How SAS Tools Helps Pricing Auto Insurance

Mattos, Anna and Meireles, Edgar / SulAmérica Seguros

ABSTRACT

In an increasingly dynamic and complex market such as auto insurance, it is absolutely mandatory to construct advanced pricing methodologies with strong predictive power in order to ensure profitability.

Due to our business being in a country with continental dimensions, finding the right price in face great risk and customer behavior disparity is a major challenge. Therefore, we developed methodologies for risk discrimination, geo classification, forecasting and optimization using SAS platforms. These have allowed us to build a more holistic view of our business and integrate pricing intelligence across risk and demand elasticity profiles.

INTRODUCTION

SulAmérica operates in multiple insurance segments, such as health and dental, auto and other property and casualty, The Company also offers life and personal accident insurance as well as asset management, private pension, and savings bonds products.

In 2016, the Company's operating revenues totaled R$16.8 billion, with R$16.0 billion in insurance revenues. On December 31, 2016, SulAmérica Investimentos, the asset management business, had R$34.2 billion in assets under management.

With approximately five thousand employees, SulAmérica conducts business through an extensive and diversified distribution network that includes more than 30,000 independent brokers, more than 20 distribution partnerships, and 16,000 bank branches accessible through bancassurance partners.

With approximately 7.0 million customers (individuals and companies) and 100% of revenues generated in Brazil, SulAmérica's administrative headquarters is located at Rua Beatriz Larragoiti Lucas, 121 in Rio de Janeiro. In 2016, 10 new branches were opened, totaling 90 in all of Brazil. The geographic distribution of the Company’s branches is shown in the figure below as follows:

THE AUTO INSURANCE MARKET

The auto insurance market is different from the traditional market such as retail, so the insurer only knows the exact final cost at end of the term of the risk. In order to price, there are a few key elements: administrative expenses, marketing expenses, financial results in addition to other expenses in order to guarantee the company's net profit.

In order to estimate the company result it is imperative to price the portfolio based on the probability of claim on each individual policy.

Due to the great diversity of customer profiles and risk behaviors, the automobile insurance pricing becomes very challenging and it is an art to be able to price with robustness and assertiveness. As a result, it is very important to develop methodologies to estimate the full operation cost.

METHODOLOGY

RISK ANALYSIS

The secret to monetize an insurance policy is to invest in a thorough risk analysis. Each policy reflect different views of risk specially the customer, vehicle and the location (zip code), among others.

It is important to organize all the different views of risk in an algorithm. For that reason, predictive models are developed. A predict model is an statistical technique that uses historical data to accurately identify patterns of behavior and thus to calculate the probability of claims.

When developing the predictive models, we identify claim behaviour patterns of each variable of the policy.

For example, to price the property, with involves the collision and theft risks, we calculate the following risks for each individual policy:

- The probability and Estimated Severity of a Theft Claim
- The probability and Estimated Severity of a Collision Claim

This way, it's possible to identify at the time of quotation what the probability is of a collision or theft to occur during the term of the policy. Based on the combination of all probabilities, we calculate the Risk Premium to be charged.

As a result, it is possible to price the customer correctly choosing in advance who you would like to keep or not in your portfolio, positioning the price strategically. Risk analysis is the foundation to insure profitability in pricing.

Our predictive models were developed by the Enterprise SAS Guide and SAS Miner. These tools help to estimate the frequency and severity of the risk. We work with a five-year historical database in order to guarantee significant volume without noise and deep analysis.

We have been able to map the risk of the entire SulAmerica Auto portfolio.

GEO CLASSIFICATION

We have developed a Geo-Classification in order to identify the different patterns of risk behaviours around the country.

Brazil is a very large country, bigger than the US mainland.

We are faced today with areas geographically close to each other but with risk levels of Theft and Collision totally different.

Geo classification is the geographic location group based on indicators and pre-established metrics, ensuring homogeneous group risks among themselves and heterogeneous among groups.

The benefit of using geo classification in the pricing process is to guarantee singleness on pricing in a region with similar risk level.

In order to achieve that, a georeference study was conducted from the claim indicators. We have been studying theft occurrences and how they were geographically distributed.

