Report for CS971 - Evolutionary Computation for Finance

Assignment 4 – Algorithmic Trading

Garry James McBride

20125102041663

University of Strathclyde 2018 – 2019

Contents

Introduction & Literature Review

The following report is about Algorithmic Trading in the financial world. The task was to choose a topic and then train a model to find predictions for the next day in relation to the chosen stock and find a problem of automatically formulating trading rules using one of the techniques. One stock was to be selected and analysed and then to find an interesting factor and experiment on. The choice int technique was very diverse and the decision was free to explore, and then using the results to then be built into trading rules which will automatically buy or sell assets at what should be the optimum point. Using either a GA (Genetic Algorithm) or a NN (Neural Network) to execute the objective, a portion of the data known as 'testing' would be used to evaluate the profitability of your final rule. The methods used must use window sizes to explore the data, since it is in fact time series.

Algorithmic Trading

Algorithmic Trading is a method used when dealing with a larger order of data, which is too large to take on at one given time. Programming plays an important part in Algorithmic Trading as it involves different variables such as: Time, Price & Volume. The data is broken down to child orders (sliced into multiple parts), and sent out to the market, this was developed so that traders would not have to constantly keep watch out of a stock and send them out in slices manually. The end goal is not to make a profit, but is actually a way of minimizing the cost, market impact and risk in an order, used widely by investment banks, a variety of funds such as: Hedge, Pension and Mutual, to complete large orders in an ever growing market that cannot support the full sizes at once due to such a large quantity and quality of data.

Research

A MCAD is used to gauge the strength of stock price movement, instead of primarily focussing on price itself, there are so many factors following different ways a stock will differ from day to day, and they require much further, in-depth analysis to how stocks change. Experimenting with this method of studying a stock or market, the outcome of knowing when and what to sell or buy based on the MCAD instead of a 'straight forward' use of price to predict the next day stock. The study will give an insight into how effective or important this in-depth way of how factors (that relate to price) are used in today's financial industry, and if it is important or as accurate.

RSI

The Relative Strength Index (RSI) uses momentum to measure speed and change of price changes or movements within a stock. Between 0 and 100, it is overbought when above 70 and thus oversold when below 30, these levels can be adjusted to better fit the security of a stock regarding the study of its trend. Swing failures and divergences are indications of signals and are flagged inside a stock option, it is also a very effective and accurate way of identifying the general trend of data being processed or studied.

The RSI is a simple formula, and is shown below in Equation 1…

$$\mathrm{RSI}_t(n) = \frac{\sum_{i=0}^{n-1} (P_{t-i} - P_{t-i-1}) \mathbf{1}\{P_{t-i} > P_{t-i-1}\}}{\sum_{i=0}^{n-1} |P_{t-i} - P_{t-i-1}|} \times 100 \qquad (1)$$

Equation 1

EMA

Exponential Moving Average (EMA) is a type of moving average (MA) which places a significant or greater weight on the vital most recent data points. Reacting more significantly to recent price changes (rather than older or less important) than a simple moving average (SMA). As well as RSI, the EMA is used to produce buy and sell signals from historical averages by means of using crossovers and divergences, using a diverse amount of EMA days that vary from objective to objective, a method used for spotting trends of market moves to gauge a stocks validity. In fast moving market, the EMA is accurate and thus more applicable.

Its limitations are an uncertainty on whether to use most recent days in a time period or more historical data. New data is believed to reflect the current trend the stocks security is moving with, but there are those who feel EMA causes a bias when used. Interpreting an EMA requires close attention to the direction of its line, and the rate of change from one bar to the next, meaning the EMA will diminish and change to zero whenever a price of a strong uptrend begins to flatten or reverse. This method is vital when spotting trends or studying the stock market.

The EMA formula is shown below in Equation 2…

$$\mathrm{EMA}_t = \left[\frac{2}{n} \times (P_t - \mathrm{EMA}_{t-1})\right] + \mathrm{EMA}_{t-1} \qquad (2)$$

Equation 2

Objective

The objective is to predict the EMA & RSI of the next day and use it to find a profit. As researched both these factors are used in the financial world and play an important part. By means of using them, how much profit can be made and are effective when applied to data they haven't seen or without the Neural Network in a more randomly selected predictions for said profit.

