Innovative Options Data Signal for Trading with Technical Indicators and VIX optimized by a GA

José Bernardo Belo
Mestrado em Engenharia Electrotécnica e de Computadores
Instituto Superior Técnico
Lisboa, Portugal
Email: bbelo95@gmail.com

June 2019

Abstract

In this thesis it is proposed the use of options data, organized in an innovative form, in a mechanism made by an investment simulator led by technical and VIX indicators using trailing stoplosses and optimized by a genetic algorithm with the objective of discover a rule that can be applied in the options market, then some tests were made in the SP500 Index options market for results comparing and analyzing. Daily options data from 2011 to 2017 on the SP500 index were used for train with a rolling window and a return of investment of 325% was achieved in test in a time period of 1200 days.

Keywords: Data Processing, Genetic Algorithm, Options Market, Technical Indicators, Trailing Stoplosses, VIX

I. Introduction

In the era of data in which we live in, many are the practical uses that can be taken from big data sets. The analysis of data is nowadays an area that gathers many academics and companies because, from data, a lot of conclusions can be taken to improve systems and services. In the financial markets, a lot of money is traded every day and the possibility of merging data analysis with profiting is fascinating since data is nowadays easier to get and share. However more than use big data sets it is very important the way data is organized and settled for the models of data analysis. It is very important to have financial knowledge. Knowing when and how to invest is an art that requires many years of experience and knowledge but it can be for everyone. Whether it's an engineer, a doctor or a postman, knowing how to invest is a quality that should be studied by everyone and, since every worker gets paid in the country we are, knowing how to reinvest can make the incomes of everyone much greater. If all the Portuguese population knew how to invest and succeed on it, certainly the country would get richer and get better conditions for all.

The main objective of this thesis is to develop a model that creates a rule on the options market, which regularly settles positive transactions and gives returns in the future.

For the objective proposed it will be used data from the Chicago Board Options Exchange from 2011 to 2017 of the option contracts and the volatility index. The model will be trained and tested for results analysis with a rolling window that will roll on the time period described.

This document is organized in five main chapters described below:

- Introduction - in this chapter the works proposed is explained along with the document structure.
- Related Work - in this chapter, it is made a view on the main themes that this work approaches and then some of the research made is scrutinized and resumed.
- Architecture - in this chapter, it is projected the architecture for the model developed and all the characteristics that were built.
- Evaluation - along this chapter are exposed the study cases developed for the architecture testing, while a set of comments is made on the results achieved.
- Conclusion - In the last chapter it is made a conclusion on the work and improvements that might be done in future works are suggested.

'The financial markets are a zero sum game where every dollar won by one investor is lost by another. Knowledge and trading tools are the differentiating factors that determine whether an investor lands on the winning or losing side' [1]. My investment experience already teach me that investing in not an easy task. In such a dynamic financial world where, nowadays everyone is connected and investing was made easier with the internet and globalization many are the variables which explain the oscillations of the markets and for achieving the success, making some good profit, it is needed not only experience but also resilience and patient. Along this work I had the opportunity to have weekly meetings with professor Rui Neves and many other students that also share with me the passion of investing. These meetings gave to me even more will to understand market behaviours and use my knowledge in programming and machine learning to analyze data and discover automatized rules that might be profitable in

the future. This work, I hope, is the beginning of a relationship between me and market analysis that will certainly stay for the future with the dream of achieving great profitable investments along my life.

II. Related Work

A. Options Market

For this subsection a synthesis is made on the definitions of call and put options along with the Greeks. An option is called a 'Call' when the holder as the right, but not the obligation, to buy an underlying asset from the writer at a specified price and time period and it is called a 'Put' when the holder as the right, but not the obligation, to sell an underlying asset to the writer at the specified price and time period [2], [3], [4]. On one hand, a Call is a bet on the underlying asset price increase and the holder can make profit by buying an asset cheaper than the real price since the contract is executable at the strike price. If the asset price doesn't get lower than the strike price during the contract the holder will only lose the premium price payed for the signing. Therefore a Call option has a limited possible lose which is the premium price and an infinite possible profit for the holder since assets can theoretically grow infinitely. On the other hand, a Put is a bet on the underlying asset price decrease and the holder can make profit by selling an asset more expensive than the real price since the contract is executable at the strike price. A Put option can be seen as an insurance. If the asset price doesn't go below the strike price during the contract the holder won't execute the contract and the premium price will be lost. Therefore a Put option has a limited loss which is the premium price and a finite possible profit since asset prices are limited at zero.

Besides executing an option contract during the contract period an investor can also sell an already bought contract in the middle of its lifetime by a bigger premium price than the one paid to the writer because of the variations on the assets price. In its lifetime an option can be in-the-money or out-of-the-money [4]. An option is in-the-money if, in a moment, the strike price is smaller than the asset price in case of a call and if the strike price is greater than the asset price in case of a put. So on, and summarizing an option is in or out of the money depending if its intrinsic value is, respectively, positive or negative.

