Paper 10081-2016

An Application of the PRINQUAL Procedure to Develop a Synthetic Index of Customer Value for a Colombian Financial Institution

Iván Darío Atehortua, Banco Colpatria Colombia; Diana Flórez Aza, Banco Colpatria Colombia

ABSTRACT

Colpatria, as a part of Scotiabank in Colombia, has several methodologies that enable us to have a vision of the customer from a risk perspective. However, the current trend in the financial sector is to have a global vision that involves aspects of risk as well as of profitability and customer loyalty. As a part of the business strategies to develop cross-sell and to loyal customer, our bank is currently focused exclusive in a Risk vision. With this paper, we propose a new methodology which allows to rank our customers from goods to bad customers with a more complete vision. In this paper, we estimate a synthetic index to know the real value of the client and it includes not only risk as profitability, loyalty and customer income. In order to generate the Index of Customer Value, we construct a synthetic index using principal component analysis. As result, we obtained a unique measure of value. When we compare our index with the current Risk Vision, we found very important opportunities and now we can focus in customers before unattended.

INTRODUCTION

Colpatria-Scotiabank is the fifth Bank in Colombia. Our principal Loan Products Credit Cards, Mortgages and Close-ends. As part of the past strategies of the bank each product tried to grow alone and we obtained good results. New policies and strategies on the bank tend to make strategies to customer level as not as product level. In this context, statistical tools are an excellent way to know and segment customers by different approaches.

In this paper, we use the component principal analysis to develop a unique joined index to measure the "value" of our customers focused in risk, profitability, loyalty and income level. The idea of this synthetic index is to know who are our best costumers but not only from a risk perspective as a joined perspective.

Currently the bank depending on the area the customer only seen from a unique perspective of profitability, balance or possession and use of each product separately, even often not taken into account customer behavior in other banks in the system financial.

DATABASE DESCRIPTION

Used data base contains 954.899 for all customers with 6 months on books and current balance greater than zero. 73 variables were analyzed which include Income estimators, financial sector, balances and credit limits, Risk and Profitability variables.

Variables where grouped in 4 segments using expert criteria:

Risk Variables

Financial System

Profitability variables

Loyalty variables

Variable selection had two principal criteria: First, were excluded variables wich had not relation with the variable segment. In this case we used the Kaiser's measure of sampling adequacy and the analysis over the correlation matrix; as a second criteria, we exclude variables with low communality, the communality for a given variable can be interpreted as the proportion of variation in that variable explained by the factors.

The treatment for variables include exclude outliers through the leverage statistic which is high for customers whose variables are significantly different. Selected variables (see Table 1).

Table 1. Selected variables for the analysis.

Loyalty
Last twelve months balance
Current balance
Current credit limit
Personal loans balance
Credit card limit
Quanto

Financial System
Credit card
number of credit products
Current balance
Personal Loans current balance
Balance of credit products

Risk
Máximum mora strip last 03 months
Maximun mora strip last 12 months
Probability that the customer is good
Score Bureau

Profitability
ROA
Margin 12 months

METHODOLOGY

In many situations we are interested in understand a complex problem and starting work, we have a lot of variables which represent specific items all interesting for the research. In this cases however, to study a big number of variables is complex and we can lost the objective of the problem. To solve this situation we can use the principal components analysis (PCA).

The PCA has as objective the reduction of the number of variables in set of variables maintaining the major quantity of information as possible. This is a mathematic procedure in which we do not assume any distribution (Diaz,2002)

The objectives of the principal components are:

- To generate new variables which show information contained in a dataset.
- To reduce the number of variables to analyze.
- To drop variables which does not provide relevant information.
- To facilitate the interpretation of the information.

The principal idea of the PCA is to calculate in a few variables called factors the major variability contained in de data. This factor are orthogonal and independents. The PCA is useful where analyzed variables have high correlation.

THE ACP ANALYSIS

Let a random vector $X' = (X_1, \dots X_p)$ with covariance Matrix Σ, and without loss of generality consider X with mean cero. We want to find the $\Gamma' = (\gamma_{11}, \dots \gamma_{1p})$ such that variance of $\Gamma' X$ is maximum. We can increase the variance as much as we increase Γ', then we maintain the restriction $\sum_{j=1}^{p} \gamma_{1j}^2 = 1$.

This problem can be written as

$$(1)\ \ (\Sigma - \lambda_1 I)\Gamma_1 = 0$$

Solution for equation (1) correspond to eingenvectors and eigenvalues for matrix Σ.

The eigenvector associated to the higher eingenvector is the first principal component and so on. Finally, we have (p) principal components associated to (p) eigenvalues.

With the calculated principal components, we can apply a linear transformation to X:

$$(2)\ \ Y = \Gamma' X$$

Y represents the same individuals but in other axes.

$$(3)\ \ cov(Y) = \Gamma' \Sigma \Gamma$$

(3) is a diagonalization for Σ, then principal components are uncorrelated. As it is a diagonal matrix, then the total variance of the factors is the sum of the variances of the principal components which is equal to the sum of the eigenvalues.