Related work

Exponential Moving Average (EMA)

Denis S. Grebenkova and Jeremy Serrorb (2014) state that price variations of stocks are turned into profits by following a trend strategy, and their exponential Moving Averages, efficiently and effectively. Rui Jiang & K.Y. Szeto1 (2003) argues the importance of Moving averages in opening and closing values of a given stock, for purposes ranging from profit or risk analysis. A new approach to moving averages has exploded in the trading industry, the new and improved method works towards heling predictions and spotting trends in data (C.Murphy, 2013).

S.Dash (2013) refers to a study where EMA were used when trading, and the results showed not an increase in accuracy for profit when predicting, but a much more efficient way of reaching an outcome, based on the help and methodology behind studying a stock before predicting and then investing. It might not be a major factor regarding investment, but it has been proved to help the process in preparing to trade or invest, thus results in more accurate results based on its help and methodology.

The Relative Strength Index (RSI)

Taylor and Allen 1992, Neely, 1997 States that the role of RSI has become increasingly popular in the financial world, amongst practitioners, to make accurate and positive investment plans or decisions. It is used in a vast amount of work in investment banks and funds across all financial banks or organizations across the world. With Fama 1965 stating that even though a large amount of studies has been carried out to examine a performance or accuracy in trading rules across the sector, historical price study cannot in fact predict future prices, thus RSI helps towards the objective goal.

With a traditional method of using buy-and-hold applications, moving averages are not as effective or as accurate (Hudson et al 1996). With the experimentation of financial trading rules taking off, Treynor and Ferguson 1985 argues that Fama 1965 is wrong, and that findings result in the opposite of what they appear to be, and that historical data prices can in fact help generate high returns than other methods. A study conducted by Brock et al 1992, backs up the afore mentioned statement that moving averages in trading, tested on Dow-Jones Industrial Average and the results highlighted and concluded that RSI will in fact outperform buy-and-hold strategy.

Arnott, 1979 states that a study was conducted comparing RSI against trading rules, and the accuracy was doubted, but pulled through to be accurate, thus the entrance to trading and introduction of RSI. The study proved that using RSI in trading benefits and secures decision when trading, improving the success of investing, and the security behind large profitable objective.

Chosen Stock

Amazon is the largest e-commerce market place and cloud computing platform in the world, a global success that changed the objective of purchasing products online and surpassed all competitors in a matter of years. It has branched out to other technological advances like the streaming of movies with its service Amazon Prime and has created its own products to access such services, known as the Amazon Firestick. After leaving his high paying job in Wall Street, where he worked in the Computer Science role in a Hedge fund, Jeff Bezos created the website to initially sell books, little did he knows this would make him the founder, creator and CEO of one of the biggest companies in the world… along with making him the most wealthiest person on the planet and the first person to gain a net worth of One Trillion Dollars.

Data Analysis

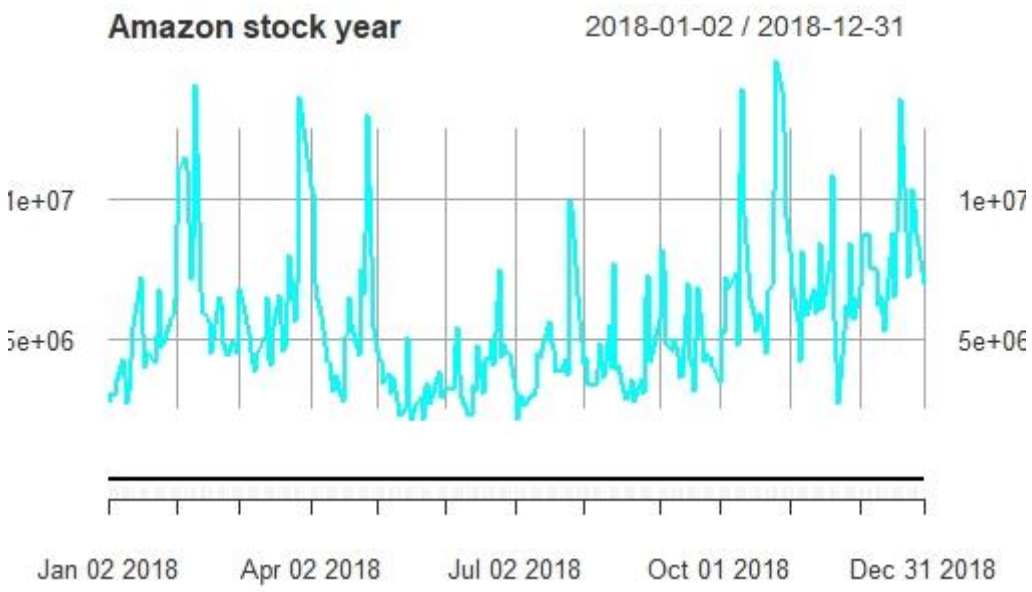

Figure 1

The stock has been chosen over the course of a year, from January 2018 – January 2019. Shown above in Figure 2 is the stock throughout the year and its trend from the dates. The Stock appears to be quite neutral, in some way jumps slightly higher before and after the summer. A curious trend and leaves questions to be answered as to why it increases in these months instead of following the trend through summer.