Below it is the calculation formula of the intrinsic value of a call option:

$$IntrinsicValue(Call) = UAP - Strike \tag{1}$$

Along it is the calculation formula of the intrinsic value of a put option:

$$IntrinsicValue(Put) = Strike - UAP \tag{2}$$

The Greeks are simply sensitivities to options risk characteristics [3]. Each one of the Greeks variables depend on different options parameters and in figure 1 it is displayed each Greek and which parameter sensitivity reflects:

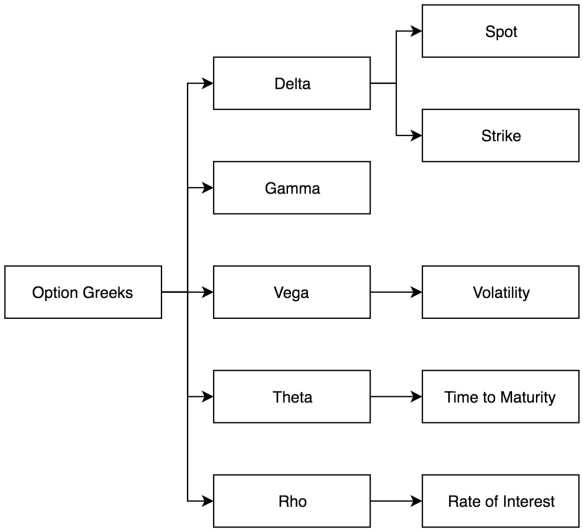

Fig. 1. The Greeks Map

- The Delta parameter is the degree of change an option premium price will move based on a 1 dollar change on the UAP. Delta values fluctuate between 0 and 1 for calls and between -1 and 0 for puts and, for example, if an option has a delta of 1 it means that its premium price will oscillate in the same rate of change as the UAP.
- The Gamma parameter is the rate of change the premium price will move based on the next increments of 1 dollar after the first one in the UAP. Gamma is the 'acceleration' of Delta through out successive increments on the UAP. Option contracts with highest Gammas are more responsive to changes in the UAP.
- The Theta or time decay is always negative and is the enemy number one of an investor because it is the amount of money the premium price will decrease every single day. Since options are wasting assets their premium price decays over the time and it is expected that options with more time to expiration have less time decay than others with less days to expiration.

Along many years, many researchers have empirically investigated the relation between options and equity markets. Anthony (1988) [6] uses daily closing prices to conclude that options lead stocks. Later Stephan and Whaley (1990) [7] conclude the exactly opposite relation finding that stocks lead option prices. This discussion is also made by Manaster and Rendleman (1982) [8], Bhattacharya (1987) [9], Vijh (1988, 1990) [10], Conrad (1989) [11], DeTemple and Jorion (1990) [12], Damodoran and Lim (1991) [13], Sheikh and Ronn (1994) [14], Kumar, Sarin and Shastri (1995) [15] and however the evidence on market interrelationships is inconclusive as to which of the two markets reflects new information earlier.

B. Technical Analysis

In finance, technical analysis is a security analysis discipline for forecasting the direction of prices through the study of past market data, primarily price and volume [16]. In 2003, David Enke and C.H.Dagli exposed in their work [17], three types of neural networks joined with technical indicators for predicting future stocks behavior. In their publication indicators like Relative Strength Index, Money Flow Index, Moving Average, Stochastic Oscillator and Moving Average Convergence Divergence are used and using as evaluation metrics the return and the risk, it is settled that *The overall results indicate that the proportion of correct predictions and the probability of stock trading guided by these neural networks are higher than those guided by their benchmarks* (traditional technical indicators strategies).

C. Volatility and VIX

Volatility refers to the amount of uncertainty related to the size of changes in an asset value. A higher volatility means that an asset value is spread out over a larger range of values. This means that the price of the asset has changed more over a time period in either direction. A lower volatility means that an asset value does not fluctuate so much and so it is steadier. Volatility on the assets directly impact the options premium prices because options on assets with higher volatility are much more risky to the writers than assets with lower volatility.

The VIX® Index was developed by CBOE [5] and is a *'benchmark index to measure the market's expectation of future volatility. The VIX Index is based on options of the SP500® Index, considered the leading indicator of the broad U.S. stock market'*.

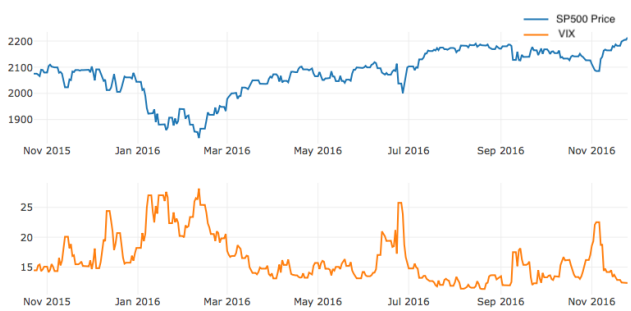

Fig. 2. VIX Indicator with SP500 Stock Price

D. Genetic Algorithm

It is needed to go back to the 70's where for the first time John Holland [19], Professor of electrical engineering and computer science at the University of Michigan, developed and built this new concept that merges Darwin's ideas [18] into process optimization. In his book [19], John Holland describes adaptation, which retracts the biological response whereby organisms evolve by rearranging genetic material to survive in environments confronting them, and overpasses this analogy to a practical mechanism. Later on, David A Coley [21], in his book, defines the Genetic Algorithms as *'...numerical optimization algorithms inspired by both natural selection and natural genetics'* recalling to Darwin's Theory [18] and, as Melanie Mitchel [20], describes GAs in four main blocks:

1) A population of members, each one with a solution to the problem;
2) A way of calculating how good or bad each member of a population is and a selection method to develop future populations;
3) A method for mixing fragments of the selected members inside a population to form a new and better one;
4) A mutation operator that gives diversity in members inside new populations.

