CRC | Credit Research Centre

# CREDIT SCORING & CREDIT CONTROL XIII

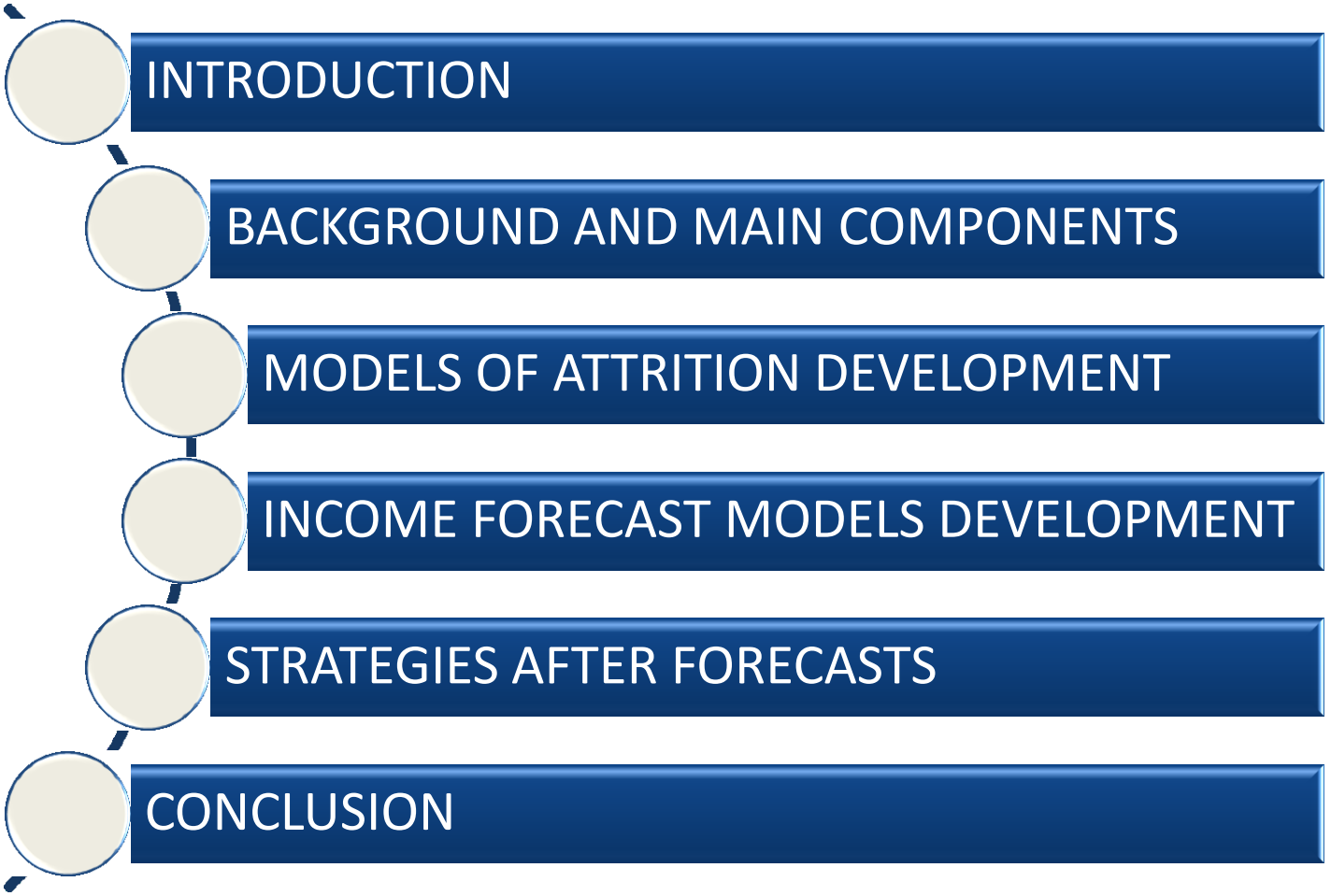
28-30 AUGUST 2013

# SCORING AND INCOME FORECAST FOR PENSION FUNDS IN COLOMBIA

**Gehiner Salamanca**
**Senior Partner**
**LiSim**

LiSim® A World of Knowledge

# AGENDA


INTRODUCTION
BACKGROUND AND MAIN COMPONENTS
MODELS OF ATTRITION DEVELOPMENT
INCOME FORECAST MODELS DEVELOPMENT
STRATEGIES AFTER FORECASTS
CONCLUSION

# INTRODUCTION

In Colombia, since 1993 the Severance funds, Mandatory and Voluntary Pensions went from state monopoly to a shared scheme with private companies (AFP) or pension funds. Now, the AFPs have about 11 million of members.

The income that the AFPs receive, come from a percentage on the managed resources. The more affiliate income the higher the value of the fund and the commission. The higher the dropout (transfer to another AFP or total or partial withdrawal from the fund), the respective commission for the AFP is reduced or disappears.