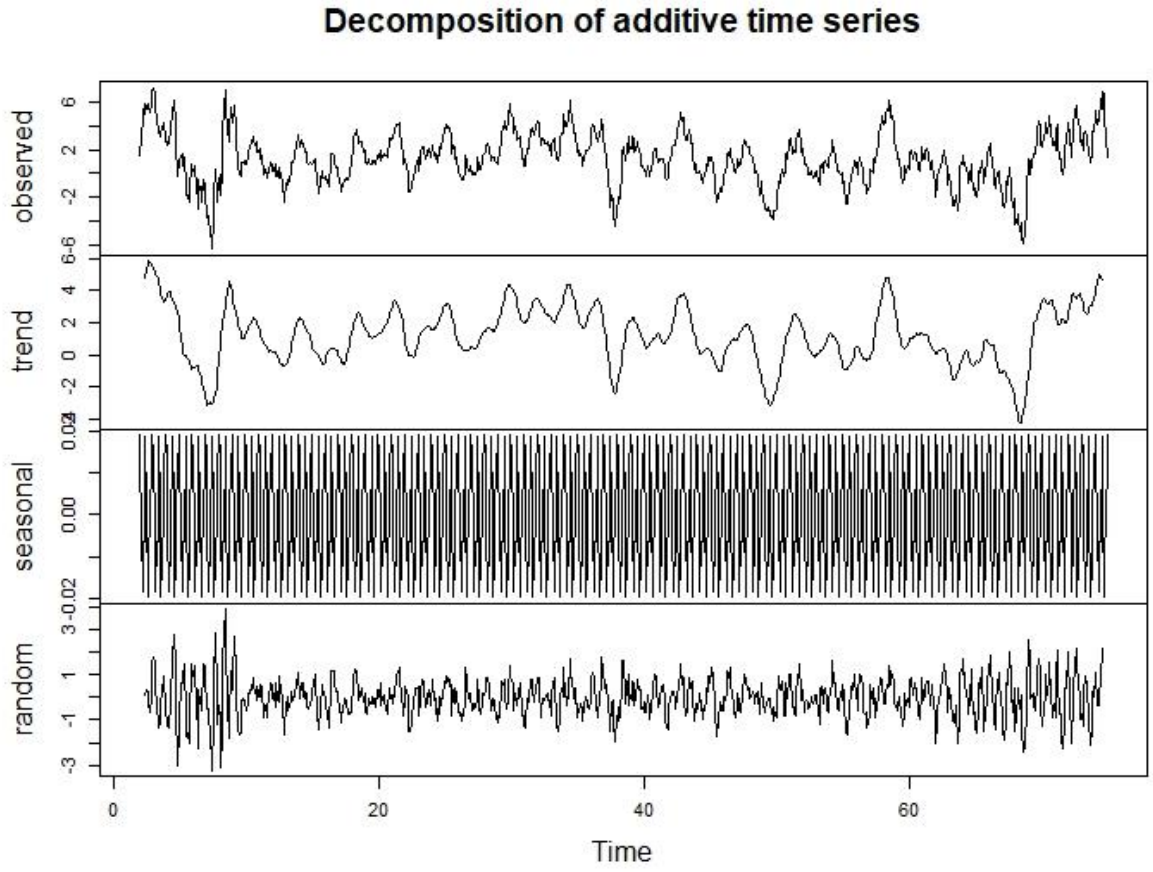

Figure 2

Trend – A complicated and useful technique which looks at past sales, it used to determine possible trends from the given data to extrapolate what could be the future values of the stocks.

Seasonality – One of the most frequently used statically patterns, which creates a predictable cyclic version of the data, this depends on the time of the year.

Cycles – Used to predict changes in the data over the period collected.

The above graph (Figure 2) shows these important aspects and of the stock and gives am more in-depth analysis to its trend, seasonality and cycle. Aspects like these, focus more on just a simple trend, and shows detail in places that could prove be relevant when beginning to analyse a given stock option. The trend is outlined in a simpler way, to see main points and averages where the stock is currently rated.

