Executive Program in Algorithmic Trading Batch 35

Applying Machine Learning Ensembles to Market Microstructure to Achieve Portfolio Optimization

Lamarcus Coleman

November 2017

Abstract

"This paper serves as an introduction to a new approach toward solving the portfolio optimization problem, namely the Stereoscopic Portfolio Optimization (SPO) framework. The key assumption of the framework is that a portfolio is the sum of 'n' microstructures. This paper shows that this new idea is a legitimate approach to optimizing quantitative portfolios and trading strategies."

In recent times, the migration to algorithmic and quantitative trading has become more profound. Many traditional asset managers have found the new environment extremely difficult to navigate, while others have found competing against better technologically outfitted institutions insurmountable and have resolved to exit the asset management industry. It is apparent that the global financial industry is evolving before our eyes. Traditional means of asset management must evolve to comply with the current conditions within the global financial industry. This paper presents a case for why today's revolution is in some ways like prior technology driven revolutions within the financial industry, but yet very different. A brief elucidation is offered of what the global financial markets may look like in the future. A review of machine learning is conducted as well as how the same can be employed to rethink portfolio and risk management. A review of market microstructure is presented as it not only has evolved with every financial market revolution, but has become of paramount importance given the current state of the global financial industry. Portfolio optimization is reviewed and a portfolio is categorized as being the sum of 'n' market microstructures.

A study is conducted in which quantitative portfolios are comprised of an initial capital value of $70 million USD. The portfolio began as four strategies, namely Microstructure, Volatility Dispersion, Statistical Arbitrage, and Trend Following, but, was later reconfigured to two of the four due to data and computational resource limitations. Two specific strategies are implemented per portfolio, namely, statistical arbitrage within the US equity markets and trend following within G-10 currency futures. Four portfolios

are constructed of these strategies. The first portfolio is an equally weighted portfolio that does not employ any optimization technique. The second portfolio is optimized by traditional methods such as the Efficient Frontier, referred to as top-down optimization. While the third is optimized by employing a bottom-up machine learning optimization approach, while the fourth combines the traditional top down approaches with the bottom-up machine learning approach to form a Stereoscopic Portfolio Optimization (SPO) framework. The bottom-up machine learning optimization approach employs the use of an ensemble on a market microstructure component with the function of augmenting the complex event processing engine should the strategy enter a predetermined regime. Each ensemble varies across strategies regarding its objective function. For instance, the objective function of the ensemble within a given Statistical Arbitrage strategy is to augment the signal generator if the probability of entering a non-cointegrated regime crosses some threshold. The determination of regimes across each strategy is determined by employing the use of Gaussian Mixture Models.

Volatility is the key market microstructure component in which this work is concerned. The ensemble method is built around volatility. The ensemble method's task is to improve the risk-adjusted returns of the portfolio by optimizing each respective strategy via its market microstructure.

K-Means clustering is employed to determine tradeable relationships within Statistical Arbitrage.

The objective of this study was to determine the degree in which a Stereoscopic Portfolio Optimization(SPO) framework, leveraging the use of machine learning ensembles and their application via the market microstructure to disparate strategy objectives, impacts portfolio optimization. To assess the performance of this effort, the Mu, Sigma, Sharpe and Sortino metrics will be used.

This study serves as an introduction for further and continued analysis. It was found with reference to optimizing K in K-Means, if one's desire of tradeable relationships, is sufficiently small, K can be derived as $\frac{1}{2}\sum n$. This work also revealed that Gaussian Mixture Models can be applied to existing trading strategies and used as an optimization framework. The study showed that while multiple regimes were generated for Statistical Arbitrage pairs, a single component was generated for each of the five currency futures traded. This required an impromptu update to this work which encompassed designing a method, within the machine learning portfolios, to augment the signal generator based on a given lag of volatility and a historical threshold. The impact of prior regimes on future optimization was also understood. The Efficient Frontier was calculated on the out of sample period of 1st Jan. 2015 to 1st Jan. 2016. Over this period the Statistical Arbitrage strategy regime had returns in excess of 36% with a Sharpe above 2. This skewed the weights of each regime grossly in favor of the Statistical Arbitrage strategy. After composing the four portfolios, it was found that Portfolio B, or the Stereoscopic Portfolio Optimization (SPO) framework implementation outperformed all other portfolios relative to their Sharpe ratios. Consistent with the idea of this work serving as an introduction to the idea of SPO, the machine learning models underwent a single implementation. In other words, the models were not reconfigured to window dress the idea. The Gaussian Mixture Models' regimes were calculated on out of sample data and used as labels within the Random Forests. The Random Forests were designed once and no recalibration of the model was made. The performance of the SPO framework, amid these conditions, displays reasoning that despite exceeding the performance of the equally weighted, Efficient Frontier, and bottom-up optimization framework portfolios, the SPO's performance was a conservative example of its potential.