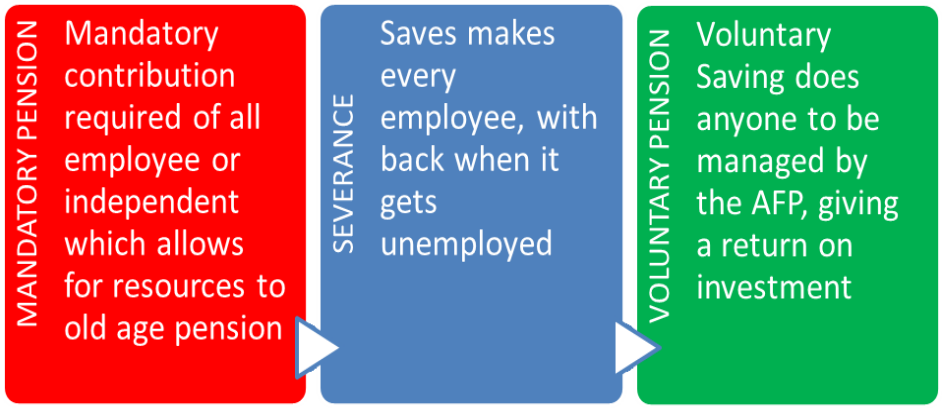
In this presentation will be shown how, in need of a pension fund in Colombia* to develop strategies to mitigate the risk of attrition of its affiliates and forecast the income they will gain in the future, LiSim managed to use Scoring and statistical methodologies to response to a need that affects more than 80% of the income of pension funds.

The paper will present the experience in the design and generation of attrition scoring models for AFP and income forecasting, also segmentation and strategies arising from them.

*Thanks to Colfondos (AFP) for allowing the development of this paper.

***DIFFERENT PRODUCTS OF THE AFP***

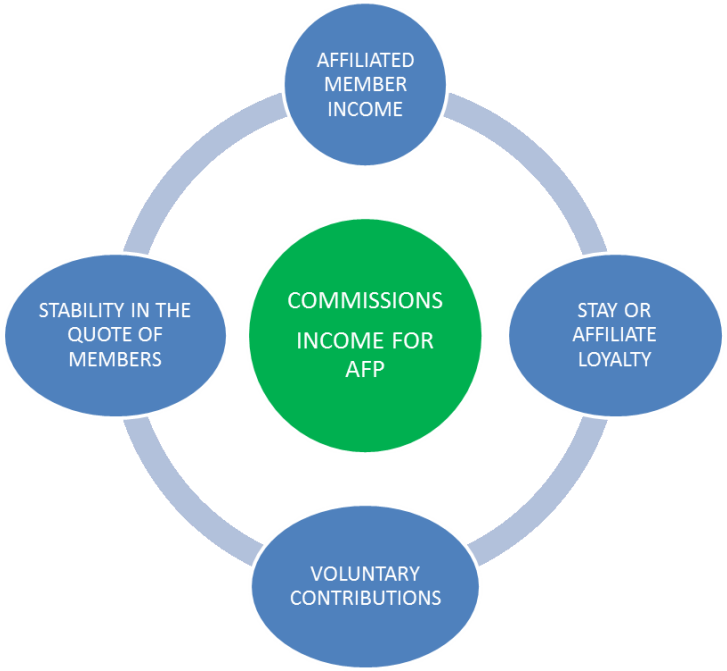
In general, the AFPs have the following products:




AFP's participation involves the administration and investment of these resources to ensure its return in the future to affiliated members.

***THE AFP´S COMMISSION SOURCES***

Keeping constant the rates or fees earned by AFP for every dollar received, the commission income generally depends on the sources shown in the graph:




Therefore, we developed a project focused on forecasting:

- Affiliated member income
- Loyalty rate of the affiliate
- Density: permanence rate of the quotes for Mandatory Pension
- Commission to be earned by the AFP from each affiliate.

This paper will focus on the design of income and loyalty forecasting models.

# AGENDA

- INTRODUCTION
- BACKGROUND AND MAIN COMPONENTS
- **MODELS OF ATTRITION DEVELOPMENT**
- INCOME FORECAST MODELS DEVELOPMENT
- STRATEGIES AFTER FORECASTS
- CONCLUSION

The aim of this model is, by calculating the dropping out probability of each affiliate, to guide the strategy and focus the resources for customer retention and improve dropout indicators for AFP.

***STATISTICAL MODEL***

The dependent variable is a dummy variable that is coded as 0 or 1 (respectively, "Deserters" and "No deserters"). Namely

$$y_{I\,i} = \begin{cases} 1, & if\ the\ ith\ individual\ is\ not\ deserter \\ 0, & otherwise \end{cases}$$

Be $y$ the number of hits of $m$ independent trials, each with success probability $\theta$, i.e. $y = \sum_{i=1}^{m} y_{I\,i}$. The random variable $y$ has then binomial distribution $y \sim Bin(m, \theta)$. The probability that $Y$ is equal to an integer $j = 0,1, \ldots, m$ is given by:

$$\Pr(y = j) = \binom{m}{j} \theta^{j} (1-\theta)^{(m-j)}$$

## Methodology

In binomial regression, $y_i$ response contains the number of events in $m_i$ trials, additionally there are $p$ predictors $x_i$ possible, including the intercept, and assume that the probability of being "no deserter" for the i-th case is $\theta(x_i)$. It can be written as:

$$(Y|X = x_i) \sim Bin(m_i, \theta(x_i)), i = 1, \dots, n$$

It is assumed that $\theta(x_i)$ depends on $x_i$ only through a linear combination $\beta' x_i$ for $\beta$ unknown. $\theta(x_i)$ can be written as $\beta' x_i$ function,

$$\theta(x_i) = M(\beta' x_i) = \frac{e^{\beta' x_i}}{1 + e^{\beta' x_i}} = \frac{1}{1 + e^{-\beta' x_i}}$$