MEAN	942785.2
MODE	numeric
MEDIAN	1691.595
RANGE	1170.51 - 14963800.00

Table 1

The Mean, Medium, Mode and Range for the stock prices in the data (Shown in Table 1). The mode is numeric, and the range shows a large difference in the lowest stock price compared to the highest.

Although the range may differ by a large amount, the average (Mean) is closer to the lowest Range, showing the stock is usually moderate and has spiked at certain times of the year, shown before in the graph where the stock had boasted before and summer.

Shown below is the head for the data:

	Open	High	Low	Close	Volume	Adjusted
2018-01-02	1172.00	1190.00	1170.51	1189.01	2694500	1189.01
2018-01-03	1188.30	1205.49	1188.30	1204.20	3108800	1204.20
2018-01-04	1205.00	1215.87	1204.66	1209.59	3022100	1209.59
2018-01-05	1217.51	1229.14	1210.00	1229.14	3544700	1229.14
2018-01-06	1236.00	1253.08	1232.03	1246.87	4279500	1246.87
2018-01-07	1256.90	1259.33	1241.76	1252.70	3661300	1252.70

Table 2

With several factors to choose from, such as the Opening, Closing, Highest, Lowest, Volume, Adjusted, EMA and RSI of the stock. It is important to show these details as it gives a feel for the data we are analysing.

	1	2	3	4	5	6
Open	1172.00	1188.30	1205.00	1217.51	1236.00	1256.90
Close	1189.01	1204.20	1209.59	1229.14	1246.87	1252.70
High	1190.00	1205.49	1215.87	1229.14	1253.08	1259.33
Low	1170.51	1188.30	1204.66	1210.00	1232.03	1241.76
Volume	2694500	3108800	3022100	3544700	4279500	3661300
Adjusted	1189.01	1204.20	1209.59	1229.14	1246.87	1252.70
EMA	1189.01	1204.20	1209.59	1229.14	1246.87	1252.70
RSI	1189.01	1204.20	1209.59	1229.14	1246.87	1252.70

Table 3

Again, for a good analysis of the data, the factors from Algorithmic Trading have their Means, Medians, Modes and Ranges below in Table 4. The Modes are all numeric value and the ranges do not differ in a massive size compared between the highest and lowest range. The averages (mean) are in between the ranges, proving these factors do not differ or increase in as large of a scale as the data itself.

	Mean	Mode	Median	Range
Open	1644.073	numeric	1623	1172.00 - 2038.11
Close	1641.726	numeric	1624.89	1189.01 - 2039.51
High	1662.84	numeric	1646.73	1190.0 - 2050.5
Low	1619.841	numeric	1600.15	1170.51 - 2013.00
Volume	5648501	numeric	4955000	2115600 - 14963800
Adjusted	1641.726	numeric	1624.89	1189.01 - 2039.51
EMA	1641.726	numeric	1624.89	1189.01 - 2039.51
RSI	1641.726	numeric	1624.89	1189.01 - 2039.51

