Paper 9600 - 2016

Evaluation of a Customer's Life Cycle Time-to-Offer Cross-sales in a Bank, Based on the Behavior Score Using Logit Models & DTMC in SAS®

Cortés Oscar, Díaz Miguel, Multibanca Colpatria – ScotiaBank, Colombia

ABSTRACT

Scotiabank Colombian division – Colpatria, is the national leader in terms of providing credit cards, with more than 1,600,000 active cards—the equivalent to a portfolio of 700 million dollars approximately. The behavior score is used to offer credit cards through a cross-sell process, which happens only if customers have completed six months on books after using their first product with the bank. This is the minimum period of time requested by the behavior model. The six months on books internal policy suggests that the maturation of the client in this period is adequate, but this has never been proven. The following research aims to evaluate this hypothesis and calculate the appropriate time to offer cross-sales to new customers using Logistic Regression (Logit), while also segmenting these sales targets by their level of seniority using Discrete-Time Markov Chains (DTMC).

INTRODUCTION

One of the main drivers of credit origination in all banks in Colombia are cross-selling. Through the offering of new products to old customers of the organization, more money is placed on the market with a considerably less risk than traditional channels (branches, website); the weighting of these originations in the case of Multibanca Colpatria can be up to 65% of all the credits issued in a single month for personal loans.

The minimum time to start making offers of new products to a customer is 6 months. After this time it begins to run the logit model of behavior which seeks to find the best customers and focus on offering new products to these; however this period of 6 months is a business rule that has not been tested yet with analytical tools. It is possible that maturity of customers is not complete at 6 months, incurring in a greater risk than cross-selling or, on the contrary, the customers's maturation time may be previously achieved and thereby this leaving money on the table, affecting profitability and customer loyalty.

The aim of this paper is to analyze the efficiency of this period of 6 months with multiple tools. Three risk and two financial variables will be analyzed to assess whether both types of variables evolve similarly; the target population of the study focuses on new customers who acquired credit cards in the window April 2014 to June 2015.

GENERAL CONCEPTS

LOGISTIC REGRESSION

The logistic regression or Logit Model is a regression method by which the objective variable cannot be modeled as a linear function, so its is a categorical response variable (In most cases a dichotomous variable); to achieve the adjustment of the variables is used the Logit transformation, dividing the variable in the right categories and calculating the probability p or event rate for each one (Kristin L. Sainani, 2014).

$$Logit = ln\left(\frac{p}{1-p}\right)$$

Then is calculated an intercept and for each one of the categories of the n variables is estimated an associated beta *(β)*. With these values the model is formulated as it follows.

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1, i + \cdots + \beta_k x_k, i)}}$$

KS

The Kolmogorov-Smirnov statistic represents the power of the model to discriminate between good and bad customers (zeros and ones in the objective variable logistic regression), for it calculates the maximum difference between the distribution of good customers regard to distribution of bad customers; the higher the KS value, the greater adjustment has the model (Bradley, 2013).

PSI

The Population Stability Index represents the variation level of the composition of a population over time; the total percentages of participation of each category in the population at time t_i are compared with their new weighings in then population at time t_{i+k}; a higher PSI implies a larger variation in the composition of the population.

MARKOV CHAIN

A Markov chain discrete time (DTMC) is a stochastic process where the probability of moving from state i to state j in a given unit of time t is modeled. These probabilities are represented through a transition matrix that meets with Markovian property (memoryless). Processes in which the probabilities of the matrix change over time are called non-stationary Markov chains.

VINTAGE

Vintages in a credit risk context shows the behavior of a portfolio. In scoring models specifically, a vintage indicates the behavior of several indicators in different periods of time after the credit is originated.

BACKGROUND

In order to achieve a global segmentation of each client according to his portfolio's composition (he mix of all his products), the costumer behavior score was built. With this model, strategies of credit line increase, overlimits, and cross-selling bids can be applied in a more effective way; optimizing the usage of the latter ones is the goal of the following sections.

The score given to each client is the result of mixing their historical internal behavioral variables such as:

- Number of payments in recent months,
- Ratio payment / balance
- Debt trend in time
- Behavior of delinquency in recent months among others.

With the resulting score model, the population is grouped into construction segments as follows. (Only for customers with 6 or more months in books):

Score_Rank	N	Bad rate	Risk
1	318351	31.1%	Higher risk
2	299947	7.0%	High risk
3	140102	3.0%	Medium risk
4	109612	1.6%	Low risk
5	107324	0.8%	Lower risk
Total	**975336**	**13.0%**	

Exhibit 1: Distribution of the score in current population.

Of which are candidates for peronal loans cross-selling the two ranges with lower risk. Currently, the portfolio of personal loans (installment and revolving) comprises 65% of cross-sells with credit card origin. Monthly, for this segment it originated 12.9 million dollars on average.

Customers who have less than 6 months books are not qualified by the model and are not part of the objective base of cross-selling. The following section will assess whether the 6 months are enough for qualification, or if it is recommendable to offer cross-selling before, or if the 6 months in books are a short time to assume that the customer is already mature.