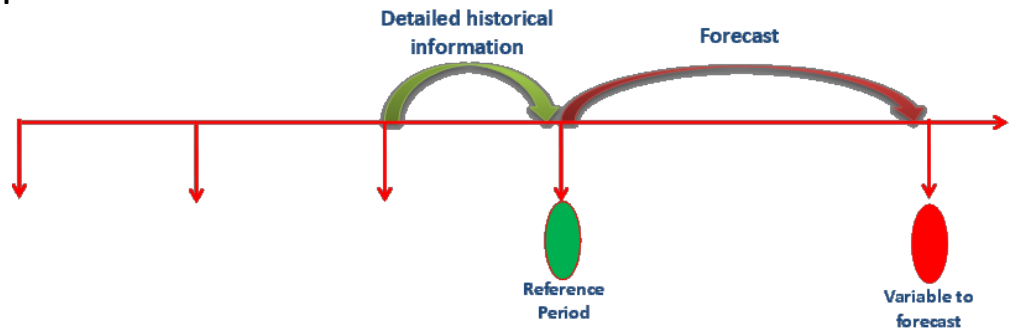
Linearity is achieved if the link is represented by the natural logarithm. Solving the above equation for $\beta' x_i$ we obtain:

$$\log\left(\frac{\theta(x)}{1 - \theta(x)}\right) = \beta' x$$

## Methodology

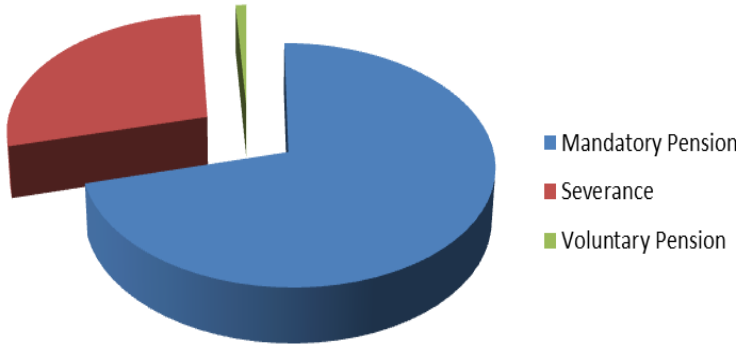
***DESIGN***

The methodology is applied to measure the probability of desertion in a period of two years for long-term products and of 12 months for shorter term, using membership and detailed historical information




Analysis was differentiated by product (Mandatory Pension, Severance and Voluntary Pension), creating independent models for each product.

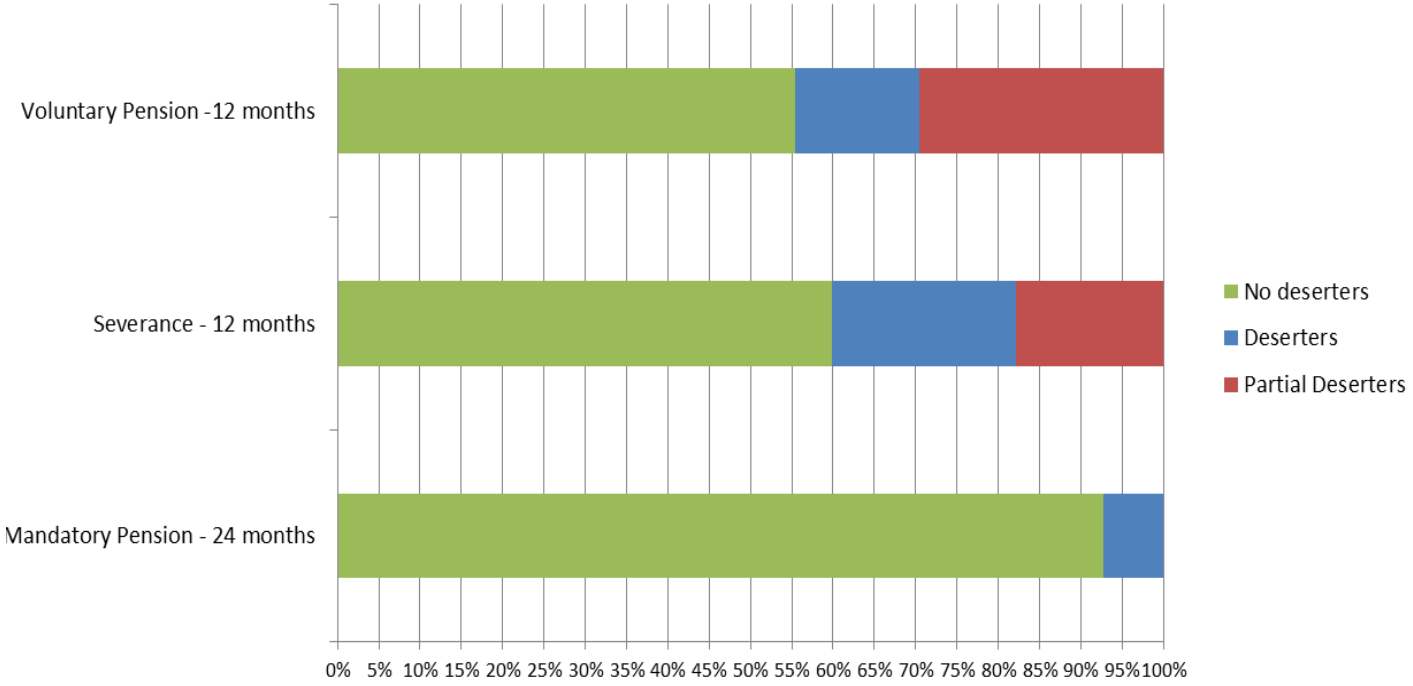
**PARTICIPATION BY PRODUCT**




As noted, most of the population is concentrated in Mandatory Pension. In this case we have more than 1.5 million records likely to develop statistical analysis using a large population.