Table 4

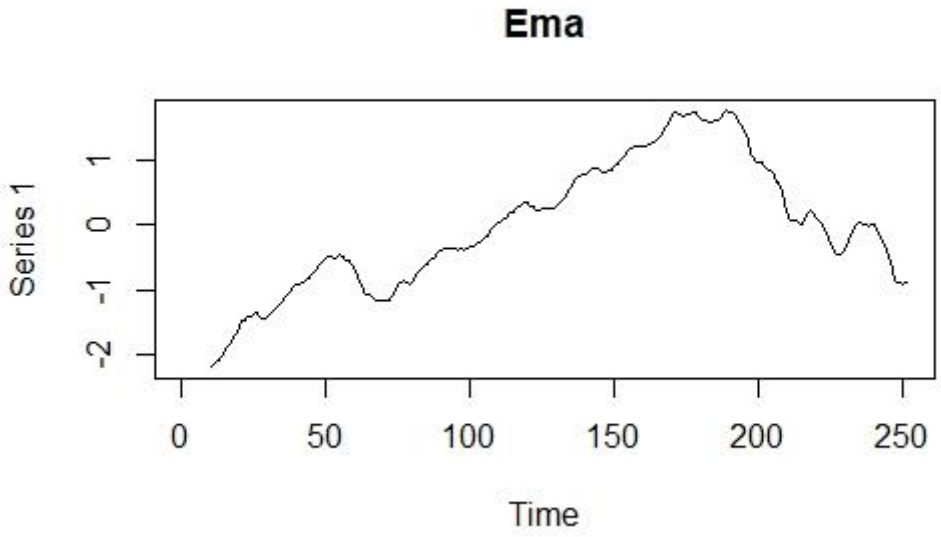

Figure 3

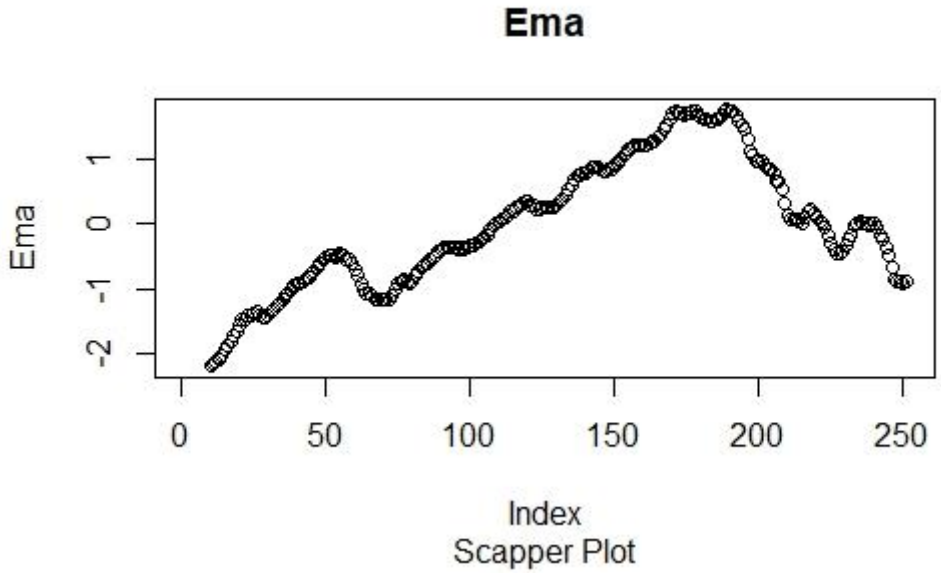

Figure 4

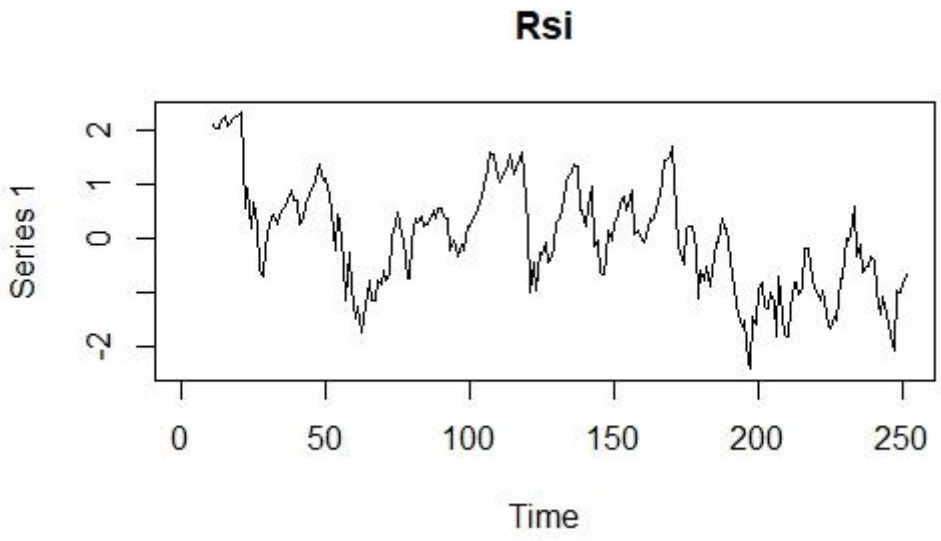

Figure 5

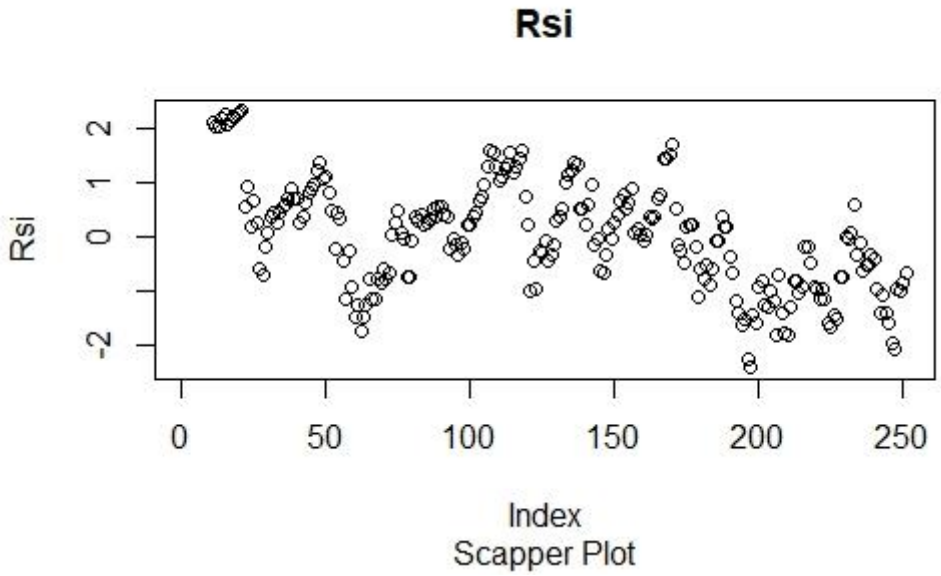

Figure 6

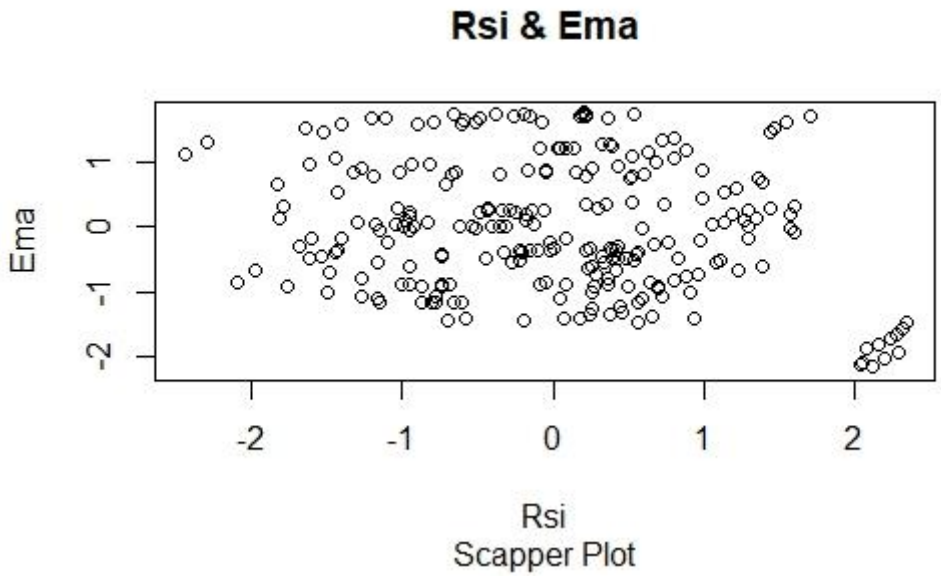

Figure 7

Shown above are the trend of the EMA in the data set. The EMA does not follow the same trend as the data (See figure 3&4), it seems to spike between numbers of 50 – 200. This will be when the stock spikes before and after the middle months of the year. The Importance of showing these numbers is to get an in-depth detail of the factors the project is concentrating on. RSI is shown where it shows a decrease throughout the year (See Figure 5&6). A Scatter plot was created to show the relationship between the two variables since it is the two the project is working with (See Figure 7). The relationship is most popular between -1 and 1 with a strange relationship accruing at 2, strangely away from the rest of the variables.

The Neural Network

What is a Neural Network?

Neural networks are used for a vast amount of machine learning techniques such as; voice recognition or computer vision. The Neural Network is fed information and passes it through layers to find its strongest matching value or asset. Other results will be calculated at the same time, and they will be somewhat accurate, or wrong, but the strongest input based on the target (the output) will be passed through and end up matching with the intended target, based on using weights and biases. The rest of the information fed to the NN will make it to the end, but they will not be the result we are looking for.

Fitness Function and Details

Using a package for the Neural Network which selected it bias automatically, the fitness of the results is based on the error value between the predictions and the actual values. The lower the error the value the stronger the prediction, based on the lowest value, that would be the error value being used going forward.

The R package “Neural Net” is used to produce the Network. The package can be seen and understood better within the code using the ‘help’ command, this gives a better insight to the package for experimentation and research for the best possible results.

