Section 1: Week 6: Problem Solving with an Algorithm

Nate Bachmeier

TIM-8110: Programming Languages and Algorithms

June 9th, 2019

North Central University

# Machine Learning in Financial Systems

An article recently appeared in Bloomberg Businessweek entitled “Computer Models Won’t Beat the Stock Market Any Time Soon.” Dewey claims this is due to (1) the market is a random walk that is ever changing; (2) there’s more signal than noise; (3) there is insufficient data; and (4) as proposed by the Efficient Market Theory (EMT) the edge is too small.

This raise follow-up questions: (1) is there validity to his argument; and (2) what characteristics would need to exist in such a model. It is not the goal of this paper to prove of disprove the position only to explore its merits within the broader state of modern scientific discovery. Obviously if Dewey’s argument can be disproven the impact would be significant to the consumers of that computer model. They could choose economic gains or minor celebrity status within the financial community, perhaps even write a best-selling book.

# Exploring the Claims Validity

## It’s a Random Walk

In 1973 Malkiel coined the term in his book “A Random Walk Down Wall Street,” which proposes that (1) all price action is purely random; and (2) an Efficient Market exists where all knowledge is instantly available to all parties. Thus, the best algorithm would be passively investing in a well-diversified portfolio.

Assuming for a moment that it truly is a random walk, this is not a limitation. These scenarios “play a central role in graph theory and in the study of combinatorics, percolation theory, classical and quantum field theory and a myriad of other applications in physics and mathematics (Rudnick, Joseph, & Gaspari, 2004).”

Rudnick describes several algorithms for associating probabilities with clusters of combinatorial results. This is accomplished through ‘generating functions’ paired with statistical and differential equations. While it is not possible to perfectly state, “in 6 months the S&P 500 will be exactly X or Y,” it is possible to assign a likelihood to both values.

## Challenges with Signal to Noise Ratio

The claim states there is too much noise on the ticker tape, which is further compounded by additional media and reporting services. First, the assumption that multiple sources adds to the complexity of modeling discounting the existing vast commoditization of big data technologies. Second, there is an assumption that the noise cannot be filtered.

Astronomical videos are frequently corrupted with ‘impulse noise’ which are bright or dark spots on arbitrary frames. This has led to the necessity for algorithms to clean up the feed.

The naïve solution is to implement a sliding window of N frames, and then take the average value of each pixel across the frame set. A better solution would be to average each pixel using only the frames where that pixel is not corrupt. This can be accomplished per pixel by (1) determine the intensity; (2) determine the nearest neighbors with same intensity; (3) construct a temporal ‘similarity filter mask’ for each neighborhood; then (4) apply the filter mask to omit any unexpected intensities (Aliakhmet, 2019).

A similar model could be applied to the financial markets, where (1) the price or volume represent the intensity; (2) each chart candle becomes a frame; and (3) multiple assets become different pixels within the frame. This may lead to a more efficient “moving average” and if not, there are many more algorithms being created within this field.

## Insufficient Data for Modeling

Financial markets have existed for hundreds of years, yet quality records are only widely available for the last 100. One approach to produce more training data is with the use Generative Adversarial Networks (GAN). GAN pairs two deep learning algorithms in a loop where (a) creates fake data; and (b) predicts if it is legitimate. Both recursively train each other resulting in high quality forgeries.

Another approach is to reach across the sciences and find similar problems in other domains. For instance, dynamically forecasting of conditional probabilities on time series is a hot topic in risk analysis. Li, Zhuang, and Shen proposed an algorithm which uses ‘partial auto correlation functions’ and ‘auto regressive integrated moving averages’ to reliably predict the expected number of terrorist bombings for the week. This is accomplished through correlations of multiple feeds within the Global Terrorism Database (GTD). It may be possible to apply their strategies to financial markets, where correlated data streams are often used to make predictions.

## The Edge Would be Insignificant

Dewey’s final claim is even if such an oracle could be built, it would not be useful as the edge is too small. This is due to Louis Bachelier’s Theory of Speculation, which states that (1) everyone knows everything; and (2) always acts rationally.

Since its publication in 1900, behavior economists have collected enough evidence to fully disprove that people are always rational, especially when money is involved.

Dalton adds that the Efficient Market Theory argument is minimizing the impact caused by participating having different time frames. During a crashing market a short-term account maybe best served by going short. However, a long-term account may see the demise as an opportunity to add to their position.