In 2008, Rohit Choudhry and Kumkum Garg [23], developed a model in which a genetic algorithm and a support vector machine Algorithm together in a hybrid model outperformed a simpler SVM model. In the work Rohit and Kumkum use technical indicators (PVT, Stochastic, Disparity, ROC, Williams and Momentum) of not only the stock price in case study but also the correlated stock price companies as features for the algorithms and show that the GA-SVM algorithm outperforms the SVM stand alone in three different stocks, TCS with a hit ratio of 61.732% against 58.09%, Infosys with 60.285% against 56.748% and in Reliance with 59.534% agaisnt 55.643%. Some years later, in 2011, Gorgulho [22] developed a model merging a GA with a technical indicator analysis. In his work, with the use of indicators as EMA, HMA, DC, ROC, RSI, OBV, TSI and MACD and a 2003-2009 data set it is settled that this model *results are promising since the approach clearly beats the remaining approaches during the recent market crash* and a ROI of 62.95% was achieved by the best GA.

In figure 3 it is shown the flowchart on a genetic algorithm as it was previously described with the Selection, Crossover and Mutation stages all together in the Next Population creation phase.

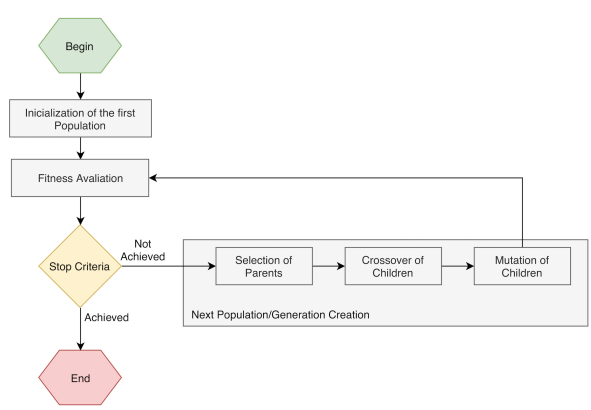

Fig. 3. Genetic Algorithm Flowchart

To conclude, all this themes were used along this project.

III. Architecture

A. Overall Structure

Along this work it is used an architecture based on four main blocks, the Data Layer, the Indicators Layer, the Investment Simulator Layer and the Optimization Layer. Each layer has its own mechanisms and together, the four layers, form an optimized trading system made to invest inside the options market. Below are presented the main layers with a small introduction to its characteristics and role inside the full process.

1) The Data Layer is the beginning of the full process and organizes the raw options data to be presented to the next layers. This process is personalized to options data and explored deeply in the Data Layer Section.
2) The Indicators Layer was based on Gorgulho's [22] system that gathers technical indicators and a grading mechanism to evaluate (daily, weekly or monthly) the best moments to long or short the stock markets. This system was used, adapted for options markets, using options market data from the Data Layer instead of the stock markets data as Gorgulho did and improved with new features that are reviewed below.
3) The Investment Simulator Layer receives the grades from the Indicators Layer and the organized data from the Data Layer and simulates investments on the options market. Both the Investment Simulator Layer and the Indicators Layer will be part of a process that will be optimized by the last layer.
4) The Optimization layer is where the Genetic Algorithm will be applied. The genetic algorithm will try to maximize the results of our process experiencing and evaluating lots of parameters combinations that are inputs in both the Indicators Layer and Investment Simulator Layer. This optimization pretends to find a rule which is applied inside the options market and profitable for the future.

In the figure (4) it is possible to see the fluxogram that describes the proposed architecture. Following this quick introduction, each layer is explored and detailed.

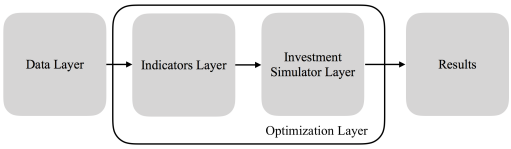

Fig. 4. System Structure

B. Data Layer

The Data Base is charged with options daily data from .csv files, gave by the Supervisor, from 2011 to 2017. Each entry represents an option contract and has 13 fields: Company Symbol, Underlying Asset Price, Exchange, Option Root, Option Extension, Type, End Date, Begin Date, Strike Price, Last Price, Bid Price, Ask Price and Volume. All this parameters were explored and presented during the Background section and will be used along the project.

An option plot is made of the values that a determined contract has along its lifetime and begins on the contract signing moment until its end. This plot is achieved by searching and gathering the values of consecutive options contracts, that only differ from each other on the beginning date.

The strategy used for the called window plot is described with real examples in figures 5, 6 and 7 correspondingly (A), (B) and (C) and as it is possible to see the plot is the result of many temporal windows put together. This strategy begins with the cropping of the full time period in windows with the same size defined as window length size. Each window corresponds to a time interval (for example the first window of the year 2011 with twenty days window length is from the 1st of January to the 21st of January). After this first step, it will be chosen, for each window, from the data base, the option contract that best fits inside it. For this selection some options parameters, that will reproduce different plots, will be used further ahead and explained below in the inputs list. When the selection is complete, all the chosen contracts will be represented in the respective window and then united in two different ways:

1) The first plot, is the union of each window with its option inside normalized to the last one which means that one by one, starting from the first to the last, windows are added to combine the final plot and each new window added is normalized to the last point of the last window assembled changing its scale. To conclude, when all the windows are joined and normalized, the last value of the final graphic is set to 10 changing also proportionally all the plot. As its seen, options value decay over time and so this graphic, as being the union of many option values along time is tendentiously decreasing. This plot will be presented to the investment layer and it will give an idea of the market tendencies rather than the absolute values of the contracts which mean that if the plot is rising or decreasing the value of the underlying asset price is also moving in the same way for a call price or in reverse way for a put price. With different inputs this plot will be represented with different options and so on with bigger or smaller oscillations due to options characteristics.
2) The second plot, is only the union of each window with no normalization and will be used for the investment simulator layer to open and close orders of option contracts by their real values.