To generate training data, inputs or windows are used to select certain parts of the data and slide. To generate a prediction, the window size must have a good amount of values to be accurate, 2 inputs would not be enough to make a prediction, so experimentation was done with using lower or higher inputs to change the range of the windows.

The neural network uses hidden layers to get an accurate result from the inputted data. The more hidden layers used does not necessarily mean the network is more accurate or faster, more layers take longer to be processed and the predication could be found with fewer layers. The Neural Network is shown below with two outputs for RSI and EMA (See Figure 8).

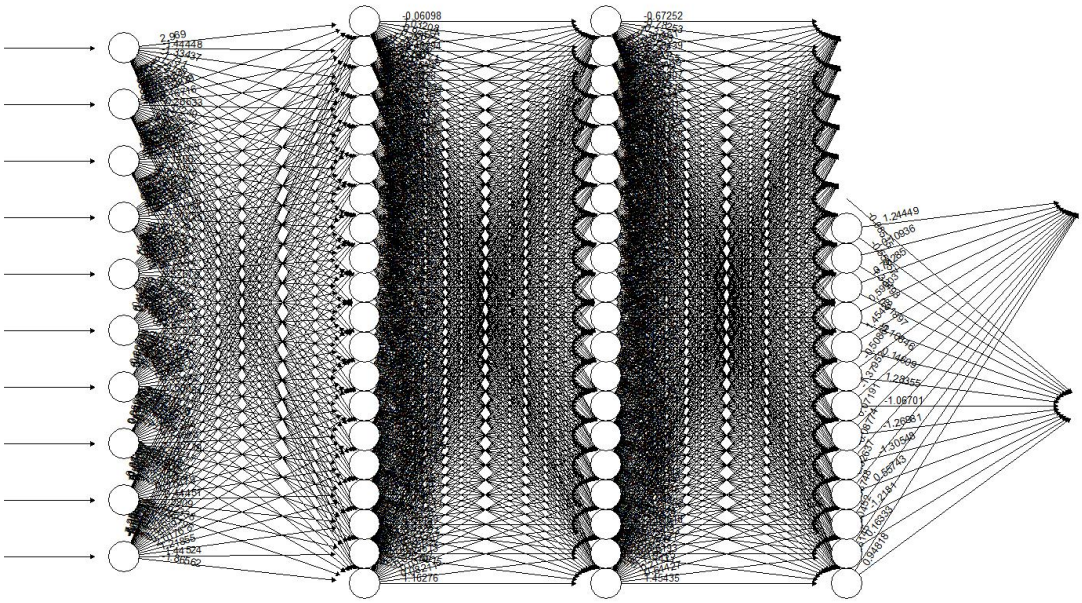

Figure 8

Shown below is the best error value (See Table 5). Changing the threshold, the number of hidden layers in the Network gave different results as shown. The lowest error value is highlighted in red, and uses 3 hidden layers of 20 – 20 – 20 with a threshold of 0.1 and a window size of 12

Inputs	*Hidden Layers*	*Threshold*	*Error*
10	15 15 15 15	0.1	3.245698
12	20 10 10	0.05	5.245689
11	20 20	0.01	5.243589
10	20 20 20	0.1	1.879356
12	15 15 20	0.01	3.258775
11	15 15 15	0.1	2.558496
12	20 20 20	0.1	1.193328

Table 5

Results

Training results

The predictions based on the training results are not as important as the testing, as the objective is for next day predictions in algorithmic trading. The results below in Table 6 show the RSI and EMA together beside the actual results, and the difference between them is quite large since both factors are added together. When doing testing the factors will be separate so a more accurate result can show.

	actual results	*neural net results*
1	-2.14827430	2.14538044
2	-2.11140120	2.10873603
3	-2.07998943	2.27287444
4	-2.02200402	2.30062096
5	-1.93981065	2.25832356
6	-1.87752314	2.03059378
7	-1.80639874	0.80918428

Table 6

Performance

Why use testing data?

By splitting data into two sections, the focus is on the first part of a higher percentage of the other, in the example of the current project, 75% is for training, and the remaining 25% is for the testing. Since the training is passed through the network or algorithm, and the results are influenced by the methods, it is important to try the method on data it has not seen, meaning the extra 25% helps determine if the code is able to perform just as good on data that could be produced at a later date instead of having to re-write everything again every time an organization receives information.