METHODOLOGY

DATABASE

14 independent databases were built (from June of 2014 to July of 2015), each one of them containing the vintage of its respective month; each database includes only the new customers to whom a credit card was delivered (first product), and for each following month after the origination, the customer is rated with the behavior model, in which the respective delinquency and balance are included

ANALYSIS TECHNIQUES

Five different methodologies will be used to find the best point in time to assume customer maturing:

1. Vintage analysis: the score, the balance, the utilization and the bad rate will be analyzed to find the equilibrium or maturity point of each variable
2. Trend analysis: The trend score's behavior will be observed (whether it should increase, decrease or stay the same) over time, and with this, the point in which the probability of each scenario will most likely remain constant will be found.
3. Markov chains: From the origination moment of the vintage, the respective transition matrices will be built in a monthly basis, over time, the steady-state matrix or the point of minimal variation of the matrix will be found
4. Optimization by regression: from the vintages each variable will be modeled using a linear regression, later, an optimization process in which the optimum point of maturity can be found will be built
5. Population stability: the distribution of the scorecard for each vintages will be evaluated on a monthly basis, to find the point in which the model's PSI reaches its minimum value or remains similar to the PSI for the overall population

RESULTS

The results for each of the selected methodologies will be displayed:

1. Vintage analysis:
    - Score: Average score for each subsequent month to origination, segmented by vintage. Figure 1: Evolution of average score by vintage over time
    - Balance: Total Balance for each month after the origination, segmented by vintage. Figure 2: Evolution of balance by vintage over time
    - Average utilization (balance / credit line) for each month after the origination, segmented by vintage. Figure 3: Evolution of average utilization by vintage over time
    - 30+ DPD : Proportion of customers with delinquency major or equal than 30 days for each month after the origination, segmented by vintage. Figure 4: Evolution of 30+ delinquency by vintage over time
    - 60+ DPD : Proportion of customers with delinquency major or equal than 60 days for each month after the origination, segmented by vintage. Figure 5: Evolution of 60+ delinquency by vintage over time