In the geo classification risk front, it was imperative to work with a tool that would enable the entire country to be broken down into large segments, ensuring uniformity of risk within each group. The technique used to develop the geo classification was the Decision Tree and it was developed in the SAS Miner tool.

FORECASTING METHODS

The forecast methods is an important component of the estimate cost. This is the only way we can predict trend of rate loss during the term of policy, which is added to the premium risk model.

In statistics, a time series is a sequence of numeric data in successive order, usually at regular intervals. Therefore, its objective is to identify non-random behaviors in a sequence of variable of interest and the observation of this past behavior can make predictions about a future decision.

For example, we can think of a time series ( ) as the combination of three components: Seasonal, a cyclic-trend ( ) and a component for all that remains, random. A time series model can be described as follows:

The SAS provides several solutions for interpretation and forecasting in time series models, but we use the SAS Miner and its specific nodes to exemplify the techniques used for decision-making.

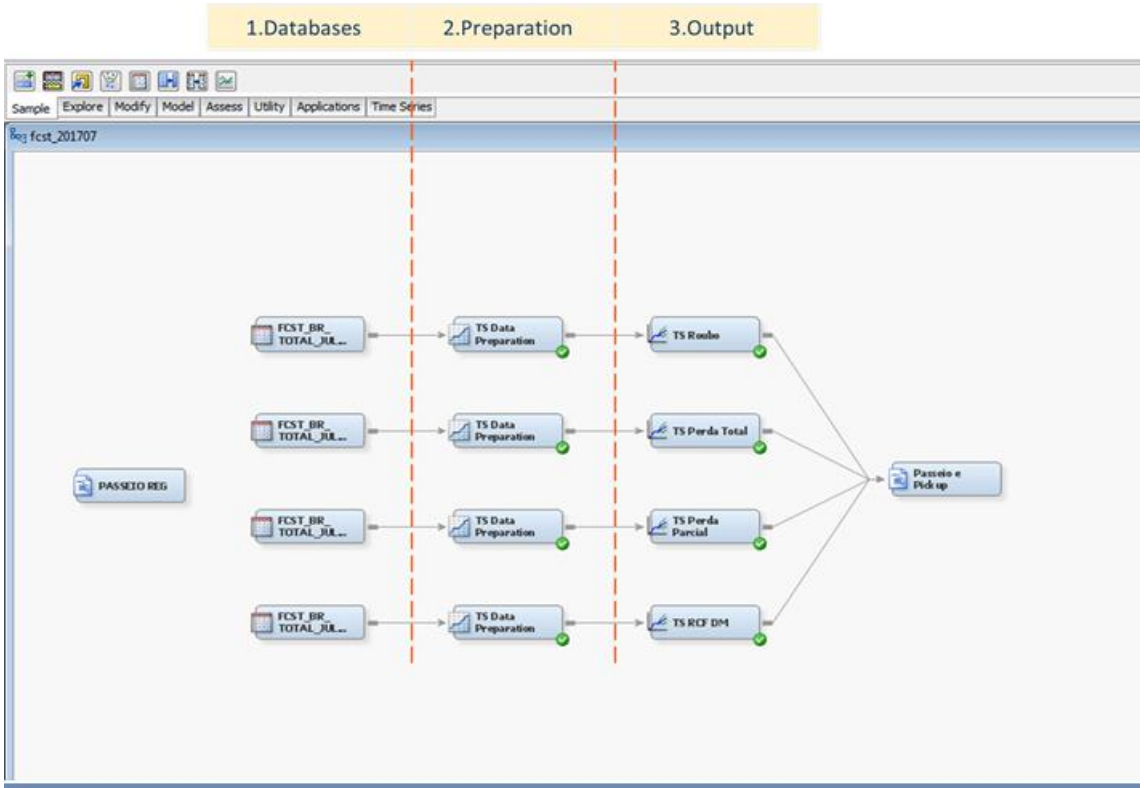

Figure x – SASMiner

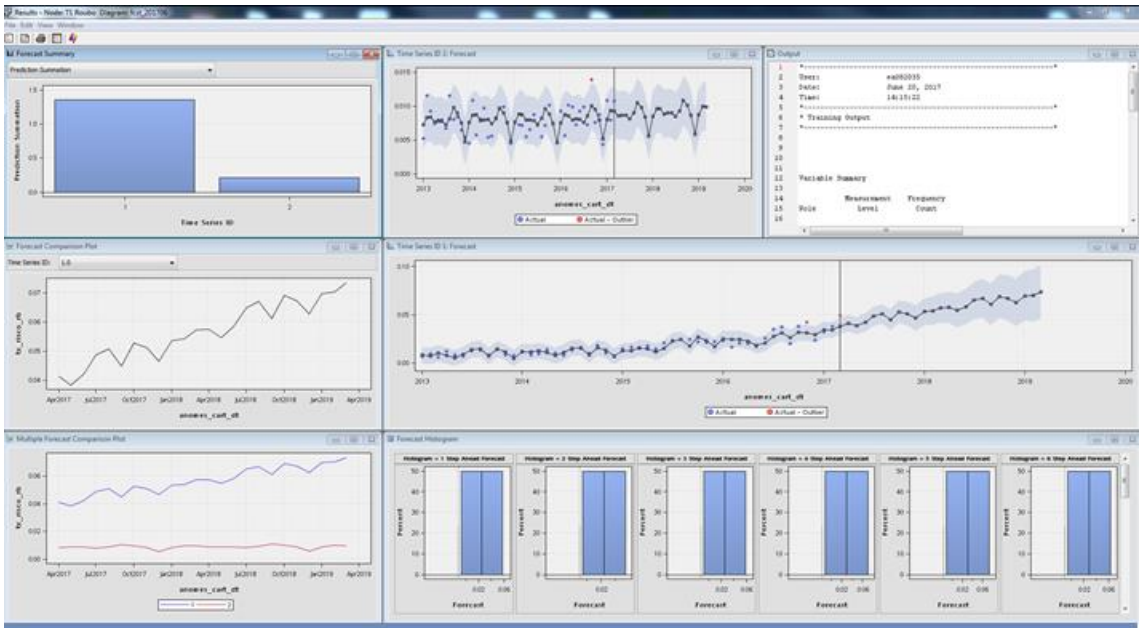

Figure x – SASMiner Output

PRICE OPTIMIZATION

Once the expected cost for each individual customer is defined, along with the probability of the closing and renewal, we need to decide the premium price associated to the risk maximizing our revenue. Given the above, we are facing an operational research problem.

Optimization, as far as the pricing process is concern, is the best result of an objective function combined with statistics and mathematical models. Its main objective is to maximize the profitability of the portfolio by developing scientific methods to solve complex problems taking into consideration the probability of risk and the profile range of the client, determining the target price for each individual police, according to business restrictions.

In general, several types of models are used in operational research such as, linear programing, non-linear programing and Monte Carlo simulation. We use linear programing technique in our model which is applied to models whose objective functions and restrictions are linear.

In order to put optimization into practice, it is necessary a robust technology, which allows several simulations to find the best solution to a problem, with a speed much faster than the human capacity.

Through SAS Operational research solution, we were able to determine the best solver with a very high processing capacity which helped us significantly in making more agile decisions.

In order to implement the model we used the SAS 9.4 along with OR/SAS solution according to following steps:

- Preparing the possible solution matrix;
- Defining the parameters of the optimization algorithms;
- Running and obtaining the solver.

Since it is a complex process of price definition and for combining several predictive models, we calculated in advance, in a matrix, all the renewal probabilities of a certain group of customers taking into consideration predetermined price variations. As a result, we achieve performance in the execution since the price is not calculated at the time of the execution, it is only consulted in the possible matrix solution.

	Fator Preço													
	1,8	1,7	1,6	1,5	1,4	1,3	1,2	1,1	1	0,9	0,8	0,7	0,6	0,5
Cliente 1	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)
Cliente 2	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)
Cliente 3	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)
.														
.														
.														
Cliente n	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)	P(x)