***GOODS/BADS INDICATOR:***

Below we find the percentage of customers who show event of partial or total desertion (of funds) by product:




In this paper we focus on the model developed for Mandatory Pension.

***STATISTICAL ANALYSIS OF DATA (CORRELATION ANALYSIS)***

It is important to consider several factors, namely:

- AFP's experience and knowledge of LiSim in similar populations, which allows multiple views.
- Likewise, customers with “no data” in a variable must not exceed 10% of the sample.
- The characteristic to be selected must belong to a minority, but considering no to create volatility risk because of low population size of the feature to consider.
- Statistical discrimination of the variable using, for example, the ref%:

$$REF\% = \frac{Population\ Bads\% - Characteristic\ Bads\%}{Population\ Bads\ \%}$$

***SELECTED VARIABLES***

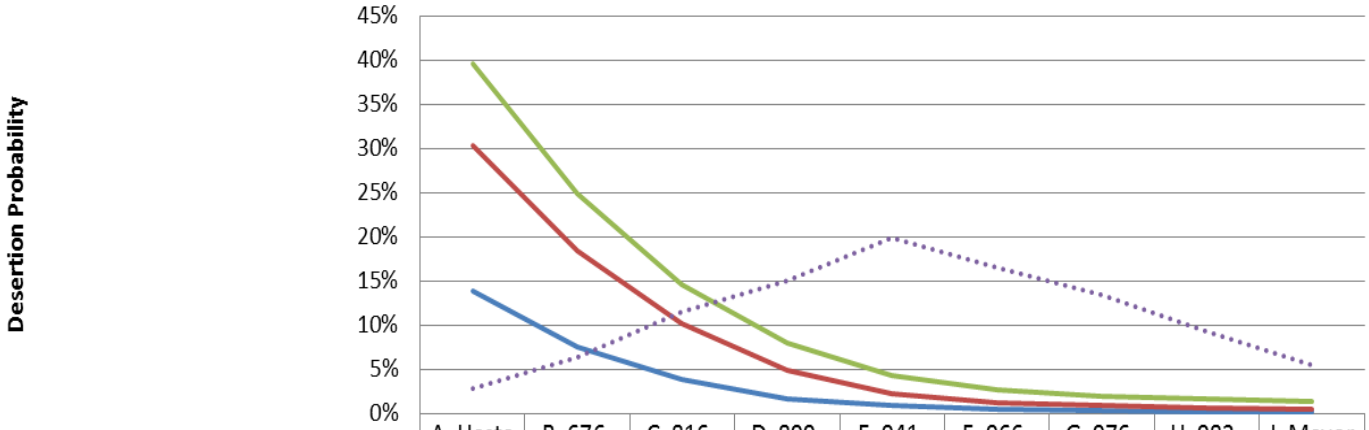
After analyzing variables which data is available, are selected those variables that are likely to enter in the score model, either to reward or punish, following the criteria in the correlation analysis.

In general were found as susceptible, more than 70 characteristics, belonging to different kinds of variables, including:

| SOCIODEMOGRAPHIC INFORMATION | EMPLOYER INFORMATION | EMPLOYER-CLIENT RELATIONSHIP | ORIGIN AFFILIATE | RELATIONSHIP WITH AFP |
|---|---|---|---|---|
| Age | Company Type | Seniority | Regional | Requests and complaints |
| Residence City | Size | | Type of member | Seniority with AFP |
| Income | Company Activity | | Sales channel | Internal segment AFP |
| Presence of beneficiaries | Economic Sector | | City sale | Fund value |
| | | | Origin Affiliate | |

## *SCORE DISTRIBUTION*

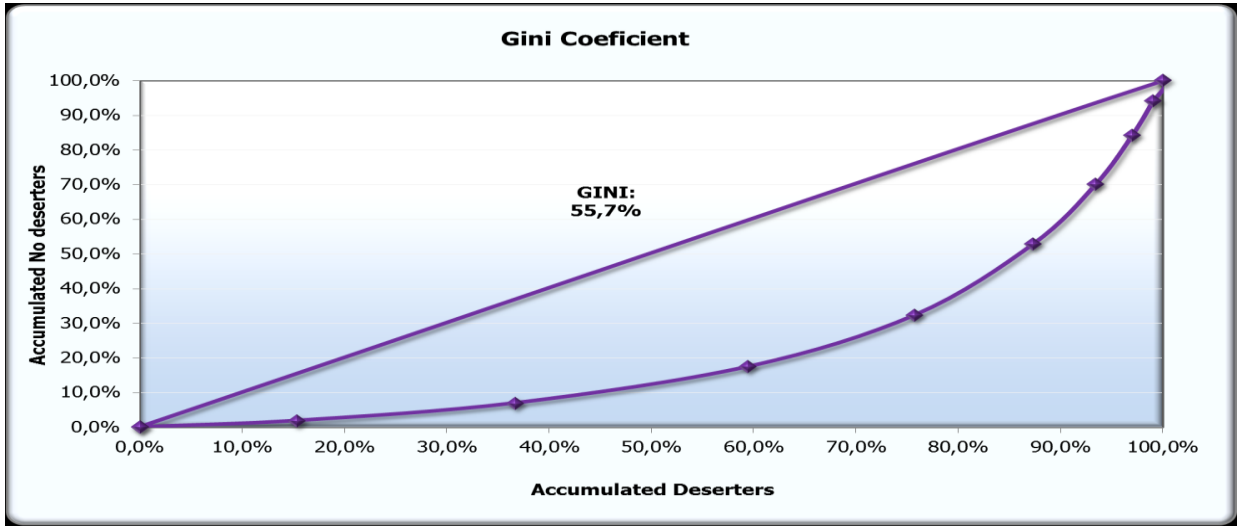
**SCORE DISTRIBUTION: Mandatory Pension - Desertion Probability**