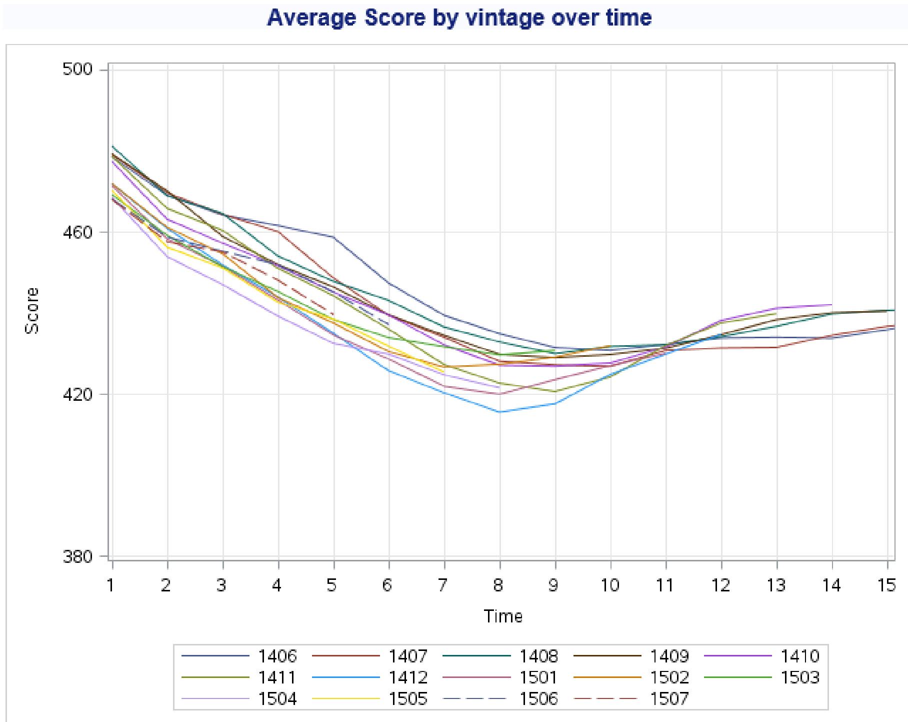

Figure 1: Evolution of average score by vintage over time

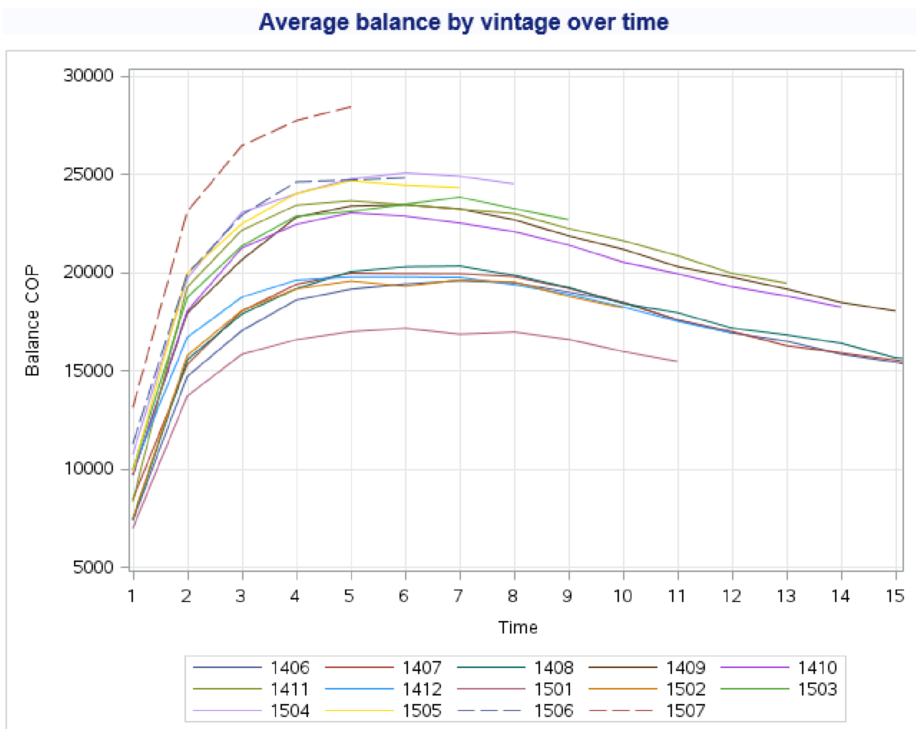

Figure 2: Evolution of balance by vintage over time

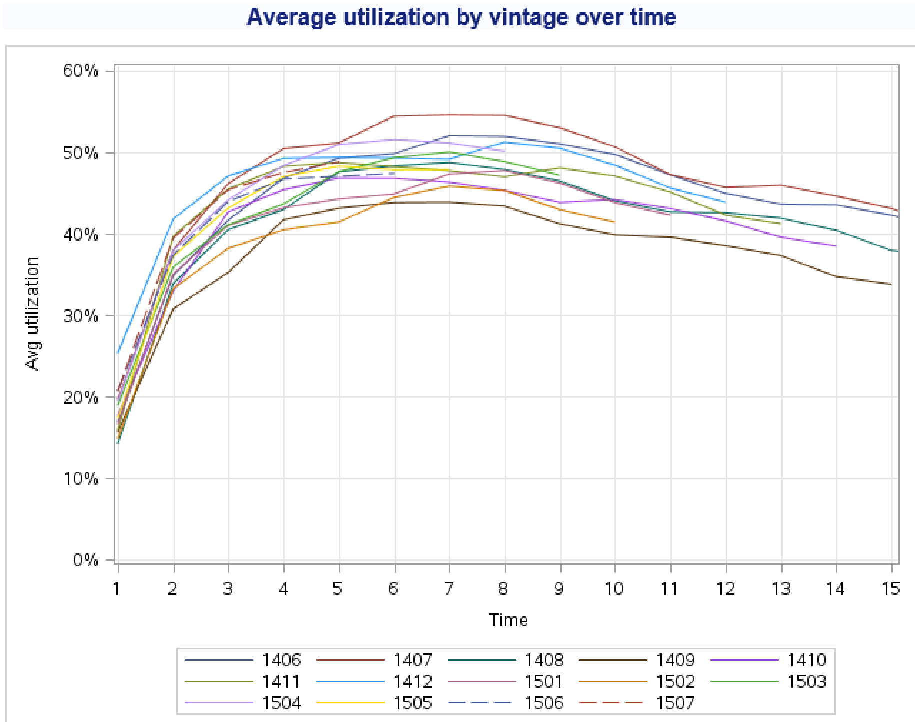

Figure 3: Evolution of average utilization by vintage over time

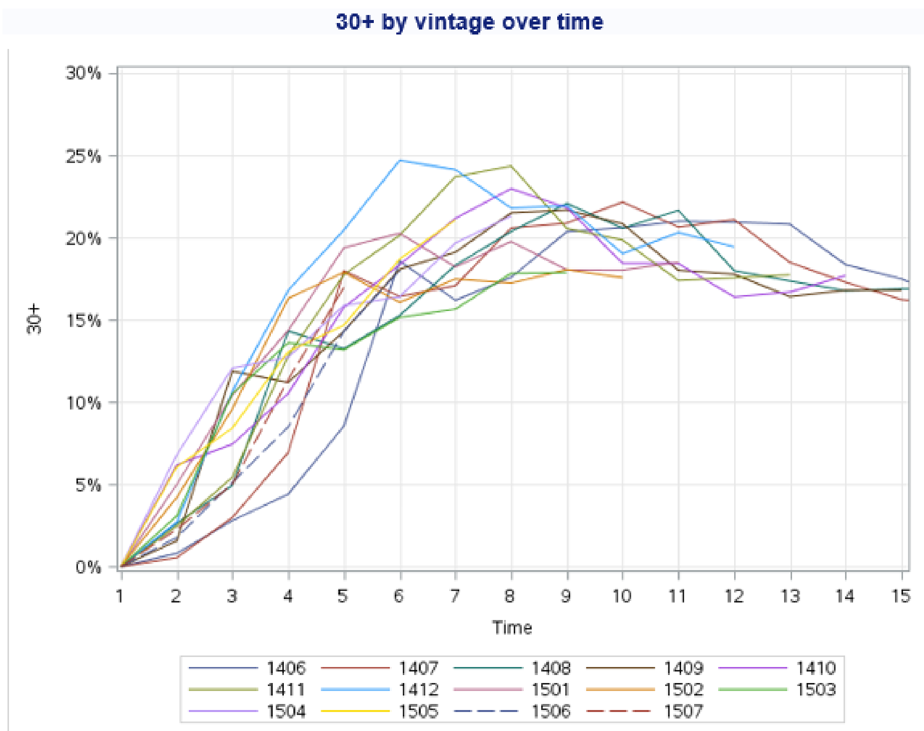

Figure 4: Evolution of 30+ delinquency by vintage over time

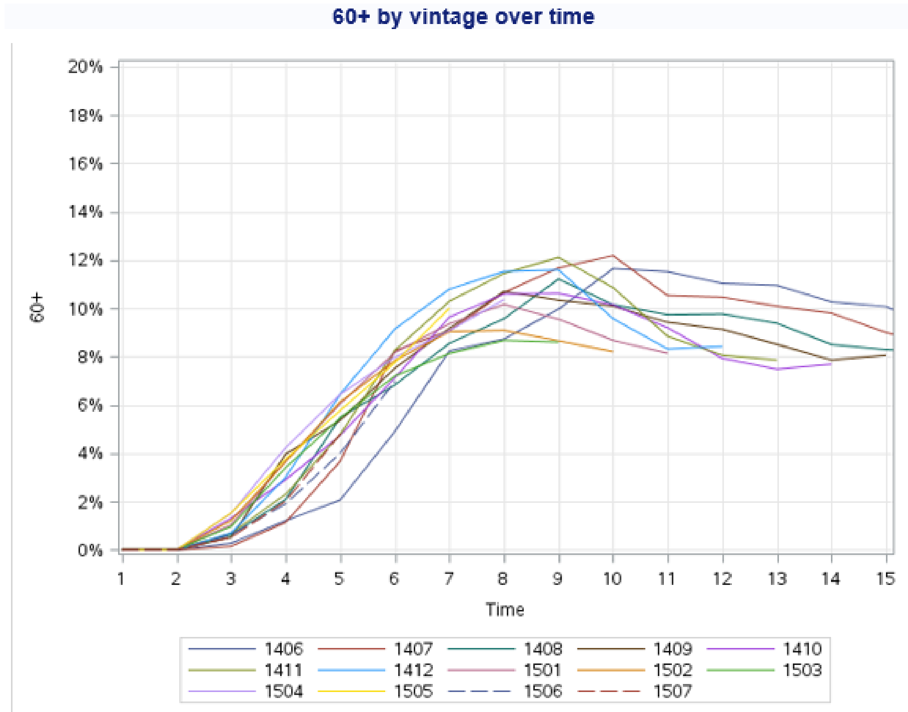

Figure 5: Evolution of 60+ delinquency by vintage over time

According to evidence, the average maturity of the variables is reached within the sixth and the ninth month. Considering the current policy of 6 months in books, it is worth pointing out that there is a great amount of customers that obtained a good score during the sixth month, but that said score could diminish on the following months (seventh to ninth), also, that the bad rate of the model (60+) still hasn't reached its maximum point, which it usually does between the eighth and the ninth month, and therefore, during the sixth month the risk can be undervalued

2. Trend analysis:

   There is a proportion of customers whose risk category increases on a monthly basis, another proportion of customers whose risk category decreases on the same period, and the remaining customers keep the same risk level; if the average weights of these movements are analyzed on a long-term basis, said movements will remain constant as the model stabilizes.

   It is observed that the distribution begins to stabilize in the eighth month, and 3-month standard deviation remains below 2% from the tenth month. Figure 6: Distribution of changes in risk over time - Figure 7: Trend of 3-month standard deviation over time.

3. Markov chains: the transition matrix for each time interval must be calculated, with each one of these the steady-state matrix should be found; due to the complex dynamics of the system, and the fact that this is a non-stationary Markov process, it is impossible to find said matrix, however, it is possible to find a close approximation to these one using the PSI, taking the transition matrix of the first period and multiplying it for the one of the second period, the resulting matrix will be multiplied for the transition matrix of the third period, and the PSI stability indicator will be calculated with the immediately preceding resulting matrix, the point in which the PSI finds its minimum point will be understood as the one belonging to the month in which the system remains stable. Exhibit 2: Transition and resulting matrix between 1 and 5 month.