Now, that the options for the plots (A) and (B) are selected and built, a third plot will be done. New options that only differ from the previously selected by a strike price difference will be searched and plotted as (B) was plotted (without any normalization). This plot can be built with different strike prices differences and it is usually parallel to the (B) plot because their contract values differences are proportional to

their strike prices differences. This plot was called a strike band (C), and will also be used in the investment layer for trading. For example in figure 7, to form the green line, a search is made for the options that differ in the strike price, from those selected for the windows of the graphs A and B (yellow line), from 1 to 30 and then the ones with the highest volume are selected and plotted.

This data organizing system will be used for different tests since it is possible to represent the price of puts or calls values along time with contracts of different conditions. So on, some inputs must be given to the data layer for plotting and are explained above along with figure 8 where the mechanism for choosing an option to its window is schematized.

1) Time Period
2) Window Size
3) Options Specifications:
   a) Options Company
   b) Options Type - Call or Put
   c) Options Contracts Duration
   d) Option Contracts Intrinsic Value
   e) Maturity Minimum Time Remaining
   f) Strike Bands Differences
   g) Volume Maximization

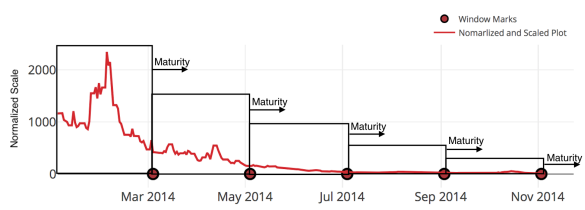

Fig. 5. Windows with the chosen, cropped and normalized put options (A)

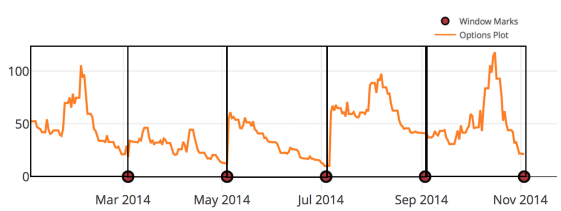

Fig. 6. Windows with the chosen and cropped put options (B)

C. Indicators Layer

The Technical Analysis is one of the main tools that investors use to study and understand better timings for market entering and as mentioned previously, technicians believe that markets price reflects all relevant information for future investments. Gorgulho [22] used the RSI, the MACD and the ROC with positive results and so those will be adapted for our problem and will be set together with another improvement indicator, the bollinger bands. Each of the below described technical indicators will require some input values for their

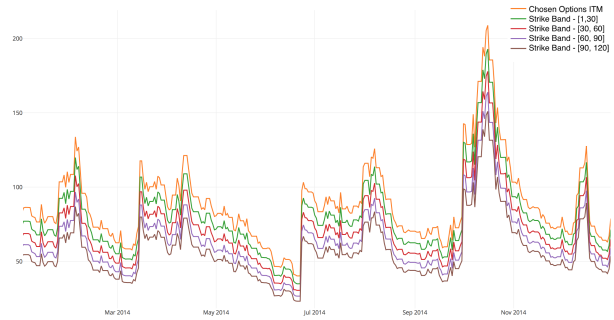

Fig. 7. Strike Bands (C)

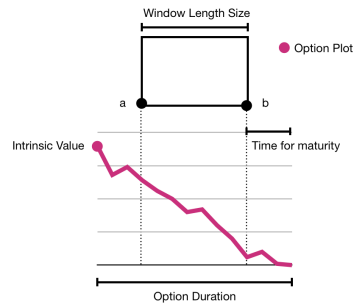

Fig. 8. Selection of an option to a window

construction and will sweep the normalized window plot giving a mark at each and every moment. During this inspection, each indicator will grade these data giving:

- High Grade: 1 point for a possible short situation (asset is going to decrease)
- Low Grade: -1 point for a possible long situation (asset is going to increase)
- 0 c.c.

Following each technical indicator is explored and reviewed.

1) Relative Strength Index: The Relative Strength Index is a momentum indicator which measures the level of recent price changes to evaluate overbought or oversold conditions in an asset. The RSI will need three parameters, a period, a maximum and a minimum thresholds to provide the evaluation.

2) Rate of Change: The Price Rate of Change is a momentum indicator that measures the percentage change in price between the current price and the price n periods ago. The ROC indicator is plotted centered in zero, with the indicator moving upwards into positive values if prices are increasing, and moving into negative values if prices stop increasing and begin to fall. Gorgulho [22] uses a point and its previous to see if a zero interception occurs in the indicator and so on grade it with a good or bad mark. In the architecture designed in this thesis a derivative parameter is also implemented so that, to see if a zero interception was made in the actual point, a comparison to its value in another moment, some points back, is made. This derivative parameter can be one and fully adopt Gorgulho's work strategy or it can be another value and so, more possible different analysis are contemplated and will

be analyzed. To conclude, two parameters are needed for the ROC, a period and a derivative parameter.

