-344, abr./jun. 2016

the client twelve months after the month of initial observation (cancelled, default, and others).

The selected variables are shown below:

Variable 1: variable divided into three categories in which each category is associated with a determined range of credit limits, according to a client's income.

Variable 2: measures the use of revolving credit over twelve months.

Variable 3: measures the inactivity of the client over twelve months.

Variable 4: measures the intensity of delay problems over twelve months.

Variable 5: measures the intensity of product use over six months, assigning greater weights to more recent months.

Variable 6: measures the usage of the credit limit over six months.

Variable 7: measures the maximum client debt over six months.

Other variables needed to run the Multi-state Markov model are as follows:

Time: time since the beginning of the transition analyses (time=0 in January 2005 and time=12 in January 2006).

Status: state of the credit card at each time.

For the logistic regression models, another variable was needed:

Performance: binary variable that indicates, depending on the model, if the event *voluntary cancelling* or *default* happened during the six or twelve months after the observation month.

5. Results

For convenience, we used the Enterprise Miner module of the SAS statistical software to adjust the logistic regression model and the MSM package, implemented in R, to estimate the Multi-state Markov model. R is a free software for statistical computing that is built collaboratively by many developers. Details of the R software can be found at R Core Team (2014). The MSM package that was developed by Christopher Jackson (Jackson, 2007) allows the estimation of Multi-state Markov models. It is important to notice that both models could be estimated by R or SAS.

5.1. Logistic regression models

To compare some of the results of the Multi-state Markov models with those of the logistic regression models with binary response, we developed 4 different models:

Model L1: Logistic regression model to estimate the probability of client default over a six-month period.

Model L2: Logistic regression model to estimate the probability of client default over a twelve-month period.

Model L3: Logistic regression model to estimate the probability of client cancellation over a six-month period.

Model L4: Logistic regression model to estimate the probability of client cancellation over a twelve-month period.

The same variables used for the Multi-state Markov model were also considered. Non-significant variables were eliminated using the stepwise method, in each model. Stepwise is a method for selection of variables in a regression model. Its first step is to identify the independent variable that is the most important to explain the dependent variable; the second variable to be included is the one that brings more information to the model, considering that the first variable is already there. At this step, the importance of the first variable in a model with the two select variables is evaluated; if its presence is not significant, the variable is removed and the processes continues with the second variable only in the model, else, the process continues in order to verify if the inclusion of a third variable would bring significant improvement in the model and so on. This method was proposed by Efroymson (1960 apud Montgomery & Peck, 1992, p. 275).

The scheme that is shown in Figure 7 illustrates the database structure used for the logistic models with response variables observed after a twelve-month period.

It is important to note that the parameter estimates of these models were obtained with the goal of predicting good clients, i.e. the higher the score, the better the client should be to the company.

5.1.1. Probability models of default in six and twelve months

The default models were estimated to identify the characteristics of a good client based on the predictor variables. Therefore, the response variable
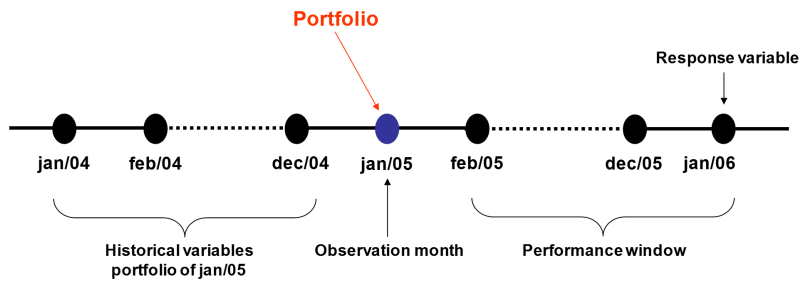
Régis, D. E. et al.
Using multi-state markov ... credit card risk. Production, 26(2), 330-344, abr./jun. 2016

PRODUCTION
PRODUÇÃO

337

Figure 7. Database structure for the logistic models. Source: the authors.

is the probability that a customer does not present default problems over a set period of observed time.

The selected variables by the stepwise method and their parameters, as well as the standard errors, are shown in Table 1 for model L1, and in Table 2 for model L2.

Overall, the results of the logistic regression models are consistent with the credit logic of the six (L1) and twelve (L2) month models.