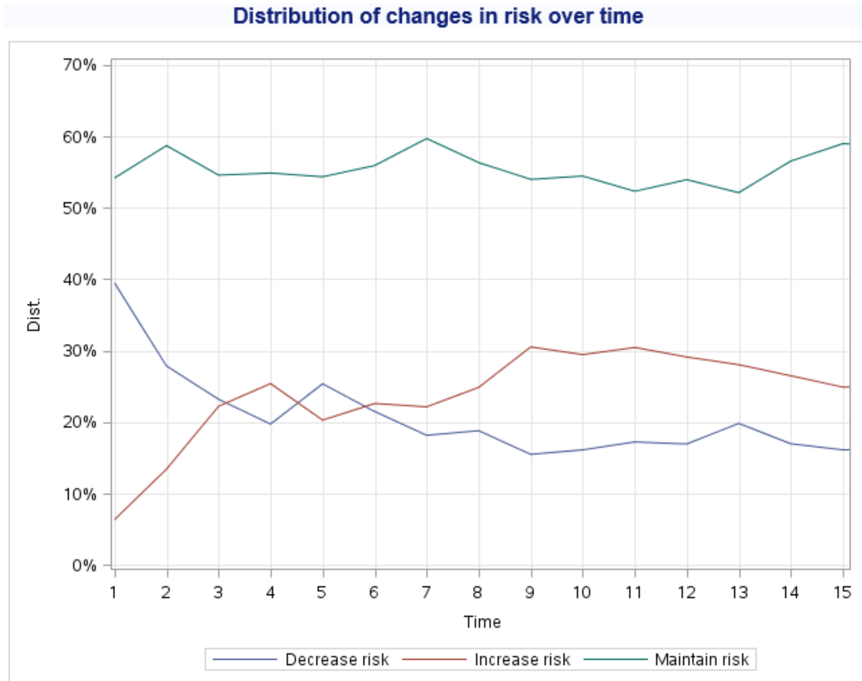

Figure 6: Distribution of changes in risk over time

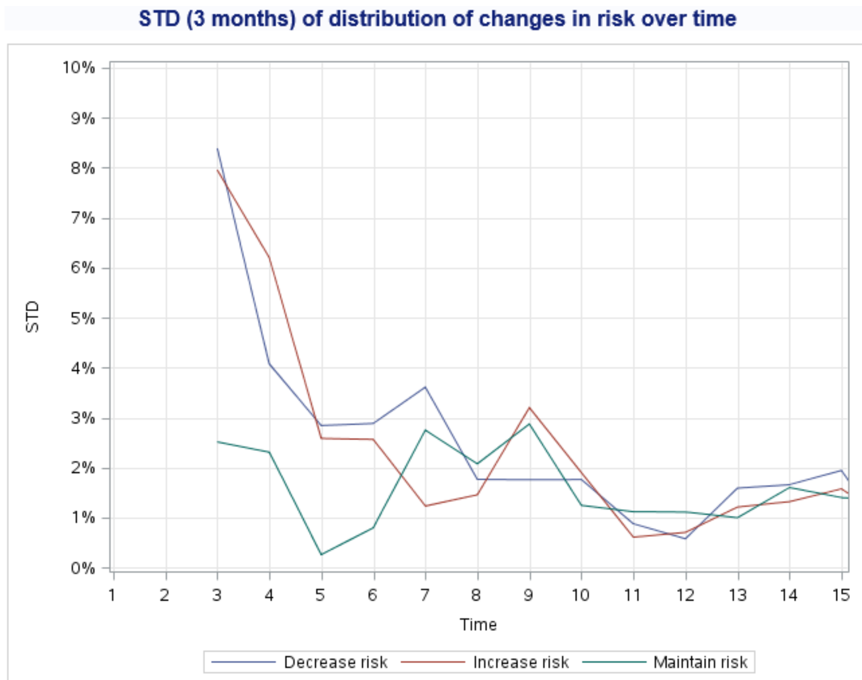

Figure 7: Trend of 3-month standard deviation over time.

Month	Transition Matrix						Resulting Matrix						PSI
	Score Rank	1	2	3	4	5		1	2	3	4	5	
1-2	1	84.8%	15.2%	0.0%	0.0%	0.0%	1	84.8%	15.2%	0.0%	0.0%	0.0%	NA
	2	25.8%	69.4%	4.6%	0.1%	0.0%	2	25.8%	69.4%	4.6%	0.1%	0.0%	
	3	0.1%	29.3%	70.6%	0.1%	0.0%	3	0.1%	29.3%	70.6%	0.1%	0.0%	
	4	0.0%	0.0%	0.0%	100.0%	0.0%	4	0.0%	0.0%	0.0%	100.0%	0.0%	
	5	0.0%	0.0%	0.0%	0.0%	100.0%	5	0.0%	0.0%	0.0%	0.0%	100.0%	

Month	Transition Matrix						Resulting Matrix						PSI
		1	2	3	4	5		1	2	3	4	5	
2-3	1	88.9%	11.1%	0.0%	0.0%	0.0%	1	77.1%	21.9%	1.0%	0.1%	0.0%	45%
	2	11.2%	82.1%	6.3%	0.5%	0.0%	2	30.7%	60.2%	8.5%	0.5%	0.0%	
	3	0.3%	8.0%	90.3%	1.5%	0.0%	3	3.6%	29.6%	65.5%	1.2%	0.0%	
	4	0.0%	0.0%	33.3%	60.0%	6.7%	4	0.0%	0.0%	33.3%	60.0%	6.7%	
	5	0.0%	0.0%	0.0%	0.0%	100.0%	5	0.0%	0.0%	0.0%	0.0%	100.0%	

Month	Transition Matrix						Resulting Matrix						PSI
		1	2	3	4	5		1	2	3	4	5	
3-4	1	62.9%	36.6%	0.5%	0.0%	0.0%	1	51.0%	45.7%	3.1%	0.2%	0.0%	80%
	2	11.6%	79.3%	8.3%	0.8%	0.0%	2	26.3%	59.7%	13.0%	1.0%	0.0%	
	3	0.4%	7.9%	88.4%	3.3%	0.0%	3	6.0%	30.0%	61.1%	2.9%	0.0%	
	4	1.5%	5.4%	49.3%	43.9%	0.0%	4	1.0%	5.8%	59.0%	34.1%	0.0%	
	5	0.0%	0.0%	0.0%	100.0%	0.0%	5	0.0%	0.0%	0.0%	100.0%	0.0%	

Month	Transition Matrix						Resulting Matrix						PSI
		1	2	3	4	5		1	2	3	4	5	
4-5	1	67.4%	32.5%	0.0%	0.0%	0.0%	1	38.7%	54.8%	6.0%	0.5%	0.0%	38%
	2	9.4%	83.1%	7.0%	0.5%	0.0%	2	23.5%	59.3%	15.9%	1.3%	0.0%	
	3	0.7%	8.3%	87.1%	3.9%	0.0%	3	7.3%	32.1%	56.4%	4.2%	0.0%	
	4	0.2%	4.6%	37.5%	56.3%	1.4%	4	1.7%	11.6%	64.6%	21.5%	0.5%	
	5	0.0%	0.0%	0.0%	0.0%	100.0%	5	0.2%	4.6%	37.5%	56.3%	1.4%	

Month	Transition Matrix	Resulting Matrix	PSI
i-j	⋮	⋮	⋮

Exhibit 2: Transition and resulting matrix between 1 and 5 month.