This study yielded a mix of challenges and surprises. In addition to data and resource limitations, the skew of the Efficient Frontier, and the currency futures isolation to a single regime, 2016 being an election year yields significant volatility toward the end of the test period. This directly impacted the equity curves of the portfolios as well as the measures of risk-adjusted return. Also, this occurrence provided an avid example of how regime shifts can catch the SPO model off guard if the training period is elongated. The Gaussian Mixture Model was trained over 2012-2015, tested over 2015-2016, and then applied to 2016-2017. Thus, the regimes identified in 2012-2015 training period were a composite of those years. This illustrates that the Random Forests model's precision is one key contingency to this ideal as the future regime could be any of those over the historical period. Also, it was found that amid massive increases in volatility unparallel to that experienced over the testing period of 2016, new regimes are created that the model has not been trained on. This illustrates that the SPO model is likely best implemented either 1) on a rolling basis (i.e. 3 month lookback) in which the regimes and model is derived from the lookback period and then used to trade a forward period of one month, and or 2) the SPO framework is used on a higher timeframe(i.e. 2 month lookback) and then the SPO model is applied to intraday data to capture the fractal nature of the markets. Though the SPO was merely introduced and not optimized in this study, the results thus far point toward one of the two aforementioned design patterns over that used in this study. However, the framework can be implemented for longer holding periods, pending a periodic recalibration.

What makes this paper unique is its unconventional approach to the portfolio optimization problem. To the best of my knowledge, no prior research has presented a similar approach to that of the Stereoscopic Portfolio Optimization (SPO) framework introduced in this text, and thus makes it the first of its kind toward solving the portfolio optimization problem. Prior researchers have applied a multiplicity of techniques to the portfolio optimization problem which encompass the use of evolutionary and genetic programming, conditional value at risk, and the classic mean variance approach, to name a few. Support vector machines, random forests, and artificial neural networks, etcetera have been applied to discrete strategy optimization problems largely involving parameter selection. Yet, it is noted that prior research, while applying machine learning to the problem of strategy parameter optimization, has done so outside of the context of the portfolio optimization problem. Also, no researcher to date has addressed the problem of portfolio optimization by implementing a stereoscopic approach that encompasses the use of machine learning ensembles applied to individual strategies within a portfolio based on an objective function, to achieve portfolio optimization, as well as simultaneously address the strategy optimization problem. By stereoscopically, I suggest that no prior work has optimized a portfolio from the classic top-down vantage of seeking to allocate capital efficiently, as well as address both the portfolio optimization and strategy optimization problems by optimizing the portfolio based on the market microstructure of each individual strategy, through the use of machine learning ensembles on the market microstructure to optimize the trading objective of each individual strategy.

As a result, this work surveys the optimization of the portfolio via 1) the microstructure of each strategy with a specific model to optimize the strategy's objective and 2) optimization of the composite of the strategies (ie the portfolio).

Another pulchritudinous trait of this work is that the study assumes the vantage of an asset manager by constructing an institutional portfolio of various strategies as well as multiple versions of the portfolio with varying influences by the machine learning ensembles' application to the market microstructure of the component strategies.

The prime objective of this study is to survey the extent that applying machine learning ensembles to the market microstructure of each strategy further optimizes the overall portfolio, the extent in which the SPO outperforms alternative approaches, and the degree in which further research is merited for the SPO framework.