5.1.2. Probability of cancellation models for six and twelve months

The cancellation models were estimated to identify the characteristics of a good client in relation to the probability of voluntary cancellation (attrition) according to the predictor variables. Therefore, the model response is the probability that a customer will not cancel his or her credit card over a set period of time.

The selected variables by the stepwise method and their parameters, as well as standard errors, are shown in Table 3 for model L3, and in Table 4 for model L4.

Overall, the results of the logistic regression models are consistent with the cancellation logic of the six (L3) and twelve (L4) month models.

5.2. Multi-state Markov model

The estimators for the models of each kind of transition, shown in Table 5, represent the relationship between the variables and the transition risk among the various states. The states *in compliance*, *revolving*, *in delay*, *voluntary cancellation* and *default* are represented, respectively, by the numbers 1, 2, 3, 4 and 5. To calculate each transition intensity between the various states, we can use (7). The table with the baselines, as well as the quantities of all the transitions observed in the development database, appears in the Appendix A.

Table 1. Default Estimates for 6 months (Model L1).

Variable	Estimator	Standard Error	P–Value
Intercept	5.4127	0.1934	0.0000
Var. 3	-0.0393	0.0159	0.0136
Var. 4	-0.2173	0.0129	0.0000
Var. 5	0.0909	0.0219	0.0000
Var. 6	-0.3191	0.0280	0.0000
Var. 7	-0.1386	0.0295	0.0000

Source: the authors.

Table 2. Default Estimates for 12 months (Model L2).

Variable	Estimator	Standard Error	P–Value
Intercept	4.6852	0.1537	0.0000
Var. 2	-0.0230	0.0100	0.0217
Var. 3	-0.0413	0.0130	0.0000
Var. 4	-0.1884	0.0117	0.0015
Var. 5	0.0849	0.0180	0.0000
Var. 6	-0.3034	0.0235	0.0000
Var. 7	-0.1432	0.0236	0.0000

Source: the authors.

Table 3. Cancellation Estimates for 6 months (Model L3).

Variable	Estimator	Standard Error	P–Value
Intercept	1.3625	0.1567	0.0000
Var. 1 / category 1	-0.5456	0.1452	0.0002
Var. 1 / category 2	0.0342	0.1254	0.7851
Var. 3	-0.0452	0.0128	0.0004
Var. 4	-0.2081	0.0137	0.0000
Var. 6	0.4766	0.0247	0.0000

Font: the authors.

Table 4. Cancellation Estimates for 12 months (Model L4).

Variable	Estimator	Standard Error	P–Value
Intercept	1.2198	0.1484	0.0000
Var. 1 / category 1	-0.5456	0.1452	0.0002
Var. 1 / category 2	0.0342	0.1254	0.7851
Var. 3	-0.0762	0.0112	0.0000
Var. 4	-0.1499	0.0122	0.0000
Var. 5	-0.0544	0.0215	0.0113
Var. 6	0.3659	0.0183	0.0000

Source: the authors.

Table 5. Multi-state Markov model estimates.