As each eigenvector represents the variance of each principal component, we can calculate the relative variance associated to each principal component:

$$(4)\ \ \frac{\lambda_j}{\sum_{j=1}^{p} \lambda_j}$$

If variables have high correlation, then expression (4) is high for the first component, then using a few components, you could explain a large proportion of the variability of the date without use too many variables.

As the covariances matrix is affected by the scale of the variables, in this document we calculated the principal componets.

The steps that we follow were:

- Join up variables whit similar characteristics into categories.
- Review each variable about the business importance, correlation, missing values and outliers.
- Transform the ordinal variables with the PRINQUAL methodology.
- Develop a principal component analysis for each category with the transformed variables.
- Review each variable about the KMO and Communality.
- Take the first principal component for create a simple index for the category.
- Group the different indices for the categories in a final synthetic index.
- Join up the final score into five groups.
- Describe the different group for verify then order of the variables that compound the category.

With the PROC PRINQUAL we used the monotone transformation for the ordinal variables, this transform permit the use of principal component on ordinal variables. The Methods of Variable Transformation that we used was Maximum Total Variance (MTV) Method (Young, Takane, and de Leeuw; 1978) is based on the principal component model, and it attempts to maximize the sum of the first r eigenvalues of the covariance matrix. This method transforms variables to be (in a least-squares sense) as similar to linear combinations of r principal component score variables as possible, where r can be much smaller than the number of variables. This maximizes the total variance of the first r components (the trace of the covariance matrix of the first r principal components).

RESULTS

The output from a principal component is a table of factor scores or weights for each variable (see Table 3). Generally, a variable with a positive factor score is associated with higher value in the group, and conversely a variable with a negative factor score is associated with lower value in the group.

The final communality estimates show that all the variables are well accounted for by the two components, with final communality estimates ranging from 0.62 to 0.91 for Loyalty, from 0.69 to 0.90 for Financial System, from 0.85 to 0.99 for Risk. (see Table 2)

Table 2. Final Communality by category

Loyalty

Final Communality Estimates: Total = 4.751434					
Last twelve months balance	Current balance	Current credit limit	Personal loans balance	Credit card limit	Quanto
0.89079586	0.84480858	0.91528464	0.62532149	0.81630534	0.65891799

Financial System

Final Communality Estimates: Total = 4.163155				
# Credit card	# number of credit products	Current balance	Personal Loans current balance	Balance of credit products
0.90332433	0.86846998	0.69363511	0.86659004	0.83113527

Risk

Final Communality Estimates: Total = 3.677031			
Probability that the customer is good	Score Bureau	Máximum mora strip last 03 months	Maximun mora strip last 12 months
0.90639876	0.99741474	0.91589283	0.8573251

Profitability

Final Communality Estimates: Total = 2.000000	
Margin 12 months	ROA
1	1

As we constructed a separate index for each variables group, we found for each group the factor scores are positive for all variables about possession and use of products and good risk behavior. Variables like bad risk behavior has a negative factor score.

For the group of variables (loyalty) the first principal component showed that for the customers the weights were concentrated on current balance, last twelve months and the credit limit. For the group of variables (Financial System) the first principal component showed that the weights were concentrated on number of credit cards in the financial system, number of credit products and balance of credit products. For the group of variables (Risk) the first principal component showed that the weights were concentrated on probability that the customer is good. For the group of variables (Profitability) the first principal component showed that the weights were concentrated on the ROA. (see Table 3)

Table3. Rotated Factor Pattern by category

Rotated Factor Pattern		
Loyalty	Factor1	Factor2
Last twelve months balance	0,936	0,12134
Current balance	0,90847	0,14011
Current credit limit	0,83819	0,46116
Personal loans balance	0,78895	-0,05039
Credit card limit	-0,00823	0,90333
Quanto	0,20219	0,78598

Rotated Factor Pattern		
Financial System	Factor1	Factor2
# Credit card	0,94052	0,13694
# number of credit products	0,86515	0,34638
Current balance	0,79579	0,24568
Personal Loans current balance	0,16075	0,91692
Balance of credit products	0,32741	0,85085

Rotated Factor Pattern		
Risk	Factor1	Factor2
Máximum mora strip last 03 months	-0,93017	-0,2251
Maximun mora strip last 12 months	-0,85726	-0,34991
Probability that the customer is good	0,90456	0,29694
Score Bureau	0,29584	0,95388

Rotated Factor Pattern		
Profitability	Factor1	Factor2
ROA	0,99728	0,07375
Margin 12 months	0,07375	0,99728

Using the factor scores from the first principal component as weights for each group of variables, a dependent variable can be constructed for each customer which has a mean equal to zero, and a standard deviation equal to one. This dependent variable can be regarded as the customer score in each group. We developed the synthetic index (Customer Value Index) whit the sum of loyalty, Risk and Profitability index and we let the Financial System like independent index because we want to join the bank index with the financial system index for develop different strategies.