Keywords: machine learning ensembles, market microstructure, statistical arbitrage, volatility dispersion, trend following, portfolio optimization, Random Forests, Gaussian Mixture Models, K-Means Clustering

1. Introduction

The portfolios constructed within this paper arise from the need of asset managers, and traders alike, to cope with and better understand the role of technology within financial markets. The hypothesis being tested here is that portfolios constructed of uncorrelated quantitative strategies that employ the use of machine learning ensembles as a means of monitoring and responding to the market microstructure, should outperform portfolios constructed by traditional means.

Market microstructure encompasses the study of price discovery, dissemination of information, and the operability of exchanges. Market microstructure study also entails the functioning of a stable market. Some common focal points of microstructure study centers around order flow imbalances and market structures. Other areas of microstructure study include order types, transaction costs, liquidity, open interest, and volatility. Within this study, the key focal area within the microstructure domain is volatility.

Portfolio optimization is the process of balancing risk and reward. Seminal works have offered mechanisms upon which to measure the variance of portfolios as well as individual strategies' performance. Modern portfolio theory, founded by Harry Markowitz, introduced a revolution within the asset management landscape. This theory suggested that a certain return be required for a certain degree of risk or exposure. Notable within this field of study are the Capital Asset Pricing Model and Efficient Frontier. This work suggests that in addition to these revolutionary developments, as a result of the dynamism of financial markets and rapid progression of automation, technological advances should be employed in innovative ways, as means of augmenting these frameworks for modern application.

Technological shifts play an important role in the evolution of markets. An avid example of this can be seen from the migration of market microstructure pre and post open outcry. Prior to the establishment of indoor trading venues, trading was conducted outside and exposed traders to the elements. In the US, orders would be shouted down from windows to the traders standing outside. In other countries, trading would convene at local coffee shops as traders monitored the ticker tape. Eventually, exchanges were constructed that served as official trading venues. The open outcry system provided auctions in which traders would conduct business in pits specific to the assets that they traded. Locals or independent traders would monitor the order flow, or the institutional traders, also known as paper for indications of order flow imbalances. Traders would compete not only for orders but positioning on or near the top step. Over time, electronic trading began to marginalize the open outcry system. This mode of trade offered a more efficient means of conducting business. The transition to electronic trading also created more competition in the marketplace as individuals who were unable to compete on the floor now were afforded the opportunity to do so via the screens. As a result of this, the market microstructure began to evolve as well. Electronic trading served as the precursor to algorithmic trading. The shift toward algorithmic trading yet again marked a shift in the microstructure and a new admission of a new class of traders. Each shift in technology results in an accompanying shift not only in the microstructure of financial markets but in the participants as well, of whom also make up the market microstructure. I term this as Recursive State Space Entanglement, or when Space A impacts Space B, thus changing the state of B, which in turn changes the Space of B, which reinforces change in both the state and space of A. Or in other words, technological innovation engenders change in the financial markets and changes the components of the market microstructure. The market microstructure is the sum of its components (of which participants are included) and thus changes in the state of the microstructure induces changes in the space of the financial markets. This update in the space of financial markets is met with new demands on the space of technology, which in turn changes the state and space of that arena, and thus the cycle continues.