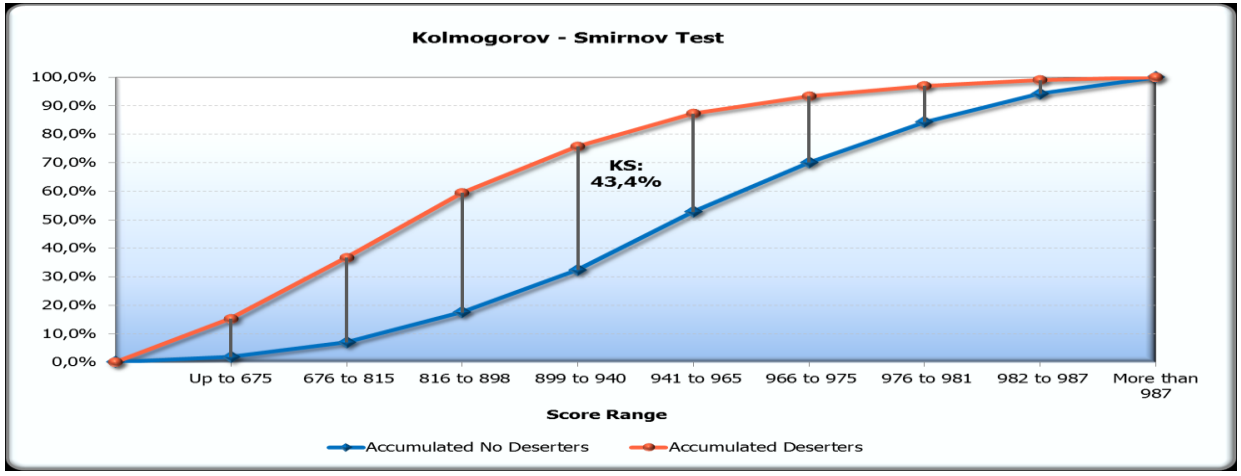



| | A. Hasta 675 | B. 676-815 | C. 816-898 | D. 899-940 | E. 941-965 | F. 966-975 | G. 976-981 | H. 982-987 | I. Mayor a 987 |
|---|---|---|---|---|---|---|---|---|---|
| 6 months Desertion probability | 14% | 8% | 4% | 2% | 1% | 0% | 0% | 0% | 0% |
| 12 months Desertion probability | 30% | 18% | 10% | 5% | 2% | 1% | 1% | 1% | 0% |
| 24 months Desertion probability | 40% | 25% | 14% | 8% | 4% | 3% | 2% | 2% | 1% |
| Total % | 2,8% | 6,3% | 11,4% | 15,0% | 19,8% | 16,5% | 13,4% | 9,3% | 5,5% |

| Score Range | Odds % | Accumulated Deserters % |
|---|---|---|
| Up to 675 | 1,5 | 15% |
| 676 to 815 | 3,0 | 37% |
| 816 to 898 | 5,9 | 59% |
| 899 to 940 | 11,7 | 76% |
| 941 to 965 | 22,7 | 87% |
| 966 to 975 | 36,0 | 93% |
| 976 to 981 | 50,7 | 97% |
| 982 to 987 | 62,4 | 99% |
| More than 987 | 72,7 | 100% |

## *TESTING THE MODEL*



# AGENDA

- INTRODUCTION
- BACKGROUND AND MAIN COMPONENTS
- MODELS OF ATTRITION DEVELOPMENT
- **INCOME FORECAST MODELS DEVELOPMENT**
- STRATEGIES AFTER FORECASTS
- CONCLUSION

## Methodology

For the calculation of the client's income was considered Mandatory Pension information. The income forecast was also carried out for other products in the AFP, in this paper is shown the developed for the Mandatory Pension.

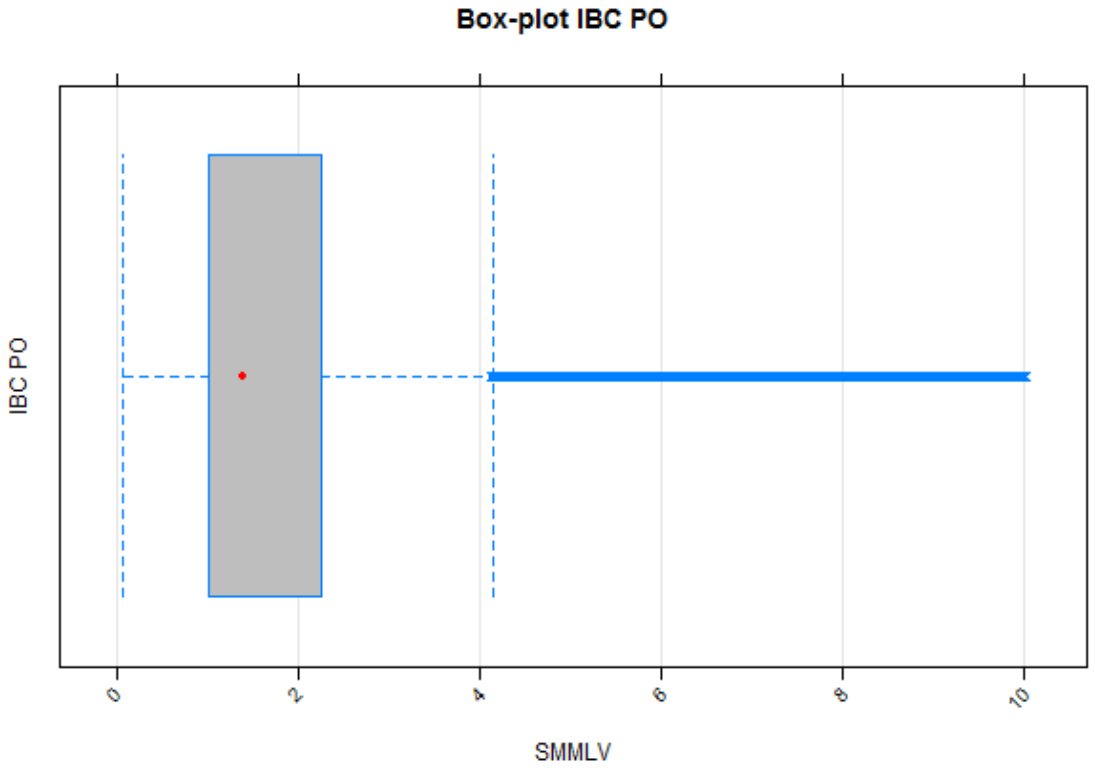
The calculation of income per member (IBC) for each of its contribution periods (n) was performed using the following methodology:

$$IBC_n = \frac{\sum_{i=1}^{m_n} S_i}{m_n} * 30$$

Where $n$ indicates the period of analysis, $m_n$ is the number of days listed on the period $n$ and $S_i$ the base salary on day $i$ of period $n$ .Through this calculation results in the IBC quote that the client has maintained monthly during the analysis period $n$.

***Box-Plot (IBC SMMLV*)***

The red dot refers to the median of the distribution and the gray box to the first and third quartile. Its position on the left side of the box indicates right asymmetry (This is because a high concentration of the population at low income), points outside the blue dotted lines refer to "outliers", which are concentrated at high values of IBC or income, demonstrating that the distribution has a heavy right tail.




*SMMLV: monthly minimum wage Colombia 2009

***STATISTICAL MODEL***

Let $Y$ be random variable with mean $\mu$ gamma distribution and variation coefficient $\phi^{-1/2}$ must be $Y \sim G(\mu, \phi)$ and its density function is given by:

$$f(y;\mu,\phi) = \frac{1}{\Gamma(\phi)}\left(\frac{\phi y}{\mu}\right)^{\phi} \exp\left(-\frac{\phi y}{\mu}\right) \mathrm{d}(\log y)$$

$$\Gamma(\phi) = \int_0^{\infty} t^{\phi-1} e^{-t} dt$$

It is observed that for high values of $\phi$ the function becomes more symmetrical about the mean and shows that as $\phi$ increases $Y$ has a normal distribution with mean $\mu$ and variance $\mu^2\phi^{-1}$

## Methodology

Let $Y_1, \dots, Y_n$ be random variables independent and identically distributed by $Y_i \sim G(\mu_i, \phi)$, i.e. it is assumed that the different variables have the same mean and coefficient of variation $\phi^{-1/2}$. It is further assumed that $g(\mu_i) = \eta_i$, with $\eta_i = {X_i}^t \beta, X_i = (x_{i1}, \dots, x_{ip})$.

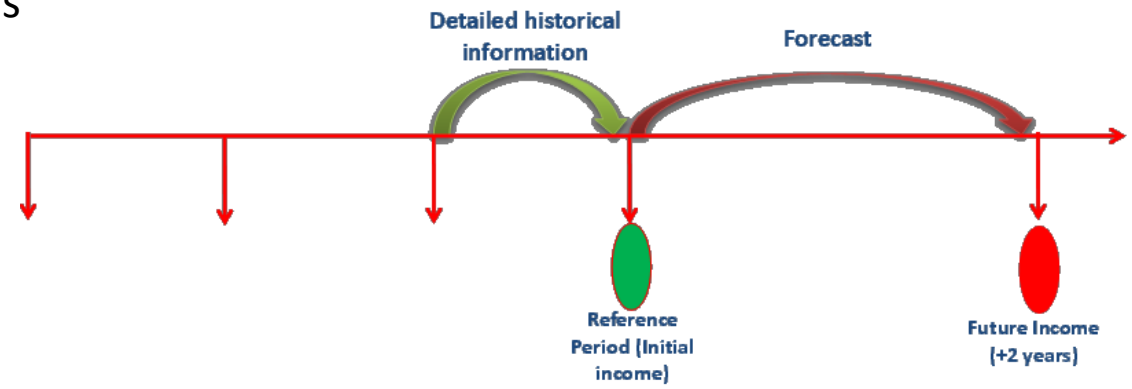
After defining the model components, it is defined which link function ($g(\mu_i)$) will be used to be modeled. The most common functions for the link $g(\mu_i)$ are the identity ($\mu_i = \eta_i$), logarithmic ($\log(\mu_i) = \eta_i$) and the reciprocal ($\mu_i = {\eta_i}^{-1}$).

