Fuzzy Moving Average System with Genetic Algorithm Optimisation for Trading Crude Palm Oil Futures

Tan Chee Wei, *A0179723U*, Laxman Singh, *A0178223E*, Ramasamy Muthuraman, *A0179756H* and Isaac Varun Kumar, *A0163230W*

Abstract* — The objective of this continuous assessment is to implement the technique proposed in one of the papers shortlisted by the lecture. Our team chose to implement the solution suggested on the paper titled "Quantified moving average strategy of crude oil futures market based on fuzzy logic rules and genetic algorithms" by Xiaojia Liu, Haizhong An, Lijun Wang and Qing Guan. As the title suggests, the authors intend to utilize a variety of moving averages functions using fuzzy logic optimized by genetic algorithm to achieve better returns in crude palm oil futures on the New York Mercantile Exchange (NYMEX). The paper talks about how different moving average can be combined to infer trading signals for buy/sell and use fuzzy logic to determine trading volume. The paper claims that fuzzy moving average strategy proposed gives stable returns when compared to moving average strategies. In our exercise, we learnt the techniques proposed and tried out the proposal to trade crude palm oil (FCPO) futures in the Bursa Malaysia Derivative Exchange. Our model managed to achieve 8,540,649.04 Malaysian Ringgits (MYR) at the end of the 6th year which is a loss of around -15% due to the overfitting of the training data. The rest of this write up discusses the approach, implementation methodology, results and learnings from this exercise.

***Index Terms*— Algorithmic trading, crude palm oil futures, fuzzy logic, genetic algorithm, moving average.**

I. Introduction

IN this work, we have developed a quantitative algorithmic trading system (ATS) based on the paper "Quantified moving average strategy of crude oil futures market based on fuzzy logic rules and genetic algorithms" that was run on Crude Palm Oil (FCPO) futures historical dataset between 03-January-2011 10:30 AM to 30-December-2016 05:59 PM with an initial investment of 10 Million MYR. The dataset given contained per minute trading volume and prices for 37 crude palm oil futures (FCPO) for the entire 6 years. There were constraints/rules (discussed later) that were to be followed while trading. The objective of the exercise is to maximize the profits at the end of given period i.e. 30-December-2016 05:59 PM. For this implementation, we have used python scripting and leveraged a trading platform called backtrader. (https://www.backtrader.com/).

II. Assumptions

One of the core assumptions in our implementation is that the dataset is clean and without any loss in data points. Even though the FCPO are traded for a three month window, the model treats all the FCPO as one dataset.

III. System Architecture and Design

A. Paper Summary

In the paper, the authors are suggesting a fuzzy moving average strategy that utilizes four different moving average methods namely Simple Moving Average (SMA), Adaptive Moving Average (AMA), Typical Price Moving Average (TPMA), and Triangular Moving Average (TMA) between two moving windows - short and long to determine the trading signals. The moving windows for the short and long period are chosen from predetermined set of values. The longer window (m) will be one of {10, 20, 50, 100, 150, 200} and the shorter window (n) will be one of {1, 3, 5, 10, 15, 20}. Similar to traditional buy/sell indicators, the authors are suggesting the crossover of the moving average trend lines as indicators for buy and sell.

The authors are proposing one extra step to mimic human way of trading by suggesting using fuzzy rules to determine the trading volume. The trading volume is determined by trading strength which in turn is determined based on the two moving average periods, the moving average technique, a fuzzy extent (how much strong the difference between the two moving average indicator is) and a random rating score. T Later we will show how this rating contributes to the trading volume. These attributes become part of a fuzzy rule.

A technique is proposed to determine the best fuzzy rule set by randomizing the fuzzy rule attributes and arrive at the optimized membership function. This is done by creating a fuzzy set of 10 rules called as an “individual” in the reference paper. There are 20 individuals that are created. Genetic algorithm is used to determine the best performing individual which will eventually be used for the final trading. Within the genetic algorithm, crossover and mutation are applied and best individual is determined by running the individual on the test data. The result of the test run serves as the fitness function. The individuals are run over 50 times to remove any randomness and bias.

B. High Level System Diagram

In our implementation backtrader platform plays a significant role. As shown in the high-level system diagram (fig. 1) below, the backtrader platform provides the infrastructure to orchestrate the loading of data, replaying in order, invoking the fuzzy rule generator, running the genetic algorithm module for fuzzy rule reconfiguration and eventually to run the chosen fuzzy set on the test dataset to return the final output.