The study of market microstructure became ever more important with the rise of high frequency trading. Manual traders, as a result of the entrance of new players, had to adjust their strategies as algorithmic trading rendered some of their strategies less profitable. A survey of the evolution of markets in relation to that of technology provides one with a keen construct upon which to forecast future shifts in market microstructure and market participants. The initial shift from open outcry to electronic trading was met with resistance by some traders. Those who failed to admit that they were part of a dynamic system and thus adjust to the changing environment were later relieved of their position as market participants. An understanding of the historical evolution of markets justifies the notion that this time things are different. The shift from open outcry to electronic trading eliminated market participants as pits closed, and thus provided no possible alternative. Many believe that the shift toward algorithmic trading will not yield a similar result as this apparent transformation will not eliminate the new venue of trade of which is electronic. This notion is fallible in that it underestimates the response by the financial market ecosystem. Though discretionary trading may theoretically remain possible, it is probable that doing so later within this existing revolution will not be practical. The markets evolve in a manner that prioritizes efficiency and seeks to provide offerings to those with economies of scale. Service providers, as the percentage of traders trading algorithmically increases, will imperatively transform their business models to meet the demands of the majority of market participants, thus inevitably creating structural disadvantages for manual traders. An example of this can be seen in the transformation of the business model of exchanges. Another highly important notion is that as the percentage of traders trading algorithmically expands, the market microstructure is likely to evolve in a manner in which algorithmic trading will become a minimum to even compete in the financial markets. Though these ideals will not eliminate the possibility of manual trading, it is certain to make doing so on a large scale unprofitable. In addition to this apparent evolution remains the fact that unlike the transition from the open outcry system to electronic trading, we are now in a recursive evolutionary environment. The markets are evolving multiplicatively. The shift toward algorithmic trading was the onset of quantitative trading and high frequency trading. The continued progression of the same induces a greater foot race toward implementing machine learning and other modes of artificial intelligence. These respective transformations are occurring simultaneously. This insinuates that the financial markets will rebalance the competitive advantages of market participants. The evolution of cloud based crowed sourced platforms will continue to expand and challenge the positioning of traditional firms. The pecking order within the retail space will change dramatically as a result, as a divide will occur between retail and professional retail traders. The discretionary firms that fail to adjust to the shift toward not only algorithmic and quantitative trading but artificial intelligence as well will see their edge reduced beneath that of professional retail traders who implement quantitative trading methodologies. As the world as a whole competes for individuals equipped with the skills necessary to implement quantitative and artificial intelligence models, competition will increase across industries for this talent. This will affect the competition landscape within the financial markets. The same will also almost inevitably eliminate late adopters of quantitative trading methodologies as these firms will not be able to compete with the likes of existing quantitative trading firms, banks, hedge funds, and other external competition from outside the financial markets.

Key to this evolutionary process will be the evolution of asset management. Due to the aforementioned exhibit of the probable continued progression of the advancement in the financial markets and thus the market microstructure of securities, asset management must employ the use of machine learning and other modes of artificial intelligence to remain competitive and do so innovatively.

Machine learning, also defined as statistical learning, is the process of applying statistical methods to real problems to derive inferences and predictions. Machine learning can be segmented into supervised, unsupervised, and reinforcement learning. The goal of the model is to reduce the prediction error. Researchers have found that the combination of multiple learners has the ability to reduce prediction error. This methodology is referred to as ensemble learning. Given the aim of combining multiple, sometimes weak models together to achieve a more optimal prediction, it appears that the objective of ensemble learning is analogous to that of portfolio optimization whose goal is to find the most optimal combination of a given set of assets within a portfolio toward the end of reducing trading error, or risk. It is thus plausible that ensemble learning could serve as an optimal means for achieving portfolio optimization.

This paper's focus is on applying ensemble learning methodologies to the microstructure of portfolio strategies as an additional means toward achieving portfolio optimization. A portfolio is viewed as being a composite of the microstructures of each asset traded. This can be viewed recursively from the portfolio, to each individual strategy and the microstructures of their respective assets. Below is an intuitive mathematical representation of this ideal.

$$P(t) = \sum_{a=w=1}^{n} | a_n w_n + e_t |$$

P(t) represents the value and composition of a portfolio at a static period of time t; a_n represents the market microstructure of each respective strategy, w_n represents the weight of the respective market microstructure; e_t is an error term and is illustrative of randomness; the absolute value signs indicate the non-linearity within the portfolio as the market microstructures can have either a positive or negative effect on the portfolio.

$$\frac{d}{dt}P = \lim_{t \to \infty} \sum | a_n w_n c_n + e_t |$$

The above equation is a novel way of stating that the dynamic value of the portfolio consists of the sum of the market microstructures times their respective weights multiplied by the correlation term plus the random variable. The correlation term here is not the correlation of the price or return series, but an intuitive example of the correlation between microstructure factors such as liquidity, volatility, etc. The correlation term is present in the second equation, but absent in the first, illustrating the role of correlation in static versus dynamic states. In the first equation, the portfolio is static, thus the relationship between the microstructures is null or muted because the state is not changing. However, in the second equation, as t approaches infinity, illustrating the constant evolution of the portfolio, the correlation term is present emphasizing the ideal that correlation is most important during dynamic states. Stated another way, traders are most concerned with the correlation of assets when a change in one asset can and or will induce a change in the other asset. If the assets are not changing, they cannot induce a change in another asset, thus the static representation of the portfolio omits a constant correlation term.