Among the most known and used transformations are those proposed by Box and Cox (1964), which becomes positive and the value observed in

$$z = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, & if\ \lambda \neq 0 \\ logy\ , & if\ \lambda = 0, \end{cases}$$

***DESIGN***

After defining the variable of the model (IBC), the form in which the forecast is designed is as follows

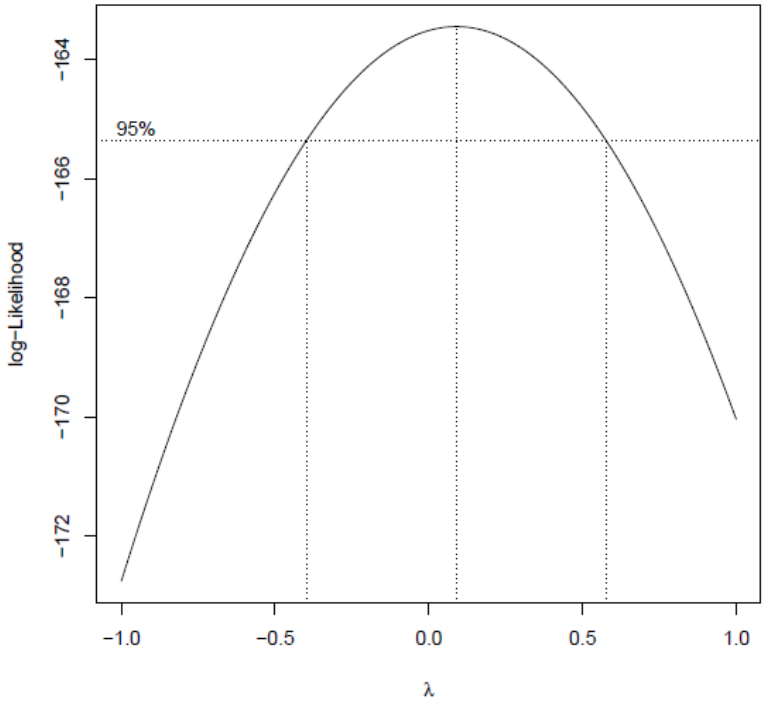



It aims to find the income or IBC at a future time, based on historical economic and demographic information since the moment of the affiliation.

To develop the model population was required to have recent income data in the analysis period, and two years later, which is the forecast horizon. The population with these characteristics is close to 500 thousand records.

Then is calculated and plotted the profile log-likelihood for different $\lambda$ in the transformation of the Box-Cox power:

For values $\lambda \approx 0$ the log-plausible is the lowest absolute value generating a better fit in the model, then the transformation cox box to analyze will be $z = logy$.

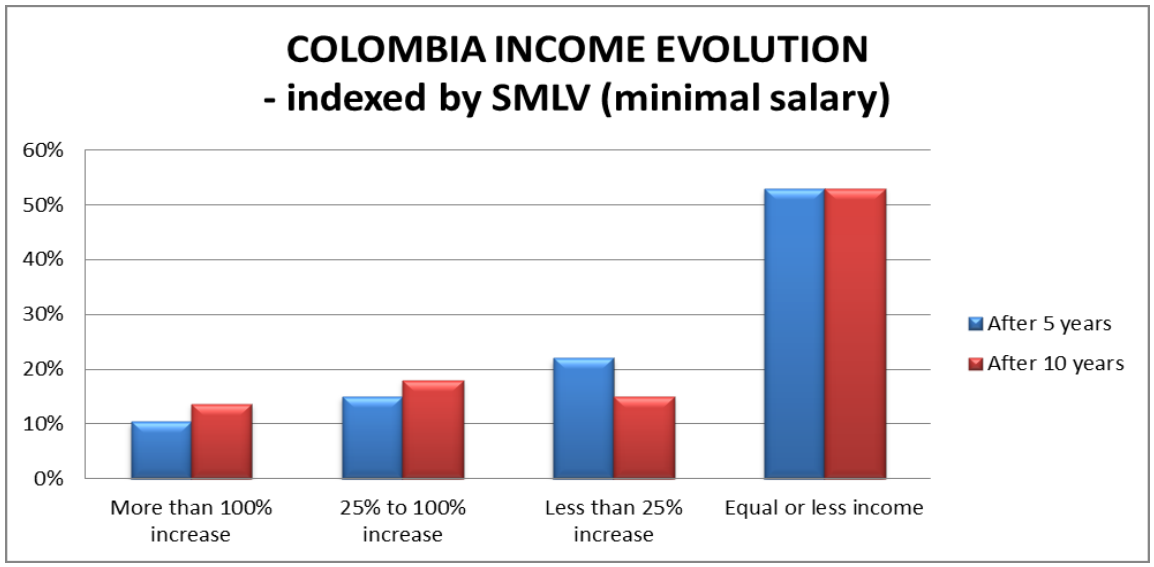


***STATISTICAL ANALYSIS OF DATA (CORRELATION ANALYSIS)***

For correlation analysis, is kept in mind the customers IBC growth in this period of time, from there, percentage of variation groups are defined.

The calculation of the average growth rate is equal to:

$$Growth\% = \frac{(future\ income - Inicial\ income)}{Initial\ income}$$



***SELECTED VARIABLES***

The development of statistical research included a thorough study of the variables for which information was found and its correlation with income information.

| AFFILIATE INFORMATION SOCIODEMOGRAPHIC | EMPLOYER INFORMATION | EMPLOYER-CLIENT RELATIONSHIP | ORIGIN AFFILIATE | RELATIONSHIP WITH AFP |
|---|---|---|---|---|
| Age | Company Type | Seniority | Regional | Relationship PQRs |
| City of residence | Size | | Type of member | Seniority with AFP |
| Initial income | Company Activity | | Sales channel | Internal segment AFP |
| Presence of beneficiaries | Economic Sector | | City sale | Fund value |
| Gender | | | Origin Affiliate | Months with collection |
| | | | | Average days quoted |
| | | | | Values historic collection |

The predicted income is compared to the real income, in order to check the predictive power of the model. 4 IBC ranges were generated as follows:

$$y = \begin{cases} Very\ low\ income \\ Low\ income \\ Middle\ income \\ High\ income \end{cases}$$