Transition	Variable	Estim.	Standard Error	P-Value	Transition	Variable	Estim.	Standard Error	P-Value
1-2	Var. 1	-0.0305	0.0306	0.1598	2-4	Var. 1	-0.0805	0.0898	0.1850
1-2	Var. 2	0.1462	0.0050	0.0000	2-4	Var. 2	0.0314	0.0143	0.0138
1-2	Var. 3	-0.0289	0.0072	0.0000	2-4	Var. 3	-0.0631	0.0218	0.0019
1-2	Var. 4	0.0447	0.0095	0.0000	2-4	Var. 4	-0.0652	0.0231	0.0023
1-2	Var. 5	0.0101	0.0090	0.1312	2-4	Var. 5	-0.0199	0.0267	0.2287
1-2	Var. 6	0.1193	0.0124	0.0000	2-4	Var. 6	-0.0230	0.0406	0.2855
1-2	Var. 7	0.1045	0.0109	0.0000	2-4	Var. 7	-0.0164	0.0394	0.3392
1-3	Var. 1	-0.0126	0.0517	0.4039	3-1	Var. 1	-0.2382	0.0606	0.0000
1-3	Var. 2	0.0638	0.0103	0.0000	3-1	Var. 2	-0.1383	0.0125	0.0000
1-3	Var. 3	0.0007	0.0105	0.4721	3-1	Var. 3	-0.0146	0.0128	0.1264
1-3	Var. 4	0.0991	0.0147	0.0000	3-1	Var. 4	-0.0328	0.0105	0.0009
1-3	Var. 5	-0.0045	0.0160	0.3890	3-1	Var. 5	0.0423	0.0221	0.0277
1-3	Var. 6	0.0490	0.0214	0.0110	3-1	Var. 6	-0.0024	0.0245	0.4602
1-3	Var. 7	0.0472	0.0179	0.0043	3-1	Var. 7	0.0218	0.0243	0.1850
1-4	Var. 1	-0.0017	0.0655	0.4896	3-2	Var. 1	-0.0113	0.0368	0.3793
1-4	Var. 2	0.0390	0.0145	0.0036	3-2	Var. 2	-0.0066	0.0069	0.1692
1-4	Var. 3	0.0362	0.0120	0.0013	3-2	Var. 3	-0.0354	0.0084	0.0000
1-4	Var. 4	-0.0606	0.0267	0.0116	3-2	Var. 4	-0.0451	0.0067	0.0000
1-4	Var. 5	-0.0892	0.0233	0.0001	3-2	Var. 5	0.0467	0.0132	0.0002
1-4	Var. 6	-0.0036	0.0275	0.4483	3-2	Var. 6	0.0223	0.0158	0.0786
1-4	Var. 7	0.0148	0.0225	0.2550	3-2	Var. 7	0.0213	0.0161	0.0937
2-1	Var. 1	-0.0041	0.0326	0.4502	3-4	Var. 1	-0.0493	0.0828	0.2761
2-1	Var. 2	-0.0883	0.0050	0.0000	3-4	Var. 2	-0.1030	0.0206	0.0000
2-1	Var. 3	-0.0223	0.0128	0.0011	3-4	Var. 3	0.0566	0.0179	0.0008
2-1	Var. 4	-0.0314	0.0075	0.0000	3-4	Var. 4	0.0459	0.0140	0.0005
2-1	Var. 5	0.0274	0.0091	0.0014	3-4	Var. 5	-0.0467	0.0501	0.1754
2-1	Var. 6	-0.0385	0.0148	0.0045	3-4	Var. 6	-0.1047	0.0360	0.0018
2-1	Var. 7	-0.0183	0.0138	0.0925	3-4	Var. 7	-0.0962	0.0343	0.0025
2-3	Var. 1	-0.2424	0.0305	0.0000	3-5	Var. 1	0.0302	0.0434	0.2434
2-3	Var. 2	-0.0254	0.0050	0.0000	3-5	Var. 2	-0.0126	0.0082	0.0636
2-3	Var. 3	0.0089	0.0066	0.0878	3-5	Var. 3	-0.0348	0.0100	0.0003
2-3	Var. 4	0.0479	0.0060	0.0000	3-5	Var. 4	0.0036	0.0079	0.3260
2-3	Var. 5	-0.0260	0.0096	0.0034	3-5	Var. 5	0.0494	0.0158	0.0009
2-3	Var. 6	0.0257	0.0140	0.0332	3-5	Var. 6	0.0415	0.0190	0.0144
2-3	Var. 7	0.0365	0.0179	0.0041	3-5	Var. 7	0.0377	0.0196	0.0271

Source: the authors.

In Table 6 we can see that all the variables are significant for at least two kinds of transitions. This fact indicates that we could have lost information if we had eliminated some of these variables.

5.3. Comparison between Multi-state Markov and Logistic Regression Models

The main purpose of this study is to test the application of Multi-state Markov models for credit card risk, a product that presents multi-state recurrent event characteristics.

An important characteristic of the Multi-state Markov model is that once the transition intensity matrix is estimated, we can easily generate several score models for several time periods, and it is possible to

order clients according to their risk profile for various purposes. In this study, in relation to credit card risk, we generated behavior score models, anti-attrition score models, delay score models and propensity score models for the use of revolving credit. To compare these with the logistic regressions, we generated default probability score models for six and twelve months which can be considered behavior score models, and voluntary cancellation models for six and twelve months, which can be considered anti-attrition score models.

To generate the default models from the intensity transition matrix we used the probabilities of transition from any non-absorbent state to the absorbent *default* state, during six or twelve months. In the case of the cancellation models, we used the probabilities

of transition from any non-absorbent state to the absorbent *voluntary cancellation* state during six or twelve months.