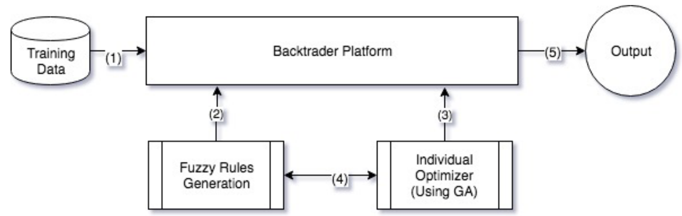

Fig. 1. High level system design

C. Exploratory Data Analysis

The dataset contains 355,959 entries of 37 FCPO. Figure 2 shows a sample of the data fields and data.

	Date	Time	DateTime	Open	High	Low	Close	Volume
97365	20121001	10:30:00	1/10/2012 10:30	2508	2516	2508	2516	6
97366	20121001	10:31:00	1/10/2012 10:31	2516	2516	2516	2516	4
97367	20121001	10:32:00	1/10/2012 10:32	2513	2513	2510	2510	31
97368	20121001	10:33:00	1/10/2012 10:33	2510	2510	2507	2507	39
97369	20121001	10:34:00	1/10/2012 10:34	2506	2506	2504	2504	30

Fig. 2. Sample data

Figure 3 shows the mid-price (average of high and low / minute) plotted for the whole dataset. Based on the chart, it is evident that the trend is downward and at time sideways.

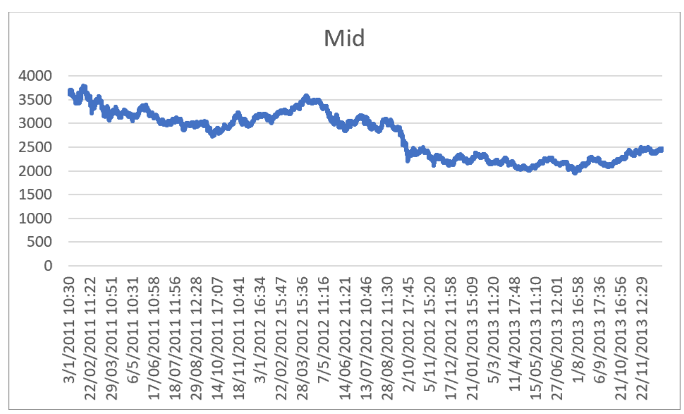

Fig. 3. Mid-price of the entire dataset

For this exercise, dataset between 03-January-2011 10:30 AM to 31-December-2013 close of business is used as the training data. And the rest of the dataset was used for testing/usage of the model generated.

D. Predefined Constraints

The following constraints were considered per requirements from the lecturer. Initial allocation of 10 million Malaysian Ringgit. Trading cost per trade is maximum of (RM 30, 0.2% x volume x price). Short (sell order) incurs additional borrowing cost: 0.01% daily interest accrual based on original market value of borrowed positions from the short day, to be deducted from balance, (immediately after short sell on the short day, and then at the beginning of a next new day onwards), until fully buying back/returning the borrowed positions. Accumulated short position is capped at 5 million Malaysian Ringgit. Maximum one buy and one sell transaction per instrument per minute.