The PSI historic looks as follows:

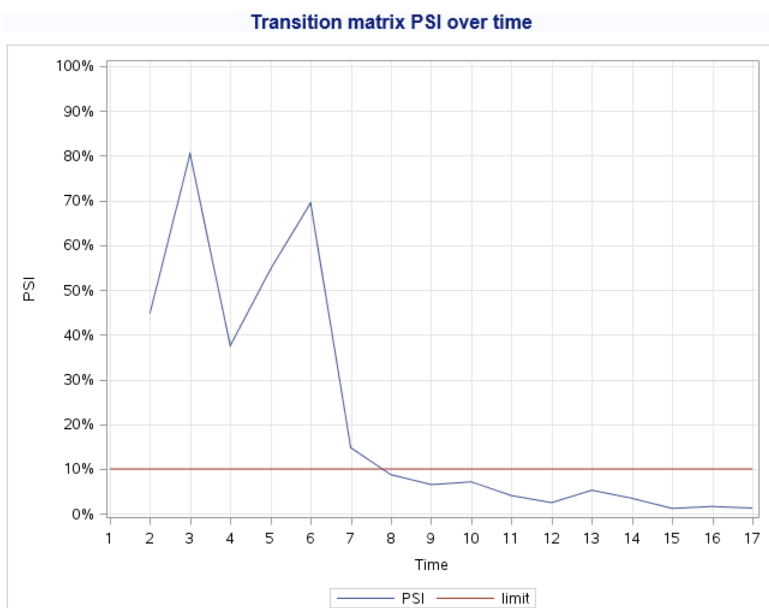

Figure 8: Historic transition matrix PSI

Taking the most common PSI limit (10%), one can say that the system finds its stable state in the eigth period.

4. Optimization by regression:

   For each of the variables (score, balance, utilization, 30+ and 60+) a fourth grade regression model was constructed using the PROC REG procedure of SAS®. With this the maximum or minimum each variable can be found and the optimum maturity point calculated.

   - Score (Figure 9: real vs predicted score)

   $$Score_{est} = 483.6 - 6.5x - 0.8x^2 + 0.1x^3 + 0.001x^4$$

   $$R^2 = 0.91$$

   - Balance (Figure 10: real vs predicted Balance)

   $$Balance_{est} = -2095 + 13641x - 2843x^2 + 247x^3 - 8x^4$$

   $$R^2 = 0.66$$

   - Utilization (Figure 11: real vs predicted Utilization)

   $$Util_{est} = -6.7 + 31.1\%x - 6.3x^2 + 0.5x^3 + 0.01x^4$$

   $$R^2 = 0.77$$

   - 30+ (Figure 12: real vs predicted 30+ delinquency)

   $$30 +_{est} = -0.3 - 1.5 + 1.9x^2 - 0.3x^3 + 0.09x^4$$

   $$R^2 = 0.91$$

   - 60+ (Figure 13: real vs predicted 60+)

   $$60 +_{est} = 3.9 - 5.5 + 2x^2 - 0.2x^3 + 0.06x^4$$

   $$R^2 = 0.94$$

   Then, the optimum value of maturity by variable is found.

	Score	Utilization	Balance	30+	60+
Optimum month	9.4	5.1	4.7	8.4	9.0
Optimum value	426.0	48%	21040	22%	11%
Adjusted optimum month	9.0	5.0	5.0	8.0	9.0
Adjusted optimum value	426.1	48%	21018	22%	11%
Observation	Min	Max	Max	Max	Max