3) Bollinger Bands: The Bollinger Bands are a momentum indicator defined by a set of lines plotted some standard deviations away from a simple moving average. One of the lines is set positively to the SMA and the other negatively. The standard deviation is proportional to volatility so when the markets become more volatile the bands widen and in reverse, they contract. This indicator will need, as parameters, a time period and two constants to manipulate the upper and lower bands distance to the SMA or middleband.

4) Moving Average Convergence Divergence: The Moving Average Convergence Divergence is a trend-following momentum indicator that shows the relationship between two moving averages. The MACD is usually calculated by subtracting a slow-period Exponential Moving Average from a fast-period EMA. The result of that calculation is called the MACD line. Then an EMA of the MACD, called the signal line, is plotted on top of the MACD line, which works as a trigger for buy and sell moments. Traders may buy the stock asset when the MACD crosses above its signal line and sell when the MACD crosses below the signal line. To conclude, the MACD indicator will need a fast and a slow period for the EMA's, a period for the signal line and a derivative parameter to be used as improvement like in ROC to see if a cross between the MACD signal and the signal line occurs.

5) VIX Indicator: The VIX Index is explained above and it will be used in the architecture of the system because it can be useful for detecting moments where the market is highly volatile and so dividing moments for go short or go long. Since options contracts are used, a fall in the stock market can result in great option price oscillations and so on it is very important to avoid high volatile moments because if an order goes wrong losses can be deadly for an investor. The VIX is used with bollinger bands and when the VIX Index value is above the upper band it triggers the investment layer only to short avoiding big stock crashes, by closing long positions, and trying to profit from an asset decrease.

D. Investment Simulator Layer

The Investment Simulator Layer receives the signal to trade (B) and the correspondent marks gave by the Indicators Layer. In this layer, all the four technical indicators individual grades will be weighted, summed and applied in an investment simulation that trades the trade signal which has the price of option contracts along time. The following rule is applied in this process as Gorgulho did:

$$\sum w_i = 1, \qquad i = 1, ..., 4 \tag{3}$$

For this Layer a fund of 1M is ready for a simulations and the total risk can only be more than 1M if the past transactions devolved profit increasing the money the fund has to invest. Both long and short positions are positions that provide income, the first is by the increasing of the asset value and the second by the decreasing. Many were the studies seen with long and short positions strategy but the problem is that this types of orders alone cannot be risk measured, so as we use options data along this thesis, instead of using long and short positioning bear and bull option spreads are used. Option spreads can be risk measured unlike an unique option order, since they have a maximum win and a maximum lost possible calculable even before orders opening. Below bear and bull spreads are described.

1) Bear and Bull Spreads: A Bull Spread consists in buying and selling, at the same time, two option contracts in the same asset that only have different strikes and this position gets an income through the UAP value increasing. On the contrary, a Bear Spread consists in buying and selling, at the same time, two options contracts in the same asset that have only different strikes and this position gets an income through the UAP value decreasing. Spreads will be used in the Investment Simulator Layer.

2) Orders: It is important in this moment to define how an order is processed and it is opened. Since we are using option spreads, this mechanism will look for moments with a grade under a minimum threshold to do a bull spread in the market and above a maximum threshold to make a bear spread in the market. This grade mechanism will be helped by the VIX which will say when volatility is too high and so a crash in market is predictable closing bull positions and making only bear spreads until volatility return to lower values. Since a window plot will be used, at the end of each window all the orders will be closed because another contract is ready for trading. Besides this way and the VIX there is only one case in which an order can be closed, a stoploss. The orders risk is defined as input for this layer. All this parameters shall be given to the investment layer and different parameters will generate different results.

3) Trailing Stop Losses: In Adam Y.C. Lei and Huihua Lib [24] studies trailing stop losses showed very good results and so on they will be used in this system. Trailing stop losses are adaptive stop losses which grow if the income from a determined position increases to a new maximum. If an order keeps increasing its profit, the stoploss will also be increasing along time avoiding to lose the profit if a crash happens. If the income from a position falls, there wont be a stop loss update and if it matches the stop loss the position is closed, avoiding more risks.

E. Optimization Layer

The use of the GA will be made with some improvements studied along the research and are described below after the Chromosome defining.

1) Chromosome: For the GA, it is needed to define the structure of the chromosome, which means, it is necessary to define the variables that the GA will give as inputs to the process. The process developed is on the indicators and investment simulator layers so on, for the first layer, it is needed the inputs for the technical and economic indicators, for example, the period of the RSI along the normalized window plot or the period for the VIX bollinger bands. For

the second layer, it is needed inputs to the thresholds that define the market entrance moments (mark1 and mark2), the weights which define the relative importance of each technical indicator grade (which the sum is 1), the net risk value of each order the stoploss measure and the strike band used for the spreads. The chromosome defined then, has a length of 24 elements and each of those elements can oscillate its value inside a predefined value range. In figure 9 the full chromosome of the GA is represented and in table I the range in which the chromosome values will be generated.