Besides the default and cancellation models, we analyzed delay models and propensity models in relation to the use of revolving credit for six or twelve-month periods. In the case of delay models we used the probability of transition in a six or twelve-month period from any non-absorbent state to the *in delay* state. The probability of transition in a six or twelve-month period from any non-absorbent state to the *revolving* state has been used for the propensity score models.

5.3.1. Performance indicators

Table 7 shows the Kolmogorov-Smirnov statistics and the Gini coefficients for each analyzed model. Figures 8 to 15 illustrate the sorting capability of clients through back test graphics, where the observations are divided into deciles. The first decile has the higher-risk clients and the last decile has the lower-risk clients. All indicators were obtained from the validation sample. The comparison was made descriptively only.

The default probability models using the transition matrix of the multi-state model for six and twelve

Table 6. Significant variables for each kind of transition.

Variable	1-2	1-3	1-4	2-1	2-3	2-4	3-1	3-2	3-4	3-5
Var. 1					x		x			
Var. 2	x	x	x	x	x	x	x		x	x
Var. 3	x		x	x	x	x		x	x	x
Var. 4	x	x	x	x	x	x	x	x	x	
Var. 5			x	x	x		x	x		x
Var. 6	x	x		x	x			x	x	x
Var. 7	x	x		x	x			x	x	x

Source: the authors.

Table 7. Performance indicators for the models.

Score	Kolmogorov-Smirnov Statistic (%)	Gini Coefficient (%)
MSM Default Score 6 months	55.5	70.8
LOG Default Score 6 months	51.5	66.2
MSM Default Score 12 months	49,2	64.4
LOG Default Score 12 months	46.6	61.4
MSM Cancellation Score 6 months	47.9	52.8
LOG Cancellation Score 6 months	52.5	59.0
MSM Cancellation Score 12 months	36.7	44.4
LOG Cancellation Score 12 months	40.7	48.2
MSM Delay Score 6 months	42.1	53.4
MSM Delay Score 12 months	38.3	47.4
MSM Revolving Credit Score 6 months	49.0	59.2
MSM Revolving Credit Score 12 months	44.8	54.8

Source: the authors.

Figure 8. Backtest graph for the default score models for 6 months. Source: the authors.

340 | PRODUCTION

Régis, D. E. et al.
Using multi-state markov ... credit card risk. Production, 26(2), 330-344, abr./jun. 2016
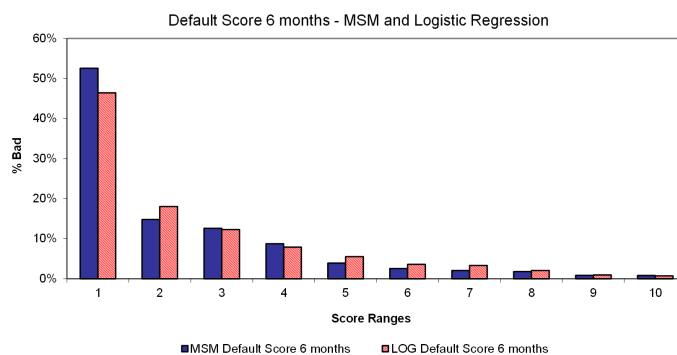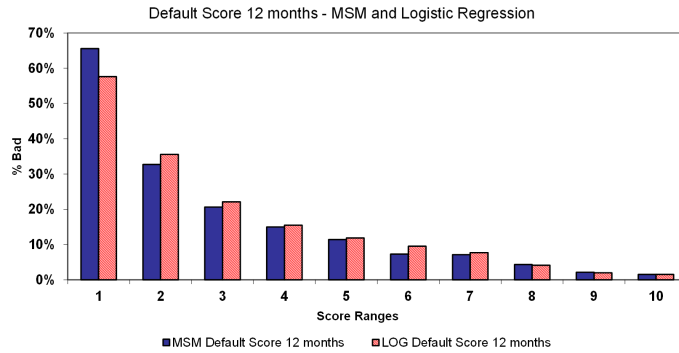
Figure 9. Backtest graph for the default score models for 12 months. Source: the authors.
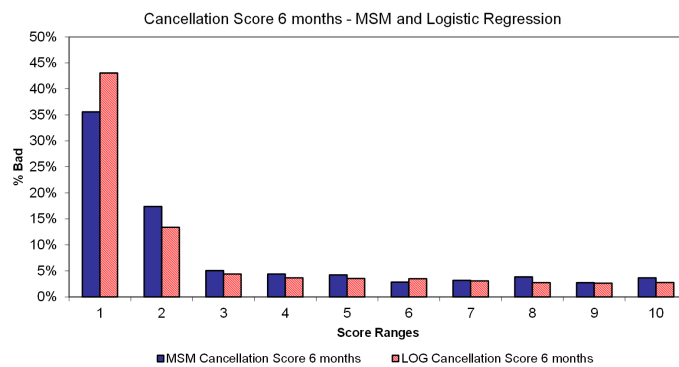