E. Detailed Design

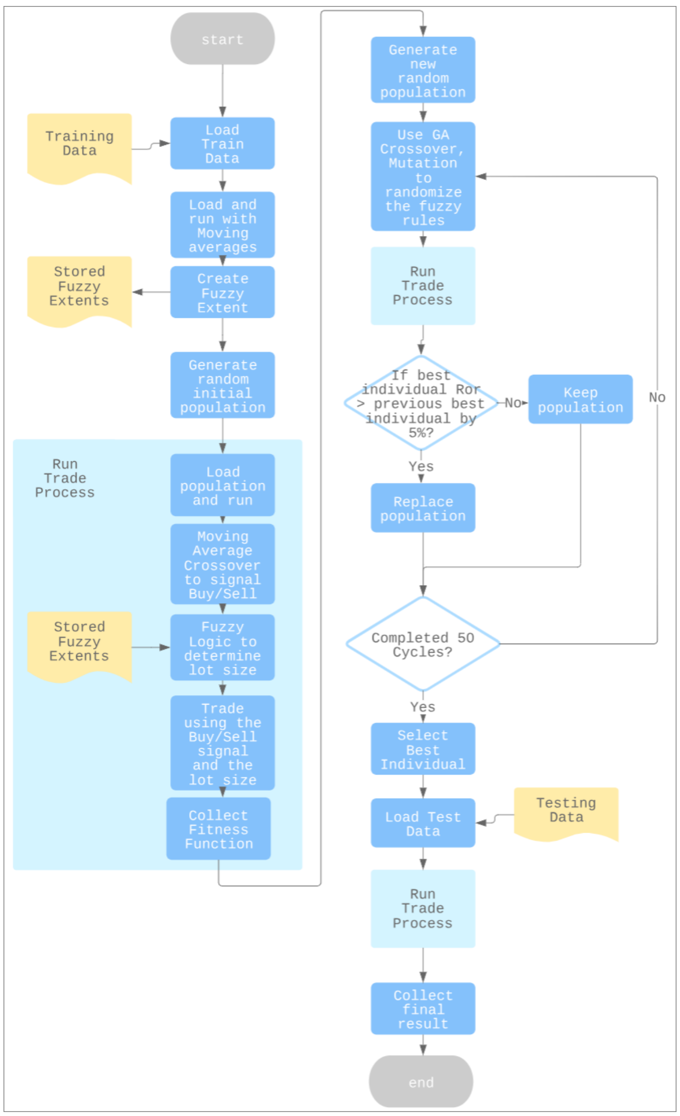

Fig. 4. Overall process and control flow

1. The Data Preparation: Monthly data provided by FCPO futures are combined into two files as training and test data. Train data file was created with transactions records starting from 02 Jan 2011 to 31 Dec 2013. Test data file was created with transactions records starting from 02 Jan 2014 to 31 Dec 2016.
2. Using backtrader platform, we load the train data and determine fuzzy extent and classify to one of the 7 fuzzy categories.
3. We generate 20 individual population with 10 rules combination for each individual. We select the top 2 best individuals with Roulette Wheel Selection (RWS), generate two new individuals and use RWS to select another 16 individuals.
4. The fitness function runs the trading simulation to determine the Rate of Return (RoR).
5. The 16 individuals undergoes mutation and crossover of the rules and individuals.
6. Determine whether best individual in the new population is better than the best individual in the current population based on if the RoR improvement is over 5%, if it is then proceed to use the new population to undergo GA and continue for 50 cycles.
7. At the end of 50 cycles, the best individual from the final population will be finalised.
8. Next, the test data is loaded and the best individual is used to trade.
9. Based on buy or sell moving average signal, the buy or sell decision will be made
10. This cycle will continue until end of test data i.e. 31 Dec 2016.
11. Calculate the RoR.

F. Fuzzy Logic Rule

To determine the trading lot size, fuzzy logic is used. This determines the strength of the buy or sell signal using the moving average crossover. For a particular moving average type, the difference between the fast and slow moving average is collected throughout the training data period and is divided into seven areas, from Extreme Low (EL) to Extreme High (EH).

These areas are determined by dividing the range between 0 (Medium area) and the maximum difference value into three equal portions (High, Very High and Extreme High). This is repeated for the low area by taking the range between the most negative value and zero and applying the same method. That forms the IF rule part of the fuzzy logic rule and we have another fuzzy logic model which will recommend a value from -1 to 1 for the THEN rule part.

The recommended value from this section will be multiplied against the initial recommend value given by the rule from the individual to calculate the average rating. For example, if the initial randomly generated rules contain a SMA 10, 20, fuzzy extent of H and a recommend value of 0.5, we will use the same SMA of 10, 20 to calculate the value difference and apply this on the fuzzy logic to get back the fuzzy extent and recommended value.

Only if the fuzzy extent is matching, let's say the fuzzy logic rule returns a H and a recommend value of 0.4 then we will proceed to use the recommended value from the rule to multiply it against the recommended value from the fuzzy logic rule, in this case, 0.4 * 0.5 = 0.2.