Weight 1	Weight 2	Weight 3	Weight4
RSI period	RSI Max Level	RSI Min Level	ROC period
ROC derivative	BB period	BB standard deviation 1	BB standard deviation 2
BB period VIX	BB standard deviation VIX 1	BB standard deviation VIX 2	MACD period 1
MACD period 2	MACD period 3	MACD derivative	STOPLOSS
ORDER RISK VALUE	MARK1	MARK2	STRIKE BAND

Fig. 9. Chromosome

Parameters	Range	Parameters	Range
Weight 1	[1, 25]	BB VIX period	[15, 50]
Weight 2	[1, 25]	BB VIX std1	[0.5, 2.5]
Weight 3	[1, 25]	BB VIX std2	[0.5, 2.5]
Weight 4	[1, 25]	MACD period1	[16, 30]
RSI period	[5, 30]	MACD period2	[5, 15]
RSI max threshold	[65, 90]	MACD period3	[2, 8]
RSI min threshold	[10, 35]	MACD derivative	[1, 4]
ROC period	[5, 30]	Strike Band	[1, 4]
ROC derivative	[1, 4]	Stoploss	[10, 50]
BB period	[5, 30]	Order Value	[1, 150000]
BB std1	[0.5, 2.5]	MARK 1	[0, 1]
BB std2	[0.5, 2.5]	MARK2	[-1, 0]

TABLE I
GA CHROMOSOME VALUES RANGE

2) Hall of Fame: The Hall of Fame mechanism is an improvement in the algorithm and it is popped when a determined number of generations occur without the appearance of a new maximum. The Hall of Fame basically saves the best solutions found along the algorithm and when called, it provides the parents for the new offspring avoiding step backs and losses of memory in the system. This mechanism permits a bigger hypermutation and random immigrants ratio because the best solutions will always be saved and will never be lost.

3) Hypermutation: The Hypermutation mechanism *adaptively introduces diversity when needed* [25] and it will be used inside the improved GA. Along the generations if a better solution doesn't get popped the mutation rate will increase until it reaches a threshold. If the threshold is reached the Hall of Fame mechanism is popped and the initial mutation ratio is again applied. This mechanism will introduce more possibilities for the algorithm to find new solutions and it prevents that the mechanism doesn't get stuck in a local maximum as it makes more random values get in the process.

4) Random Immigrants: Along with Hypermutation, this mechanism introduces new random chromosomes during the optimization making the algorithm more capable of discover new optimum solutions avoiding local minimums and reproducing new unseen solutions for the process.

IV. Evaluation

A. Data Split

For the case studies, a rolling window mechanism will be used, which means that options data from 2011 to the end of 2017 will be divided in three different sets, each with 75% train data and 25% test data. Since the data used is the plot made of the union of windowed options, the rolling window measure used will be a multiple of the options window length so that each train and test sets end in a contract option window end. Since the option windows length will be 100 days the rolling window measure will be 1200 days for the train data and 400 days for the test data. In figure 10 it is possible to see how data was divided with a representation of the scheme for the rolling window used for train and test data.

Along the data a media between the ask and the bid price will be used since it represents the median point between the minimum value a seller is willing to sell a contract and the maximum value a buyer is willing to buy a contract.

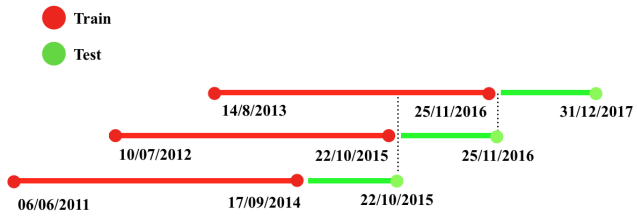

Fig. 10. Data Split Scheme for Rolling Window

B. Used Strategies

In this thesis, two strategies were developed for results comparison and for the first two study cases both strategies will be done five times for accuracy measure. Below are described the two optimization processes used:

1) Standard Genetic Algorithm - based on the background description with fixed mutation and elitism selection.
2) Improved Genetic Algorithm 1 - Based on the background description plus the State-of-the-Art improvements described:
   - Random Immigrants with 10% rate.
   - Hypermutation which means an oscillating mutation rate with increment steps of 7% capped at 28%.
3) Improved Genetic Algorithm 2 - Based on the background description plus the State-of-the-Art improvements described:
   - Random Immigrants with 10% rate.

- Hypermutation which means an oscillating mutation rate with increment steps of 7% capped at 28%.
- Hall-of-Fame implementation with the best results along the generations.

C. Evaluation Metrics

Along the research made for this work it was noticed that many academics used the Return on Investment as a main metric for results comparison and so on it will be used. The ROI metric is a percentage that indicates the gain or loss generated on an investment relative to the amount of money invested and it is calculated as in (4).

$$ROI = \frac{Gain - Cost}{Cost} \tag{4}$$

Along with ROI, listed below, are also other evaluation metrics used for study cases analysis:

1) Number of Spreads - It indicates the number of spreads made along the time period. The number of spreads opened can be a measure of each spread individual risk since the approach designed conceives an initial fund with limited money.
2) Percentage of Successful Spreads - It indicates the percentage of successful spreads and it is important to measure the accuracy of the used model since, even with positive results, a test might have a low successful order percentage.
3) Percentage of Unsuccessful Spreads - It indicates the percentage of unsuccessful spreads and, as the last feature, it can show if the model is steady or unsteady.
4) Average Spread Profit - It indicates how well are spreads being made. This feature is the profit made divided by the numbers of order and the highest, the better the results are.
5) Biggest Spread Profit - It indicates the best spread made on a specific test. This can be a good feature to compare in which conditions different tests get good returns.
6) Biggest Spread Lost - It indicates the worst spread made on a specific test. This can be a good feature to see how and when the mechanism fails an order.
7) Biggest Draw Down - It indicates the maximum fall that happens along the investment in the portfolio made of the money in the fund and the actives owned.
8) Number of Stoplosses - It indicates the number of positions closed by a stoploss

V. Case Studies

Along this thesis it was created a model to plot option contracts values along time by selecting options for certain parameters in the data layer. This implementation was done and for both, the first and the second case studies, it was developed different put options alignments. For the last study case it was used a different mechanism for the trailing stoploss in the investment simulator layer. In all the case studies usually OTM put options were used, even for the strike bands of the options chosen. So on, it is important to define how put contract time to maturity reacts along the lifetime of a put option.