Figure 10. Backtest graph for the cancellation score models for 6 months. Source: the authors.
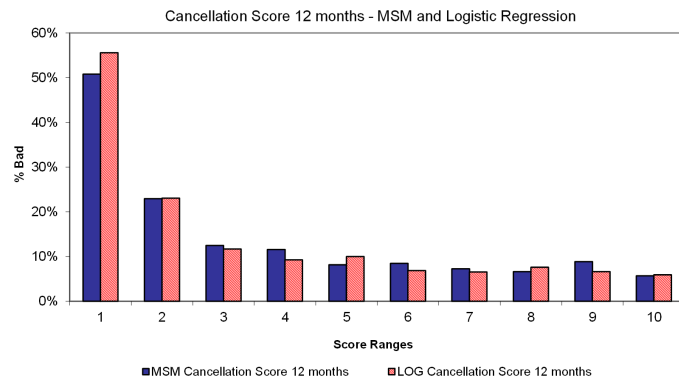

Figure 11. Backtest graph for the cancellation score models for 12 months. Source: the authors.

months showed better results when compared with the results obtained from the default probability models using logistic regression, as shown in Table 7. In the case of cancellation probability, the models obtained through logistic regression showed better results, as shown in Table 7.

In Figure 8, we can verify that the default models for six months showed a consistent order, with a small advantage for the model obtained through the transition matrix. We observe that there are greater percentages of bad clients in the lower scores and that these percentages consistently fall as we progress to the greater scores. The Kolmogorov Smirnov statistic and the Gini index also indicate that the Multi-state Markov model is better.

The default models for twelve months, despite displaying poorer performance when compared with the six-month model, also obtained good results,
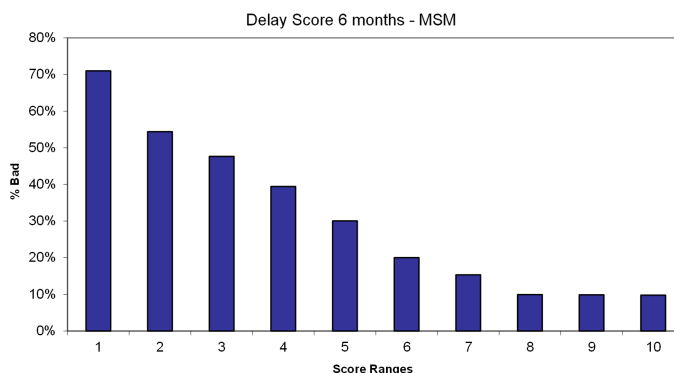
Figure 12. Backtest graph for the delay score model for 6 months. Source: the authors.
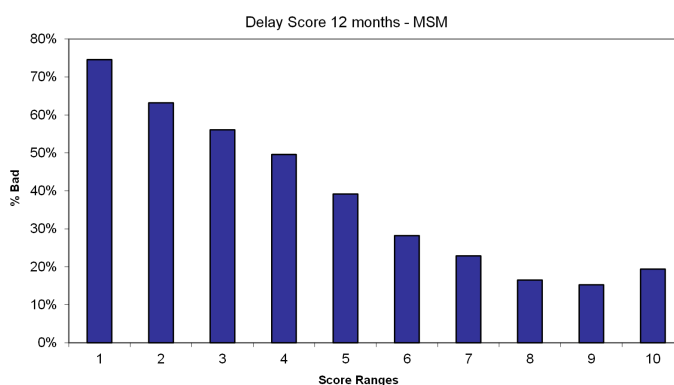