The shift toward algorithmic and quantitative trading and the resulting effect on the market microstructures of securities serves as an avid representation of the need to seek more revolutionary methods of constructing optimal portfolios.

The remainder of this study is organizes as follows: 1) a review of the literature is provided partitioned into market microstructure and portfolio optimization sections; 2) the approach of this study is offered followed by an explanation and empirical look at each strategy component,(it is noted that while late in the study, the Volatility Dispersion strategy was removed, yet the theoretical components of the strategy have been moved to an Appendix B.) 3) Section2: Data, provides the data used in the study, 4) Section3: Methodology, provides an illustration of the study, and 5) Section4: Conclusion, summarizes this work.

Literature Review

Market Microstructure

The study of market microstructure is an examination of how a given market is structured and designed. "Market microstructure is the branch of financial economics that investigates trading and the organization of markets."(Harris, L. 2003)

Joel Hasbrouck (2007) identified the following as constituting a typical analysis of market microstructure: 1) sources of value and reasons for trade, 2) mechanisms in economic setting, 3) multiple characterizations of prices, 4) liquidity, 5) transparency, and 6) econometric issues. Hasbrouck (2007) also offered a variety of queries yet needing resolution in the study of market microstructure. Inclusive to these were, "What are optimal trading strategies for typical trading problems?", "Exactly how is information impounded in prices?", "How do we avoid market failures?", "How do we enhance the information aggregation process?", and "What can market/trading data tell us about long-term risk?"

The financial markets are dynamic. From a fundamental view, asset prices are believed to reflect the relationship between supply and demand and or the value of the underlying security. Over shorter time horizons, this is often not the case as market participants re-price securities based on a variety of factors that in some instances may not be even remotely pertinent to the underlying value of the asset. This dissemination of information and the accompanying process of price discovery present trading opportunities. "Contrary to Efficient Market Hypothesis, the basis of Market Microstructure is that asset prices need not reflect full information. Frictions such as competition for order flow, rules and trading protocols and information asymmetry cause friction." (Gupta, A. 2017)

Financial markets can assume a variety of structures. "The trading rules and the trading systems used by a market define its market structure." (Harris, L. 2003) Trading rules dictate the level of information and transparency within a market. An understanding of market structure is a prerequisite for profitable trading. Trading strategies vary across different asset classes as a result of the varying market microstructure of those respective assets. "The trading strategies that are successful in one market often do not work well in markets with different structures."(Harris, L. 2003)

Traders can participate in two types of trading sessions, continuous trading and call markets. During continuous trading sessions, traders are tasked with structuring capital allocation around the hours in which the market is open. In contrast, within call markets, trades may take place whenever a market is called. An example of a call market is the auctioning of US Treasuries. The auctions take place when called, or scheduled by the government. "Many continuous order-driven exchanges open their trading sessions with call market auctions and then switch over to continuous trading. These markets also use calls to restart their trading after trading halts."(Harris, L. 2003) The primary advantage of continuous markets is that they offer traders flexibility in arranging trades. "The main advantage of call markets is that they focus the attention of all traders interested in a given instrument at the same time and place. When buyers and sellers search for liquidity at the same time and place, they can easily find each other."(Harris, L. 2003)

Markets can be categorized as being order driven or quote driven. "In order driven systems, all trades are arranged by using order precedence rules to match buyers to sellers and trade pricing rules to determine the prices of the resulting trades."(Harris, L. 2003) Markets that are order driven are dependent upon the interaction between buyers and sellers. Priority is disseminated on the basis of price and time in the order book. The order book is often referred to as the limit order book as market participants display orders indicating where they are willing to buy and sell. This is often referred to as providing liquidity. Liquidity takers are those participants who prioritize execution over price and thus offset the orders of liquidity providers. These participants pay the bid-ask spread while liquidity providers receive the bid-ask spread. Size priority is another facet of order priority. This encompasses orders being executed according to a pro rata rule. Filled orders are allocated based on a rank of order size. Pro rata execution allocates order fills to depth at the same price on a pro rata basis. "Order-driven market structures vary considerably. Some markets conduct single-price auctions in which they arrange all trades at the same price following a market call. Other markets conduct continuous two-sided auctions, in which buyers and sellers can continuously attempt to arrange their trades at prices that typically vary through time. Still others conduct crossing networks, in which they match orders at prices taken from other markets." (Harris, L. 2003)