1) Delta and Time to Maturity - the delta of an OTM put option approaches 0 as the option approaches expiration and for ITM put options approach -1.
2) Time Decay and Time to Maturity - The further a put option is from the expiration, the smaller is the time decay.

In the first case study, the options window plot will be made with options closer to maturity compared to the options used for the second case study which are further to maturity. With this maturity difference, the chosen options for the first study case will overall be less exposed to delta and more exposed to time decay than those used on the second study case and this gap on maturity will influence all the optimization process. For the third case study, options used for the first case study will be reused but now with a trailing stoploss on the option prices rather than on the stock prices. For the first two study cases an absolute trailing stoploss was used on the stock price and for the third a percentage trailing stoploss is used closing positions when a loss is seen in a value range of 10% to 20%. All this three case studies will use the same strike band margins used for the plot 7 and will result in some conclusions when compared. Along this paper it will only be shown the first and the second case studies. The third result was the one with the biggest ROI but it was more interesting for this paper to watch the cases which try to discover which option contracts are better for the model built. I extremely recommend reading the full work done.

A. Study Case 1 - Options closer to maturity

For the first case study, the Data Layer is set with the parameters on table II. In table III are exposed the evaluation metrics for the best results represented in figure 11.

Parameters	Values
Window Size	100
Strike Difference Minimum	-20
Strike Difference Maximum	0
Duration Contracts Minimum	0
Duration Contracts Maximum	300
Maturity	100

TABLE II
Data Layer First Case Study Parameters

Evaluation Metrics	Improved GA	Standard GA
Number of Spreads	277	267
Percentage of Successful Spreads	53.3	52.4
Percentage of Unsuccessful Spreads	46.7	47.6
Average Spread Profit	0.98	0.92
Biggest Spread Profit	20.4	20.9
Biggest Spread Lost	-3.9	-6.2
Biggest Drawdown	-31.6	-28
Number of Stoplosses	142	124
ROI	263.69	245.4

TABLE III
Evaluation Metrics for the First Study Case

From table III and figure 11, it is possible to conclude that the Improved GAs outperform the Standard GA, achieving a bigger average ROI and a percentage of successful spreads slightly bigger for the best results achieved which can be a signal that improvements made like Random Immigrants, Hypermutation and the use of an Hall of Fame can be good features to set in a GA for a better performance. The best solution found was achieved with the Improved GA2.

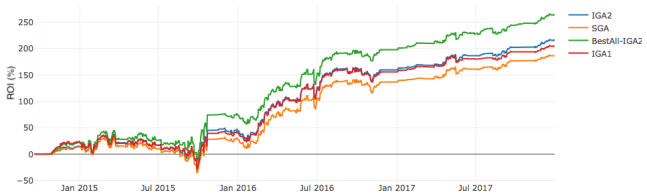

Fig. 11. ROI for the two Strategies

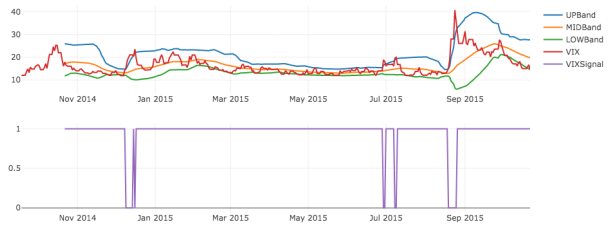

Fig. 12. Bollinger Bands on VIX for the first case study

It also is seen that most of the positions opened are bull spreads since drawdowns in the asset are very quick and only cut by the VIX indicator by a small time period which takes the opportunity of the model to take profit from bear moments but consecutively profit from bull spreads. Looking to figure 12, it is concluded that many asset drawdowns, in 2015, were avoided but in the last represented, on September of 2015, the VIX is to slow to avoid a big drawdown that is observable in the ROI chart.

B. Study Case 2 - Options further to maturity

For the second case study it is used options further to maturity and the data layer inputs are on table IV. Again, the best solution found was achieved with the Improved GA2 and the Improved GAs outperform the Standard one. In this case options have, after each window, more than 200 days to maturity. This study case will be used for comparison to the first one seen before and as first comment it is imaginable that, since options further to maturity are less exposed to time decay, the profits for this case shall be smaller. What should be really seen is the percentage of successful spreads done by each strategy since even that the options traded are different, when a fall on the underlying asset happens, all the put contracts will increase their values and in reverse, decrease.

Parameters	Values
Window Size	100
Strike Difference Minimum	-20
Strike Difference Maximum	0
Duration Contracts Minimum	0
Duration Contracts Maximum	500
Maturity	200

TABLE IV
DATA LAYER CASE STUDY PARAMETERS

In figure 13, it is possible to see the ROI along time achieved on the best results and the evaluation metrics are shown in the table V for these case.