# General Structure of Desired Algorithm

There is enough evidence to conclude that a model could be created, and that it could potentially beat the market with an acceptable level of consistently. The core algorithm would likely consist of steps (1a) acquire a new quote frame; (1b) apply noise filter; (1c) compute desired moving averages; (2) determine market condition; (3) determine the accounts desired state; (4) execute rebalancing if a threshold has been exceeded.

## 1. Preparing Next Quote Frame

Each frame represents a snapshot of all relevant asset metrics, with an individual pixel containing the pricing and volume information. Like previously described algorithms, each pixel needs to be transformed based on the filtered moving average.

Additional investigations are required to determine if the average should be projected as a relative delta change or literal value. It might also be advantageous to compute the moving averages multiple times with different lengths to represent the views of different time frame participants. Many existing systems already leverage 20, 50, and 200 day moving averages to account for this.

## 2. Determine Market Condition

A market is defined as a collection of assets, such as a basket of technology or international stocks. Each markets condition can be represented as a real number between -1.00 (short) to 1.00 (long). This number can be calculated from market internal metrics, which are often simple aggregates.

A more advanced system could also include a weighted graph with the correlations (Beta) between assets encoded as edge distances. For instance, bonds and equities move in opposite directions, perhaps a sharp move in bonds acts as a signal to change the desired equity position.

## 3. Determine Desired State

There are five dimensions that lead to profitability (alpha) in financial markets. These are represented as the tuple (Beta, Delta, Vega, Gamma, Theta). Beta is the deviation of correlated assets; Delta is gain from 1$ increase in price (direction); Vega is 1% increase in the volatility; Gamma is increase rate of delta gained per 1$ increase; and Theta is cost (interest) of 1 day passing.

The market condition of the desired time frame determines the optimal configuration of the tuple. During an uptrend money can be made from aligning with the trend (Delta), reducing size as it reverts downtrend (Gamma), an understanding that a positive move up decreases volatility (Vega), and not paying interest on the net position (Theta). This desired state would be described by the tuple (0, +Delta, -Vega, -Gamma, 0).

## 4. Transition to Desired State

If the current state exceeds the threshold of desired state; then a rebalance strategy is required. The rebalancing occurs through a multi-dimensional version of the Knapsack Problem, which attempts to select the maximum value from the fewest items.

The available items are listed below with each dimension being a real number between -1.00 and 1.00. For example, 60 deltas can be acquired by (a) 60 shares of stock, (b) 6 x 10 delta calls, (c) 1 x 60 delta calls, or (d) selling 1 x 30 Delta put + buying 1 x 30 delta call. The minimization of the other dimensions introduces more constraints and causes the algorithm to select the most desirable combinations.

* Long Call (+delta, +vega, +gamma, -theta)
* Long Put (-delta, +vega, +gamma, -theta)
* Long Stock (+delta, 0, 0, +dividend - margin)

## Other Considerations

While optimizing the net position there are other dimensions that could be taken into consideration. For instance, having a bias toward the fewest short legs possible increases buying power allowing for more positions at the account level.

The account time frame would impact the number of different positions that should be placed across the account. If a day trader needs to retreat they have less time than an annualized account. This is compounded by each reverted position encountering slippage, which can be minimized with patience.

An ideal algorithm for determining the desired state should not be greedy. If it merely makes the best trade available right now, there will be excessive trades. Each trade will also not receive the best price possible. For instance, it’s cheaper to hedge a down move before the move occurs. It may be possible to mitigate requiring more complex algorithms by using the Beta correlations of the account against leveraged futures. For example, a typical 150k equity account can be exactly hedged using one short S&P Future (-150k notional value).

# References

Aliakhmet, K. (2019). Temporal G-Neighbor Filtering for Analog Domain Noise Reduction in Astronomical Videos. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS, VOL. 66, NO. 5, MAY 2019*, 868-872.

Dalton, J. (2013). *Mind over Markets.*

Dewey, R. (2019, May 21). *Computer Models Won’t Beat the Stock Market Any Time Soon*. Retrieved from Bloomberg Businessweek: https://www.bloomberg.com/news/articles/2019-05-21/computer-models-won-t-beat-the-stock-market-any-time-soon

Li, S., Zhuang, J., & Shen, S. (2017). Dynamic Forecasting Conditional Probability of Bombing. *Risk Analysis, Vol. 37, No. 7, 2017*, 1287-1297.

Rudnick, Joseph, & Gaspari. (2004). *Elements of the Random Walk: An introduction for Advanced Students and Researchers.* Cambridge University Press.

Smith, T. (2019, May 10). *Random Walk Theory.* Retrieved from Investopedia: https://www.investopedia.com/terms/r/randomwalktheory.asp