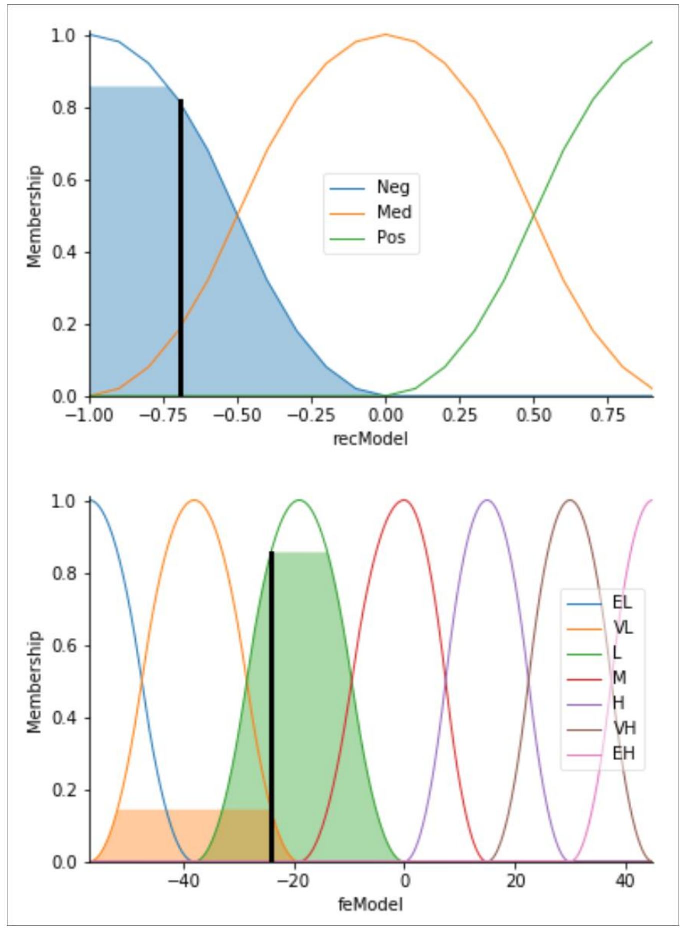

Fig. 5. Fuzzy membership function for antecedent and consequent

G. Genetic Algorithm

Generate the initial population: Initial population contains randomly generated 20 individuals. Each individual contains 10 rules. Single rule will look like as {Moving average, n= lower time period, m= higher time period, Fuzzy extent, recommend value (-1 to 1)} e.g. {TMA, 20,150, HIGH, 0,3}.

Fuzzy extent for each rule will be decided in the population using fuzzy logic along with a recommended value. Genetic algorithm (GA) will be divided into three parts.

1) Selection: GA selects two best individuals from the provided population using roulette wheel selection (RWS). This will create 2 individuals for the new population. Second step in selection, involves randomly selecting 0.8 of the existing population to add up in new population. This will create 16 individuals for new populations i.e., 0.8 * 20 = 16. Third part of the new population is generated randomly. It is 2 individuals for this population as we already have 18 individuals.

2) Crossover: 16 randomly selected individual from old population is passed to crossover. In crossover, parents are selected those are below 0.7 probability. Using parents, offspring are generated taking crossover point as random value between 5(half of the rules in an individual) and 10 (total rules in an individual).

3) Mutation: Mutation is performed on the 16 individuals received from crossover. In mutation, individuals are selected with below 0.01 probabilities. Then, randomly three numbers will be generated between 0 and 10 (total rules in an individual). These three rules in the selected individual are replaced by new randomly generated rules.

Final population is returned to calculate rate of return and performance of individuals to select the best individual.

H. Return on Revenue Calculation

We will be using the following formula as per the paper to implement the rate of calculation but with a slight modification. As the cost is already handled by the backtrader platform, we will exclude that from the calculation. However, for the risk-free rate, we will be following Malaysia's central bank government securities yield with an interest rate of 3.28% to calculate the formula.

$$rreturn = (profit + riskfree - cost)/capital.$$

Variable rreturn means rate of return. Variable profit indicates the money earned from the trade of crude oil futures contracts. Variable riskfree means risk free return. Variable cost indicates the cost of handling charge. Constant capital means the initial capital the trader has in beginning.

IV. Results & Discussion

After the 50th training run, we're able to generate the convergence plot as shown below. From the result we're able to achieve a maximum fitness function using the rate of return of around 1.317 or 32% return. However, we also noticed that the graph is rather volatile with huge swing rotating between the two extremes and suffering slight loss for one run. This can be attributed to the GA portion using the roulette wheel selection to pick the individuals for the new population.

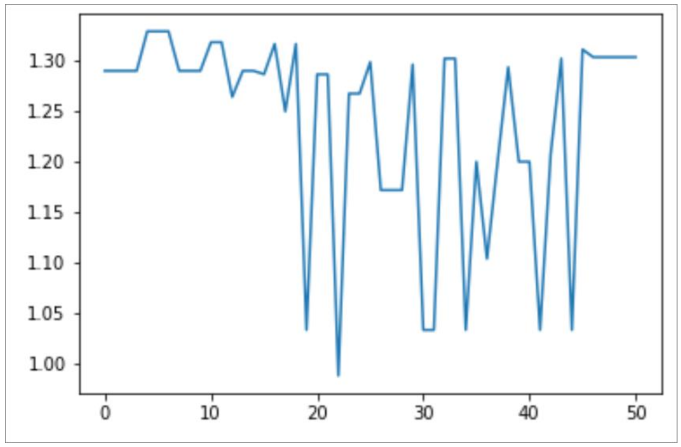

Fig. 6. Convergence plot

After the fiftieth run, the below individual is the best one that the algorithm has selected, and analysing the rules generated, we're seeing there's a bias towards short selling.