Exhibit 3: optimum times of regression models

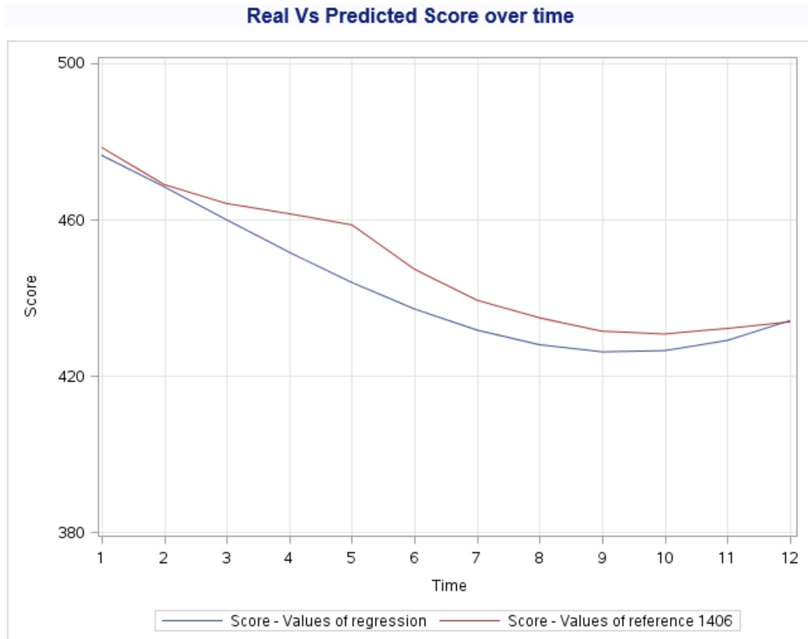

Figure 9: real vs predicted score

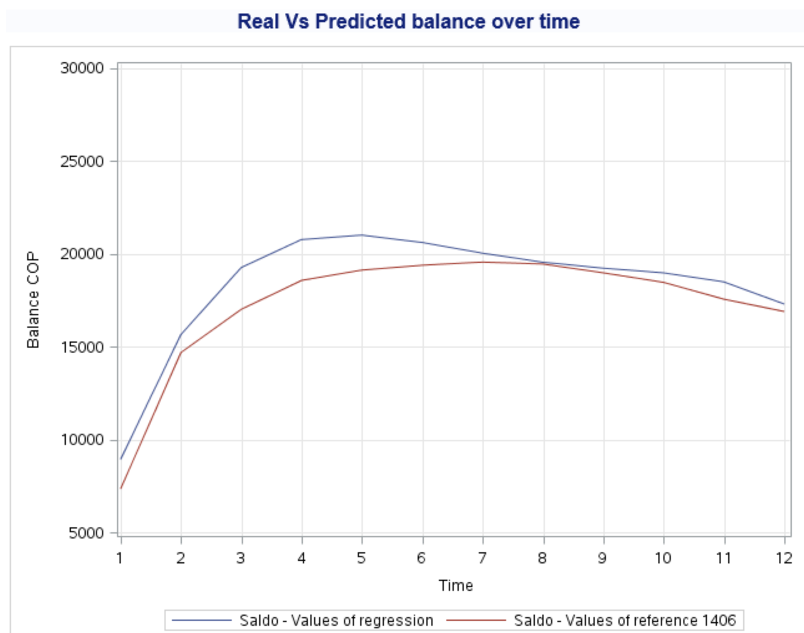

Figure 10: real vs predicted Balance

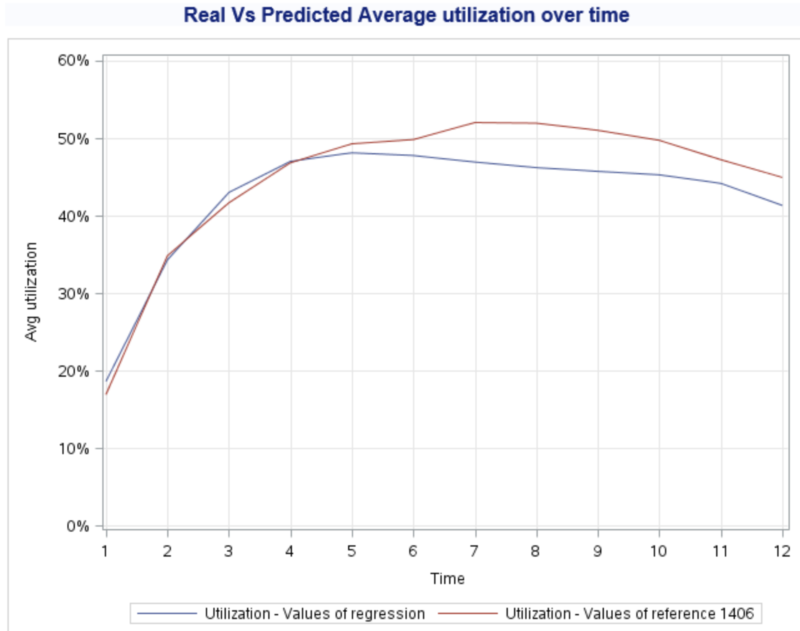

Figure 11: real vs predicted Utilization

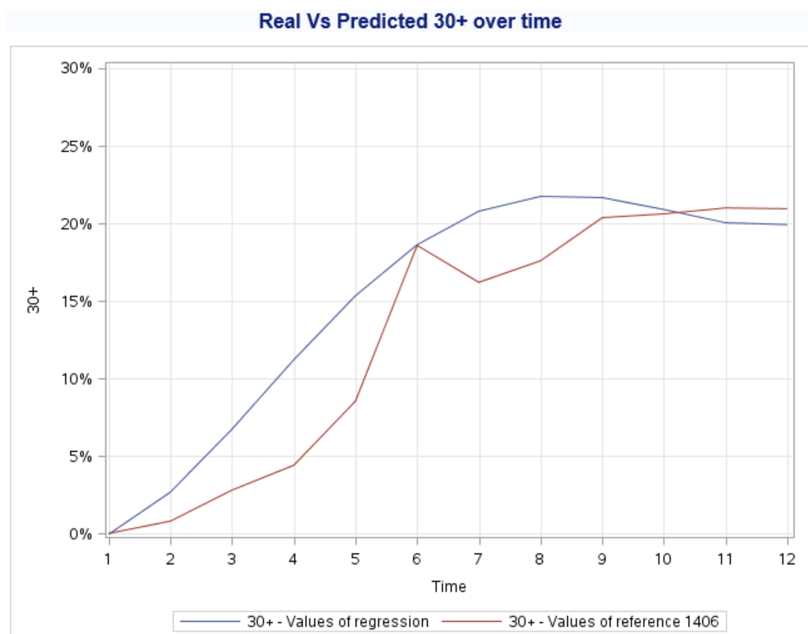

Figure 12: real vs predicted 30+ delinquency

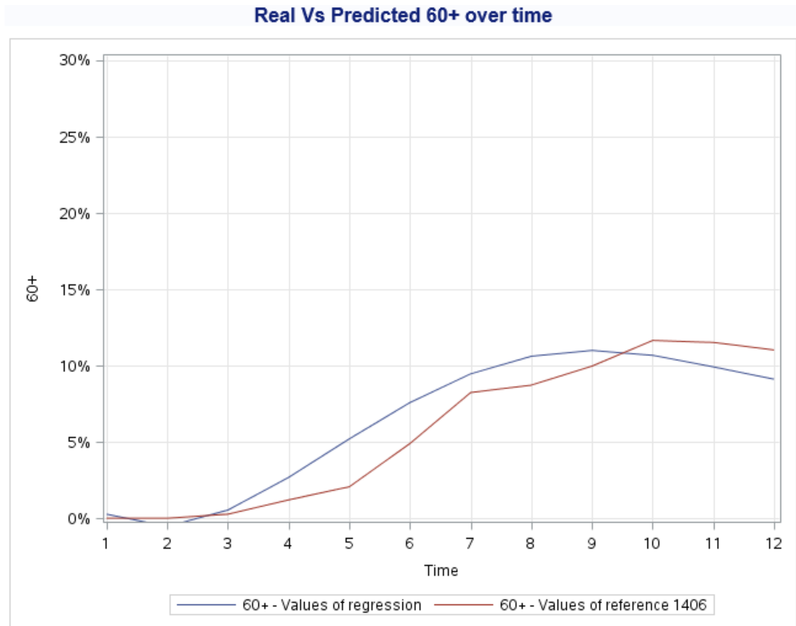

Figure 13: real vs predicted 60+

5. Popoulation Stability:

If the distribution of score ranges is compared with the generals of the population with credit cards in the same month, the PSI can be calculated, indicating a variation of new customers compared to the entire population. Clearly the point where the change in the distribution of new customers, compared to other customers with more seniority is minimal, can be interpreted as the point of maturity. In this case our new customers (MOB = 0) are compared with customers who are between one and a one a half years of seniority (12 <= MOB <= 18). (Figure 14: PSI over time)

In this case the limit PSI=10% is reached the ninth month after the origination.

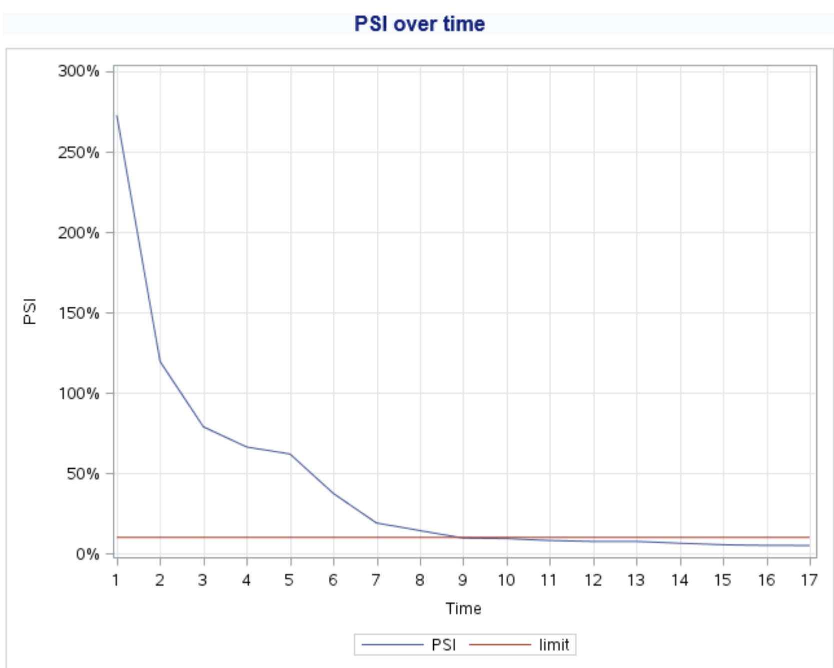

Figure 14: PSI over time

CONCLUSION

Through five analysis techniques could be observed that the maturation of the risk variables (score, and bad rate) is reached around the 8 and 9 months, on the other hand the financial variables (balance and utilization) mature in a shorter time, about the sixth month after origination period. Observing the model as a system, it is clearly evidenced through Markov chains and the population trend that the maturation is achieved between 8 and 10 months.