Figure 13. Backtest graph for the delay score model for 12 months. Source: the authors.
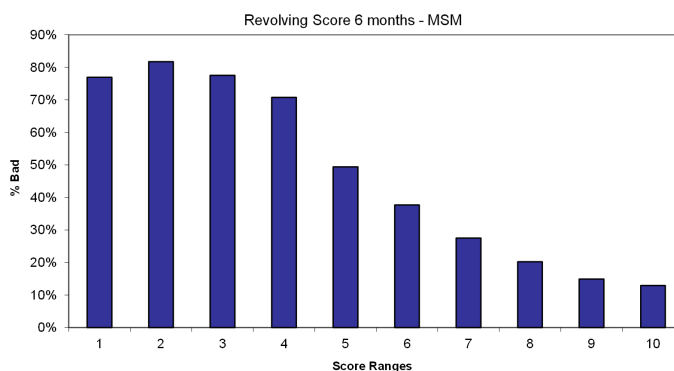

Figure 14. Backtest graph for the revolving credit score model for 6 months. Source: the authors.

again with the advantage going to the Multi-state Markov model, as shown in Table 6 and in the graphs of Figure 9.

For the six-month cancellation models, we observe satisfactory results for both kinds of models, with the advantage, in this case, going to the logistic regression models. Figure 10's backtest graph shows that the models display better ordering in the lower scores and that, in the greater scores, the cancellation

percentage does not fall in an accelerated fashion. As for the twelve-month models, we observe good form for the cancellation percentage declines, as we can see in Figure 11, although the Kolmogorov Smirnov and the Gini index display worse results when compared with the six-month models.

The delay model for six months obtained from the Multi-state Markov model transition intensity matrix showed a consistent ordering as we can see
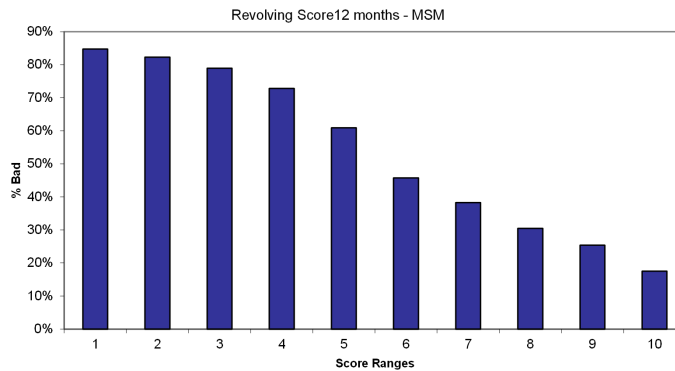
342 | **PRODUCTION**

Régis, D. E. et al.
Using multi-state markov ... credit card risk. Production, 26(2), 330-344, abr./jun. 2016

Figure 15. Backtest graph for the revolving credit score model for 12 months. Source: the authors.

from Figure 12, where clients with lower scores have a greater probability of delay.

In the case of the twelve-month cancellation model obtained from the Multi-state Markov model transition intensity matrix, we still have some degree of client ordering, in spite of the fact that the three largest deciles do not show consistent ordering, as shown in Figure 13.

The six-month revolving credit model obtained from the Multi-state Markov model transition intensity matrix still presents some degree of ordering, as we can see in Figure 14, despite the observed inversion at the lowest decile in relation to the penultimate decile. Strategies to encourage product use or client retention can be derived from this model, where the higher the score the lower the propensity to use revolving credit.

In the case of the revolving credit model for 12 months, the backtest graph shown in Figure 15 indicates a model with some discriminatory power.

Except for cancellation, the results observed for MSM models were as good or better than those obtained for logistic regression models. Other applied studies may be conducted in order to verify the stability of this conclusion. The main advantage of MSM models is the possibility to estimate the probability of transitions, for any time horizon, from the same multivariate model.

6. Conclusion

This work presents the application of a Multi-state Markov model in continuous time to credit card risk, taking advantage of the characteristics of multi-state events and recurrent events that can be observed in the relationship between customers and credit card companies. We tested the performance of default, cancellation, delay, and revolving credit score models obtained by using the transition intensity matrix.

We compared these default and cancellation score models with estimated logistic regression models with binary response, using the same database and the same variables, and verified that the models obtained from the Multi-state Markov model showed better results than the logistic models in the case of the default score models, and worse results in the case of the cancellation score models.

An interesting characteristic of the multi-state Markov model is the fact that, once the intensity matrix has been estimated, we can easily obtain several models for various times and purposes. This kind of model can be tested for any product with multi-state recurrent event characteristics. This versatility is, in fact, the main advantage of this family of models over the logistic regression. To obtain the same result from logistic models it would be necessary the estimation of several and independent regression models. The Markov model presented in this paper uses all multivariate structure of the data to obtain the parameter estimates and the results may be used to make simulations of the clients conditions in different time horizon.