Evaluation Metrics	Improved GA	Standard GA
Number of Spreads	271	290
Percentage of Successful Spread	0.56	0.53
Percentage of Unsuccessful Spreads	0.44	0.47
Average Spread Profit	0.79	0.6
Biggest Spread Profit	9.6	9.4
Biggest Spread Lost	-4.85	-4.81
Biggest Drawdown	-14	-15.7
Number of Stoplosses	132	124
ROI	213	173.9

TABLE V
EVALUATION METRICS FOR STUDY CASE

The ROI and the average spread profit of the first study case are bigger than on the second case because options traded in the first case suffer more the effect of time decay and so for bull positions the spread is decreasing and the overall profit is bigger but the percentage of successful spreads is a little bit bigger for the second case study which might indicate that options used for the second study case, with more time to maturity, can make the model create a better rule for the proposed objective.

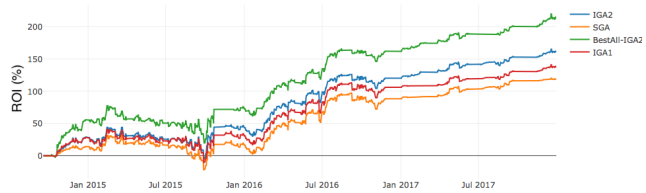

Fig. 13. ROI for the two Strategies

VI. CONCLUSION

In this project it was designed an optimized model to trade in the options market by creating a rule with train data that revealed positive results on the test data. The model is made of an investment simulator layer where technical indicators, VIX and trailing stoplosses trigger entering and exiting moments for bull and bear positions. Both the investment simulator layer and the technical indicators layer are optimized by a genetic algorithm that thrives to discover the best input parameters with the objective of discover patterns that happen along data and fulfill the desired objective of finding a rule that shows to be profitable.

REFERENCES

[1] Jeff Augen, *The Option Trader's Workbook*, New Jersey, FT Press, 2009.

[2] John C. Hull, *Options, Futures and Other Derivatives*, Pearson Education Inc, Boston, 2012

[3] Guy Cohen, *Options Made Easy* FT Press, USA, 2013

[4] George Fontanills, *Trading Options for Dummies*, Indianapolis, Wiley Publishing, 2008

[5] http://www.cboe.com/

[6] Anthony, Joseph H., The Interrelation of Stock and Options Market Trading-Volume Data, The Journal of Finance, vol.43, pp. 949-964, 1988

[7] Stephan, Jens A. and Whaley, Robert E., Intraday Price Change and Trading Volume Relations in the Stock and Stock Option Markets, The Journal of Finance, 1990

[8] Manaster, Steven and Rendleman Jr., Richard J., Option Prices as Predictors of Equilibrium Stock Prices, The Journal of Finance, vol.37, pp. 1043-1057, 1982

[9] Bhattacharya, Mihir, Price Changes of Related Securities: The Case of Call Options and Stocks, The Journal of Financial and Quantitative Analysis, Vol. 22, No. 1 (Mar., 1987), pp. 1-15

[10] Vijh, Anand M., Liquidity of the CBOE Equity Options, The Journal of Finance, vol.45, pp. 1157-1179, 1990

[11] Conrad, Jennifer, The Price Effect of Option Introduction, The Journal of Finance, vol. 44, pp. 487-498, 1989

[12] Detemple, Jerome and Jorion, Philippe, Option listing and stock returns: An empirical analysis, Journal of Banking and Finance, vol. 14, pp. 781-801, 1990

[13] Damodoran, Aswath and Lim, Joseph, The effects of option listing on the underlying stocks return processes, The Journal of Banking and Finance, vol. 15, pp. 647-664, 1991

[14] Sheikh, Aamit M. and Ronn, Ehud I., A Characterization of the Daily and Intraday Behavior of Returns on Options, The Journal of Finance, vol. 49, pp. 557-579, 1994

[15] Kumar, Raman and Sarin, Atulya and Shastri, Kuldeep, The impact of index options on the underlying stocks: The evidence from the listing of Nikkei Stock Average options, Pacific-Basin Finance Journal, vol. 3, pp. 303-317, 1995

[16] Kirkpatrick and Dahlquist, Technical Analysis: The Complete Resource for Financial Market Technicians, (Financial Times Press, 2006) page 3

[17] Thawornwong, Suraphan and Enke, David and Dagli, Cihan Neural Networks as a Decision Maker for Stock Trading: A Technical Analysis, Approach Article in International Journal of Smart Engineering System Design, October 2003

[18] Charles Darwin, *The Origin of Species*, London, 1859.

[19] John H. Holland, *Adaptation in Natural and Artificial Systems*, 1975.

[20] Melanie Mitchel, *An Introduction to Genetic Algorithms*, Massachusetts Institute of Technology, 1996.

[21] David A Coley, *An Introduction to Genetic Algorithms for Scientists and Engineers*, University of Exeter, World Scientific, Singapore, 1999.

[22] A. Gorgulho, R. Neves, and N. Horta, *Applying a GA kernel on optimizing technical analysis rules for stock picking and portfolio composition*, Expert Systems with Applications, Vol. 38, No. 11, 2011, pp. 14072–14085

[23] Choudhry, Rohit and Garg, Kumkum, A Hybrid Machine Learning System for Stock Market Forecasting, World Academy of Science, Engeneering and Technology International Journal of Computer and Information Engineering, vol. 2, N°3, 2008

[24] Lei, Adam Y.C. and Li, Huihua, The value of stop loss strategies, Financial Services Review, vol. 18, N°1, pp. 23-51, 2009

[25] Helen G. Cobb and John J. Grefenstette, Genetic Algorithms for Tracking Changing Environments, 1993