Testing results top 6

The results shown below in Table 7 are the predictions shown on the unseen data and have proven to be accurate. The Neural Network has done well when predicting what the EMA and RSI are going to be, with RSI being considerably more accurate. The first 6 predictions are shown.

	predicted_ema	***predicted_rsi***	***actual_ema***	***actual_rsi***
1	-0.2972818	1.559493	-1.406797	1.604046
2	-0.2651285	1.375789	-1.643263	1.522010
3	-0.4696279	1.217382	-1.526411	1.460710
4	-0.3699700	1.151442	-2.294535	1.297051
5	-1.3134819	1.235963	-2.442489	1.127746
6	-0.9851542	1.623702	-1.445300	1.057531

Table 7

The fitness value or all predictions is shown below in Table 8 for both EMA and RSI. Since RSI has proven to be somewhat more accurate than EMA, the project shall concentrate on the impact and valuable presence the EMA has to offer when used in the financial sector.

	Error
Prediction & Actual EMA	0.9485829
Prediction & Actual RSI	0.4318058

Table 8

From the predictions, we can tell when and what t buy or sell for a total of 59 iterations and see what profit has been presented. The model has been given an unlimited amount of debt, so that it runs for longer to see if a profit can be made after a while. Instead of using a weight (amount of money) the model might have run out of before reaching past a small amount of iterations of buying and selling. A for loop was used so that it knows to sell on the last iteration to stop it from buying when already made a profit, and in the case of being in debt, more money cannot be spent. The for loop uses the fitness function to determine when to buy or sell based on the money of the stock and the error value together. If the stock is prediction is higher for the day after, the method will sell and it is lower prediction for the day after, it will buy.

51	Sell
52	Sell
53	Sell
54	Sell
57	Sell
56	Buy
58	Sell

59	Sell

Table 9

With the process of the above table (Table 9), the method has sold all its stocks after 59 iterations and made no profit, instead it is in debt of the number shown below in Table 10.

Stocks Left	0
Profit	- $235.0701

Table 10

With the method failing to make a profit, it is important to see if it can randomly find profit without a neural network, with the amount of iterations as before. The reason for the fall in profit could be because of the iterations, it may have been positive but because it is waiting to reach 59, by then it could have gone into debt, causing concern for the model that has been implemented.

Comparison

Using a for loop again for the random comparison, with the same method as buying low and selling high for the EMA of the stock the next day, it was randomly run 13 times and jumped between very low debt, and a fair amount of profit in Table 11. With this method producing a better profit, it was not as accurate as its money is all over the place regarding being negative and positive. The best result for profit has been highlighted in red, and is surrounded by very diverse results, an unreliable method, but has proven to beat the Neural Network in this assignment… eventually.

	Stocks Left	***Profit***
RUN 1	0	56.14001
RUN 2	0	168.2698
RUN 3	0	-385.9999
RUN 4	0	-551.0498
RUN 5	0	-239.71
RUN 6	0	-81.15991
RUN 7	**0**	**179.85**
RUN 8	0	-648.4099
RUN 9	0	-15.64038
RUN 10	0	105.7301
RUN 11	0	-18.2898
RUN 12	0	106.1398
RUN 13	0	-21.80981

Table 11

References

Arnott, R. 1979. Relative Strength Revisited. The Journal of Portfolio Management, 5(3), pp.19-23.

Brock, W., Lakonishok, J. and LeBaron, B. 1992. Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47: 1731–64.

C. Murphy, 2013. Moving averages tutorial, " http://courses.jmsc. hku.hk/jmsc7008spring2012/files/2010/02/MovingAverages.pdf, 2013.

Denis S. Grebenkova and Jeremy Serrorb. 2014. Physica A: Statistical Mechanics and its Applications, Volume 394, 15 January 2014, Pages 288-303

Fama, E. F. 1965. The behavior of stock-market prices. Journal of Business, 38: 34–105.

Hudson, R., Dempsey, M. and Keasey, K. 1996. A note on the weak form efficiency of capital markets: the application of simple technical trading rules to UK stock prices — 1935 to 1994. Journal of Banking and Finance, 20: 1121–32.

Neely, C. J. 1997. Technical analysis in foreign exchange market: a layman's guide. The Federal Reserve Bank of St. Louis Review, 79: 23–38

S. Dash, "A comparative study of moving averages: simple, weighted, and exponential, " http://www.tradestation.com/education/labs/analysis- concepts/a-comparative-study-of-moving-averages, last accessed October 13, 2013.

Taylor, M. P. and Allen, H. 1992. The use of technical analysis in the foreign exchange market. Journal of International Money and Finance, 11: 304–14.

Treynor, J. L. and Ferguson, R. 1985. In defense of technical analysis. Journal of Finance, 3: 757–73