From the risk perspective it can be said that originations in cross-selling can be considered premature if made before the month 9 because they would have offered products to customers whose risk is not yet mature and may increase in the next three months; if the strategies are executed after the nine months with the maximum bad rates, the model will reach its good power of discrimination and originations will be more successful in terms of credit lines and PCL.

From the perspective of marketing and finance, it can be said that the originations on cross sales made after the ninth month may not be the most appropriate. This point of view seeks to identify the most profitable customers in terms of balance and utilization, which matures from the fifth month. A possible balance between risk perspective, contrasted with the other points of view can be put into consideration for its analysis in a future research.

REFERENCCES

Bradley, A. P. (4 de 2013). ROC curve equivalence using the Kolmogorov–Smirnov test. *Pattern Recognition Letters*, 470-475.

Kristin L. Sainani, P. (12 de 2014). Logistic Regression. *PM&R, 6*(12), 1157-1162.

Madhur Malik, L. C. (2012). Transition matrix models of consumer credit ratings. *International Journal of Forecasting, 28*, 261-272.

ACKNOWLEDGMENTS

This research was made posible through the academic and business support of Model Management of Multibanca Colpatria, special acknowledgment to the manager, the directors and all professional of the team.

Your comments and questions are valued and encouraged. Contact the author at:

Oscar Javier Cortés Arrigui
Banco Colpatria – ScotiaBank, Bogotá, Colombia
Universidad Nacional de Colombia, Bogotá
Cell phone: +57 3043611731
ojcortesa@unal.edu.co
www.linkedin.com/in/ojcortesa

Díaz Rodriguez Miguel Angel
Banco Colpatria – ScotiaBank, Bogota, Colombia
Cell phone: +57 3152666888
ma.diaz216@uniandes.edu.co
www.linkedin.com/in/migueldr