Method	Short-Term	Long-Term	Extent	Rating
tma	10	100	H	-0.5
tpma	5	10	VL	0.1
sma	5	150	VL	-0.2
tpma	20	200	VL	-0.6
sma	15	20	L	-0.8
ama	5	100	EL	-0.2
tma	20	200	EH	0.4
sma	20	100	EH	0.7
tma	10	50	EL	-1
ama	5	100	VH	0.1

Fig. 7. Best individual configuration

And using this best individual to trade on the testing data, the result was negative with a loss of -15%. The below plot shows the trading action that commences for the best individual and the final result of MYR 8,540,649.04. Upon further analysis of the 50 best individuals gathered during the training period, we noticed the best individuals pick for each round were slowly orienting towards a heavy short selling strategy. And based on the plot for the testing data, the trend was initially down but the middle portion and the final part of the testing data were on horizontal trend and eventually and upward spike trend proceeded. This has caused the down trend data overfitted model to perform poorly.

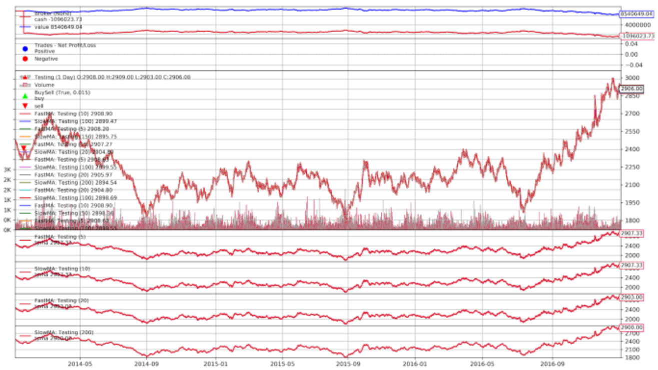

Fig. 8. Training and testing trading plot

V. Conclusion

Our simulation concluded with a NAV of MYR 8,294,017 that is a loss of MYR 1,705,983. Based on further analysis of the outcome using the charts plotted (figure 8) along with another deep analysis of the training and test dataset explains the results. We observed the following – the trend over the training years show a downward or sideway price movement. This explains why the model sells most of the time during the test period even though the market turned bullish. In short, train data being not a combination of up trends and down trends price movements, made the model biased towards down trends.

We also speculate that improvement could be achieved by considering market volume movement during buying and selling. The paper touched about this but didn't explain quite well on applying this. Another options that we considered but couldn't implement on time is to cover up to short sell. The idea is to apply this strategy on top of the current strategy where 3 days old or aged short sell transactions will be fulfilled by buying futures within T+2. Last but not least, if we have more training data containing up and down trend, we could achieve better results.

References

[1] Quantified moving average strategy of crude oil futures market based on fuzzy logic rules and genetic algorithms by Xiaojia Liu, Haizhong An, Lijun Wang and Qing Guan

[2] An intelligent hybrid trading system for discovering trading rules forthe futures market using rough sets and genetic algorithms by Youngmin Kima, Wonbin Ahnb, Kyong Joo Ohb, David Enkec.

[3] Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review Yong Hua, Kang Liub, Xiangzhou Zhanga, Lijun Sub, E.W.T. Ngaic, Mei Liud.

[4] Fuzzy logic, trading uncertainty and technical trading by Nikola Gradojevic, Ramazan Gençay.

[5] Incorporating Markov decision process on genetic algorithms toformulate trading strategies for stock markets by Ying-Hua Chang∗, Ming-Sheng Lee

[6] Backtrader: https://www.backtrader.com/docu/index.html

Appendix

A. Code, setup and execution steps

Prerequisites:

Git	backtrader
Anaconda 1.9.2	scikit-fuzzy
Jupyter Notebook 5.5.0	ta-lib
Spyder 4.0	pandas
Python 3.6	numpy
matplotlib	pytz
datetime	setuptools

Setup:

Copy Scripts/skfuzzy/controlsystem.py to $ANACONDA\Lib\site-packages\skfuzzy\control.
This is required to use the get antecedent function custom implementation.

Running the solution:

Run the following commands on a terminal.

- git clone https://github.com/steve7an/PortfolioU.git
- cd PortfolioU/Scripts

- jupyter notebook
- Open bt2.ipynb on notebook
- run all the cells in order

B. Output

Refer to https://github.com/steve7an/PortfolioU/blob/master/Scripts/bt2.ipynb for the script and its outcome.