Figure x – Example: Possible Solutions Matrix

Once the matrix is developed, the next step is to configure all the input parameters of the algorithm optimization, including the objective function and the business rules.

```
%macro otimizacao ( Cenario , Regional , Regional_Aux , Sinistralidade_Min ,
Sinistralidade_Max , Prob_Retencao_Min , Prob_Retencao_Max );

proc optmodel;
        set <str> NUM_DOCS;
        set <str, num> PARES;
        num Margem{PARES};
        num Premio_Esperado{PARES};
        num Sinistro_Esperado{PARES};
```

```
        num Tx_Ret_Estimada{PARES};

        read data base_completa (where=(regional = "&Regional."  and &fatores_pares.))
into    PARES=[NUM_DOCUMENTO    Fator]    Margem    Premio_Esperado    Sinistro_Esperado
Tx_Ret_Estimada;
/*      print Margem Premio_Retido Sinistro_Retido Tx_Ret_Estimada;*/

        NUM_DOCS = {setof{<i,j> in PARES} i};
/*      put NUM_DOCS;*/

        var Fator_Escolhido {PARES} binary;

        impvar Sinistralidade_numerador = (sum{<i,j> in PARES} Sinistro_Esperado[i,j] *
Fator_Escolhido[i,j]);
        impvar Sinistralidade_denominador =  (sum{<i,j> in PARES} Premio_Esperado[i,j] *
Fator_Escolhido[i,j]);

        con     r1:     Sinistralidade_numerador     >=     Sinistralidade_denominador     *
&Sinistralidade_Min;
        con     r2:     Sinistralidade_numerador     <=     Sinistralidade_denominador     *
&Sinistralidade_Max;

        impvar  Prob_Retencao_Total  =  (sum{<i,j>  in  PARES}  Tx_Ret_Estimada[i,j]  *
Fator_Escolhido[i,j]) / card(NUM_DOCS);

        con r3: Prob_Retencao_Total >= &Prob_Retencao_Min;
        con r4: Prob_Retencao_Total <= &Prob_Retencao_Max;

        con r5 {i in NUM_DOCS}: sum {<(i),j> in PARES} Fator_Escolhido[i,j] = 1;

        max Margem_Total = sum{<i,j> in PARES} Margem[i,j] * Fator_Escolhido[i,j];

        solve;

/*      print Fator_Escolhido;*/
/*      expand;*/

        create data base_completa_saida_&cenario._&regional_aux. (where=(Fator_Escolhido
= 1)) from [NUM_DOCUMENTO Fator] Fator_Escolhido;

run;
```

The last step is to develop the result output. So that after running, it is necessary to check the effectiveness of the optimization.

The OPTMODEL Procedure

Problem Summary	
Objective Sense	Maximization
Objective Function	Margem_Total
Objective Type	Linear
Number of Variables	964797
Bounded Above	0
Bounded Below	0
Bounded Below and Above	964797
Free	0
Fixed	0
Binary	964797
Integer	0
Number of Constraints	28505
Linear LE (<=)	2
Linear EQ (=)	28492
Linear GE (>=)	11
Linear Range	0
Constraint Coefficients	6716466

Performance Information	
Execution Mode	On Client
Number of Threads	1

Solution Summary	
Solver	MILP
Algorithm	Branch and Cut
Objective Function	Margem_Total
Solution Status	Optimal within Relative Gap
Objective Value	8855519.3032
Iterations	93315
Best Bound	8855524.0539
Nodes	1
Relative Gap	5.3646232E-7
Absolute Gap	4.7506549973
Primal Infeasibility	0
Bound Infeasibility	0
Integer Infeasibility	0

CONCLUSION

The objective of this project is to present a complete and innovative way of pricing automobile insurance. In order to achieve assertiveness in pricing, statistical and mathematical techniques are fundamental to guarantee the required profitability.

Within the insurance process, we looked for tools that would enable us to develop and apply Risk Analysis, Geo-Classification, Forecasting Methods and Pricing Optimization, guaranteeing the expected assertiveness as well as the capacity and speed to support our processes.

We have found all these solutions at SAS in the following tools:

- SAS Guide: We use it to manipulate all the databases and for the development of predictive models.
- SAS Miner: We use it for the development of GLM models as well as for the GEO-Classification and Forecasting Methods.
- SAS/OR: We use it for Optimization Techniques.

Our next challenge is to automate some of these analysis processes through machine learning in order to allow the machines to recognize behavior patterns, enabling continuous learning and the application of the behavior changings to the new scenarios in an independent way.

REFERENCES

SulAmérica | Relações com Investidores - http://ri.sulamerica.com.br/

MORETTIN, Pedro Alberto; TOLOI, Clélia Maria de Castro. Análise de séries temporais.