In this paper, due limitations on the data base, only one year to follow the changes on the states of the clients, we only compared the models performance for 6 and 12 months ahead. As expected, the larger the time horizon, the worst is the performance of the models (Table 7).

Better behavior models and anti-attrition models, using logistic regressions and Multi-state Markov models could be developed to take greater advantage of client behavioral characteristics and analyze several other variables such as market behavioral information which can be easily obtained from credit bureau firms such as ACSP and Experian-Serasa. The application of more appropriate selection techniques, in both cases, would also provide better performing models.

A study of the effect of several variables in transitions between company ratings, as the ones

Régis, D. E. et al.
Using multi-state markov ... credit card risk. Production, 26(2), 330-344, abr./jun. 2016

PRODUCTION
PRODUÇÃO

343

provided by rating agencies, or in transitions between internal company ratings, would be a good area for future research.

References

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: a review of literature. *Intelligent Systems in Accounting, Finance & Management*, *18*(2-3), 59-88. http://dx.doi.org/10.1002/isaf.325.

Andreeva, G. (2006). European generic scoring models using survival analysis. *The Journal of the Operational Research Society*, *57*(10), 1180-1187. http://dx.doi.org/10.1057/palgrave.jors.2602091.

Araujo, E. A., & Carmona, C. U. M. (2007). Desenvolvimento de Modelos Credit Scoring com Abordagem de Regressão de Logística para a Gestão da Inadimplência de uma Instituição de Microcrédito. *Contabilidade Vista & Revista*, *3*, 107-131.

Blackwell, M., & Sykes, C. (1992). The assignment of credit limits with a behavior-scoring system. *IMA Journal of Management Mathematics*, *4*(1), 73-80. http://dx.doi.org/10.1093/imaman/4.1.73.

Chi, B.-W., & Hsu, C.-C. (2012). A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model. *Expert Systems with Applications*, *39*(3), 2650-2661. http://dx.doi.org/10.1016/j.eswa.2011.08.120.

Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. London: Chapman and Hall.

David, W. H., Stanley, L., & Susanne, M. (2008). *Applied survival analysis: regression modeling of time to event data* (2nd ed.). New York: John Wiley & Sons.

Finlay, S. (2010). *Credit scoring, response, modelling and insurance rating: a practical guide to forecasting consumer behavior*. London: Palgrave Macmillan.

Gupta, J., Wilson, N., Gregoriou, A., & Healy, J. (2014). The effect of internalization on modelling credit risk for SMEs: evidence from UK market. *Journal of International Financial Markets, Institutions and Money*, *31*, 397-413. http://dx.doi.org/10.1016/j.intfin.2014.05.001.

Hoper, M. A., & Lewis, E. M. (1992). Behavior scoring and adaptive control systems. In L. C. Thomas, J. N. Crook & D. B. Edelman (Orgs.), *Credit scoring and credit control* (pp. 257-276). Oxford: Claredon Press.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley & Sons.

Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: regression modeling of time to event data* (2nd ed.). Hoboken: John Wiley & Sons.

Hougaard, P. (1999). Multi-state models: a review. *Lifetime Data Analysis*, *5*(3), 239-264. http://dx.doi.org/10.1023/A:1009672031531. PMid:10518372.

Jackson, C. H. (2007). *Multi-state modeling with R: the MSM package version 0.6*. London: Imperial College. Retrieved in 16 July 2013, from http://rss.acs.unt.edu/Rdoc/library/msm/doc/msm-manual.pdf

Jackson, C. H., Sharpless, L. D., Thompson, S. G., Duffy, S. W., & Couto, E. (2003). Multistate markov models for disease progression with classification error. *The Statistician*, *52*(2), 193-209. http://dx.doi.org/10.1111/1467-9884.00351.

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. http://dx.doi.org/10.1080/01621459.1985.10478195.

Kay, R. (1986). A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics*, *42*(4), 855-865. http://dx.doi.org/10.2307/2530699. PMid:2434150.

King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, *9*(2), 137-163. http://dx.doi.org/10.1093/oxfordjournals.pan.a004868.

Kuhn, I. N. (2009). *Administração financeira de negócio*. Ijuí: Editora Ijuí.