From the Customer Value Index we developed five groups of customer whit different behaviors. In next Table we show the means for each variable where group number 4 is the group with higher index and 0 is the group with lower index (See Table 4).

Table 4. Means for each variable.

	Groups				
Variable	0	1	2	3	4
Last twelve months balance	2.052.567	3.000.067	3.178.244	2.903.191	10.490.513
Current balance	2.073.359	3.062.542	3.254.421	3.093.272	12.254.601
Current credit limit	4.255.542	6.391.995	6.778.287	7.315.431	15.801.527
saldo_cons	314.666	407.158	698.013	1.064.107	5.991.250
cupo_tdc	3.392.780	5.298.661	5.256.812	5.218.476	3.477.842
quanto	3.770	4.195	3.981	4.004	4.118
Maximun mora strip last 03 months	1,68	0,56	0,30	0,09	0,09
Maximun mora strip last 12 months	2,06	0,97	0,65	0,45	0,27
Probability that the customer is good	0,6077	0,8256	0,8898	0,9271	0,9393
Score Bureau	641	723	741	747	685
ROA	-1891,9%	13,1%	22,2%	198,8%	245,7%
Margin 12 months	85.961	274.021	404.807	437.324	776.785
# Credit card	1,6	**1,9**	1,7	1,7	1,7
# number of credit products	2,7	**3,0**	2,8	2,7	2,7
Current balance	2.618	**3.356**	2.961	2.467	2.670

Personal Loans current balance	5.599	**7.325**	6.504	5.618	5.646
Balance of credit products	13.356	**16.746**	14.690	13.026	12.812

We should review the customer potential looking for customers with low index in the bank and high index in the financial system may be for a portfolio purchase strategy. For example 16.4% (green zone) of the customer have a good deepening with other banks but a bad deepening with our bank (potential customers) while another 16.6% (blue zone) of the customer have a good deepening with our bank but a bad deepening with other banks (loyalty customers) (See Table 5).

Table 5. Index bank vs Index Financial System

	FINANCIAL SYSTEM INDEX (Quantity)					
BANK INDEX	**0**	**1**	**2**	**3**	**4**	**Total**
0	37.775	37.210	36.442	35.117	33.885	180.429
1	33.106	32.746	35.475	37.366	41.715	180.408
2	35.851	34.548	36.699	36.649	36.658	180.405
3	38.638	36.430	36.174	35.686	33.468	180.396
4	37.992	36.499	35.629	35.591	34.694	180.405
Total	183.362	177.433	180.419	180.409	180.420	902.043

	FINANCIAL SYSTEM INDEX (% POP)					
BANK INDEX	**0**	**1**	**2**	**3**	**4**	**Total**
0	4,2%	4,1%	4,0%	**3,9%**	**3,8%**	20,0%
1	3,7%	3,6%	3,9%	**4,1%**	**4,6%**	20,0%
2	4,0%	3,8%	4,1%	4,1%	**4,1%**	20,0%
3	4,3%	4,0%	4,0%	4,0%	3,7%	20,0%
4	4,2%	4,0%	3,9%	3,9%	3,8%	20,0%
Total	20,3%	19,7%	20,0%	20,0%	20,0%	100,0%

CONCLUSION

The synthetic index of value can help to make decisions for different areas of the bank like the commercial area, the risk area, the business support area among other. Usually we have a lot of variables in our data warehouse which have different formats as ordinal, interval or categorical. In these cases is important to have a methodology that allow us to analyze this type of data for this reason is important the new procedures developed by SAS ®.

Whit these synthetic index the bank can develop strategies for different types of customers in some cases for **deepen** customer relationship with low index bank and high index in the financial system, **retain** customer with high index in the bank or may be **let go** customers with low index bank taking into account the different dimensions of customer value.

The usefulness of the index is that for the bank that can be used to make increases in quotas, differentiated interest rates offered or offer new products. With the index, it is also possible to segment the population in a more appropriate manner. Today segmentation Bank takes into account only the income level of the customer, however some customers with high incomes today are treated as VIP but are high risk and low profitability while other customers are treated as Pioneers but have high profitability and low risk.

REFERENCES

The PRINQUAL Procedure. SAS®.
https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#prinqual_toc.htm.

Becerra Avella, 2010. “Comparison between multiple factor analysis (MFA) and principal component analysis for qualitative data (Prinqual) methods for derivation of indices - See more at: http://www.bdigital.unal.edu.co/3029/#sthash.TRFLah0k.dpuf”.

SAS/STAT® 9.2 User’s Guide The PRINQUAL Procedure (Book Excerpt).

CONTACT INFORMATION

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

Iván Darío Atehortua
Banco Colpatria
7456300 Ext 3165
atehori@colpatria.com

Diana Flórez Aza
Banco Colpatria
7456300 Ext 3692
florezdi@colpatria.com