Results

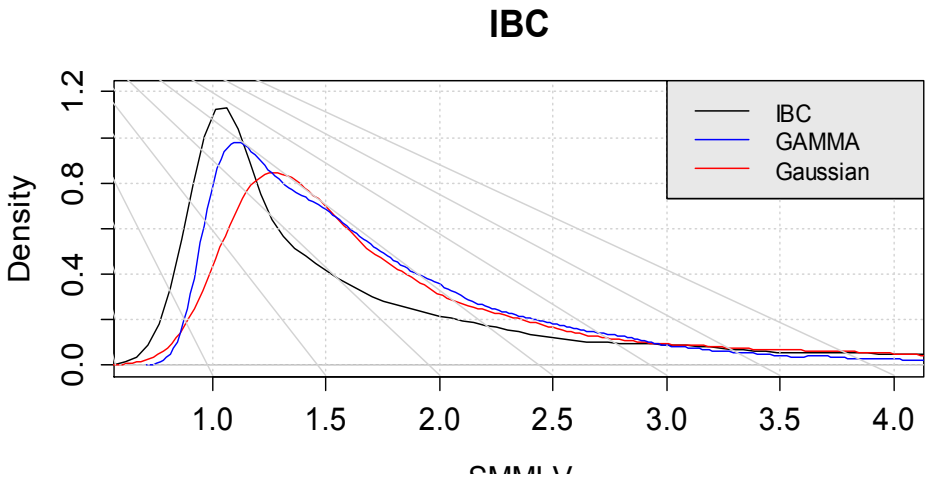
***REAL VS. FORECAST***

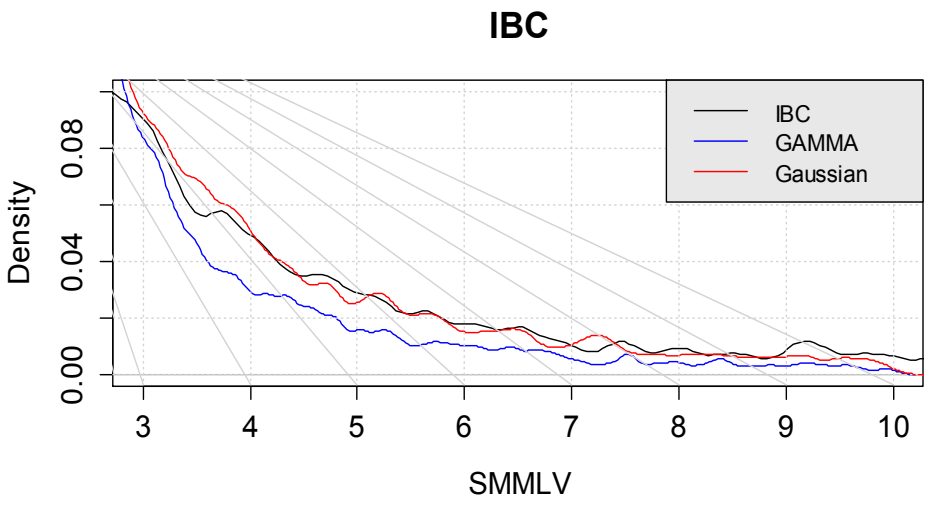
| | | Forecast | | | | |
|---|---|---|---|---|---|---|
| | | Very low income | Low income | Middle income | High income | Total |
| Real | Very low income | 61% | 7% | 0% | 0% | 68% |
| | Low income | 7% | 16% | 0% | 0% | 23% |
| | Middle income | 0% | 3% | 2% | 0% | 6% |
| | High income | 0% | 0% | 0% | 2% | 2% |

| | |
|---|---|
| Efectivity | 80,3% |
| 1 range difference | 18,6% |
| 2 range difference | 1,1% |

The predictive power of the model maintained by keeping the ranges established an accuracy level of 80.3% calculated on the sum of the diagonal (Customers whose forecast model matches the real income) of the total population.

***MODEL COMPARISON***



# STRATEGIES AFTER FORECASTS

Considering the score bands obtained for the attrition and income models, have raised the following groups, which indicate the risk of desertion of affiliates according to their classification and corresponding strategies for each level of risk.

| | 24 months desertion probability | | | |
|---|---|---|---|---|
| 24 months Forecasted income | High risk | Middle risk | Low risk | Total % |
| High income | 32% | 18% | 15% | 1% |
| Middle income | 27% | 12% | 6% | 2% |
| Low income | 17% | 6% | 2% | 49% |
| Very low income with option of no income | 26% | 5% | 1% | 21% |
| Sub total | 21% | 6% | 2% | 72% |
| Low fund value | | | | 28% |

| Range | Strategy |
|---|---|
| | * Proactive Retention<br>* Satisfaction Commercial management - Cross-selling<br>* High priority in requests and complaints |
| | * Middle priority in requests and complaints<br>* Commercial management of proactive nformation - cross selling |
| | * Satisfaction Commercial management - Cross-selling<br>* Middle priority in requests and complaints |
| | Electronic brochure |
| | * Satisfaction Commercial management - Cross-selling |
| | No management priority |
| | Client Recovery |

# CONCLUSION

The developed Scoring and income forecasting models have allowed pension funds to use statistics to organize marketing decisions, business and customer service strategies. Also, to include the option of new products to link in the long term customers who are more likely to drop out, products focused on income level forecasted to each customer

Additionally the institution was motivated to maintain a database which allows them to continue with this type of statistical analysis for decision-making. To the extent that if they have a more historical reliable database it could be increased the forecast horizon to more than two years.

# THANKS

gsalamanca@lisim.com