Lewis, E. M. (1992). *An introduction to credit scoring* (2nd ed.). San Rafael: Athena Press.

Marshall, G., & Jones, R. H. (1995). Multi-state models and diabetic retinopathy. *Statistics in Medicine*, *14*(18), 1975-1983. http://dx.doi.org/10.1002/sim.4780141804. PMid:8677398.

Menard, S. (2010). *Logistic regression: from introductory to advanced concepts and applications*. Los Angeles: Sage Publications Inc.

Montgomery, D. C., & Peck, E. A. (1992). *Introduction to linear regression analysis* (2nd ed.). New York: John Wiley & Sons.

Paes, A. T., & Lima, A. C. (2004). A SAS macro for estimating transition probabilities in semiparametric models for recurrent events. *Computer Methods and Programs in Biomedicine*, *75*(1), 59-65. http://dx.doi.org/10.1016/j.cmpb.2003.08.007. PMid:15158048.

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved in 4 November 2014, from http://www.R-project.org/

So, M. M. C., & Thomas, L. C. (2010). Modeling and model validation of the impact of the economy on the credit risk of credit card portfolios. *The Journal of Risk Model Vaidation*, *4*, 93-126.

Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, *16*(2), 149-172. http://dx.doi.org/10.1016/S0169-2070(00)00034-0.

Thomas, L. C. (2009). *Consumer credit models: pricing, profit and portfolio*. Oxford: Oxford Press.

Thomas, L. C., Edelman, D. B., & Crook, J. N. (2002). *Credit scoring and its applications*. Philadelphia: Siam.

Thomas, L. C., Ho, J., & Scherer, W. T. (2001). Time will tell: behavior scoring and the dynamics of consumer credit assessment. *IMA Journal of Management Mathematics*, *12*(1), 89-103. http://dx.doi.org/10.1093/imaman/12.1.89.

Thomas, L. C., Oliver, R. W., & Hand, D. J. (2005). A survey of the issues in consumer credit modelling research. *The Journal of the Operational Research Society*, *56*(9), 1006-1015. http://dx.doi.org/10.1057/palgrave.jors.2602018.

Trench, M. S., Pederson, S. P., Lau, E. T., Ma, L., Wang, H., & Nair, S. K. (2003). Managing credit lines and prices for Bank One credit cards. *Interfaces*, *33*(5), 4-21. http://dx.doi.org/10.1287/inte.33.5.4.19245.

Van den Poel, D., & Larivière, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, *157*(1), 196-217. http://dx.doi.org/10.1016/S0377-2217(03)00069-9.

Wang, H. Y. F., & Ikeda, A. A. (2004). *Análise do mercado de cartão de crédito brasileiro*. São Paulo: FEAUSP. Retrieved in 30 January 2015, from http://www.ead.fea.usp.br/Semead/7semead/paginas/artigos%20recebidos/marketing/mkt09-_analise_de_mercado_do_cartao_de_cr%E9dito.pdf

344 **PRODUCTION**

Régis, D. E. et al.
Using multi–state markov ... credit card risk. Production, 26(2), 330-344, abr./jun. 2016

Appendix A. Baselines and Transitions.

Baselines

Table 1A shows the baselines for each kind of transition.

Table 1A. Baselines for transitions.

Transition	Estimator	Standard Error	P-Value
1–2	0.0145	0.0015	0.0000
1–3	0.0111	0.0006	0.0000
1–4	0.0160	0.0036	0.0000
2–1	0.3542	0.0395	0.0000
2–3	0.1630	0.0181	0.0000
2–4	0.0313	0.0107	0.0018
3–1	0.2645	0.0555	0.0000
3–2	0.2648	0.0358	0.0000
3–4	0.1494	0.0509	0.0017
3–5	0.1103	0.0181	0.0000

Source: the authors.

Transitions

Table 2A shows the observed quantity of each kind of transition between the possible states in the development base of the Multi-state Markov model (1 – In compliance, 2 - Revolving, 3 – In delay, 4 - Voluntary cancellation and 5 – Default). The quantities of Table 2A do not reflect the real quantities of transitions of the company from which the data was obtained.

Table 2A. Observed quantity of each kind of transition between the possible states in the development base of the Multi-state Markov model.

			To			
		1	2	3	4	5
From	1	57.662	4.617	1.760	912	0
	2	4.580	18.616	3.675	157	0
	3	1.100	2.808	3.111	340	1.680

Source: the authors.