A quote driven market is constituted as one in which buying and selling is transacted between traders and dealers or market makers. The dealers quote where they are willing to conduct business. "Either traders negotiate with the dealers themselves, or their brokers, acting as their agents, negotiate with the dealers."(Harris, L. 2003) Quote driven markets are also called dealer markets because the dealers serve as liquidity providers. If Trader A would like to buy a security and Trader B would like to offer the same security, in a quote driven or dealer market, the two would not be able to transact business with each other. Both Trader A and Trader B would have to conduct business with the dealer. The dealer would take trades based on their inventory. Hypothetically, Trader A would have to buy from the dealer and Trader B would in-turn have to sell to the dealer. The dealer serves as an intermediary within pure quote driven or dealer markets. "In some dealer markets, traders can trade with each other without the direct intervention of a dealer. Although these are not pure quote-driven markets, they are still known as quote-driven markets because dealers supply most of the liquidity and arrange all the trade."(Harris, L. 2003)

Markets in which brokers match buyers and sellers are termed brokered markets. Liquidity providers are classified as being two types of traders within brokered markets, concealed traders and latent traders. Concealed traders desire to be active in the markets but avoid revealing their trading intentions as a means of averting market impact. "Latent traders do not know that they want to trade until brokers present them with attractive trading opportunities."(Harris, L. 2003) Selectivity affords latent traders with the ability to defer costs until a suitable trading opportunity arises. "The most important brokered securities markets are those for large blocks of stocks and bonds. Although these securities may trade in very liquid markets for small sizes, brokers must find suitable counterparties for most large blocks."(Harris, L. 2003)

Hybrid markets are those in which elements from each type of market are visible. An example of a hybrid market can be seen in the New York Stock Exchange. The NYSE essentially conducts business as an order-driven market. However, it also requires its specialist or dealers to be the last resort of liquidity. This gives the NYSE characteristics of both order-driven and quote driven markets.

The behavior of financial markets can be surveyed as the interaction between its various participants. Market participants can be classified as traders or speculators, hedgers, brokers, dealers, market makers, etc. Traders or speculators study the order book dynamics and look for statistical probabilistic opportunities to execute orders. Brokers execute order instructions on behalf of clients for a transaction fee. Market makers are responsible for providing liquidity to ensure a fair and stable marketplace. "There are four types of market making firms: (1) Retail market making firms own a retail brokerage network that serve individual investors. The order flow from this segment facilitates liquidity provision for the company's stock; (2) Institutional market making firms specialize in executing large block orders for pension funds, mutual funds, insurance companies, and asset management companies, among others; (3) Regional market making firms focus on both companies and retain investors of a particular region. The regional market maker gives companies and retail investors the benefit of expert in-depth knowledge of stocks of a particular area of the country, providing more extensive coverage than might be available elsewhere; (4) Wholesale market making firms provide liquidity services for institutional clients as well as for other broker-dealers who are not registered market makers in a stock . They provide liquidity for a company's stock by being an important source of trade facilitation for retail, institutional, and regional firms. " (Krishnamurti,C.,2009)

"Liquidity is the ability to trade large size quickly, at low cost, when you want to trade. It is the most important characteristic of well-functioning markets."(Harris, L. 2003) All market participants play a role in the liquidity within a given market. Traders either take or provide liquidity; brokers arrange liquidity, and dealers supply liquidity. Liquidity is a force that enables profitable trading. Liquidity is measured as a means to test the feasibility of an execution system or whether or not a specific trading strategy is practical. "Liquidity is the object of bilateral search. In a bilateral search, buyers search for sellers, and sellers search for buyers. When a buyer finds a seller who will trade at mutually acceptable terms, the buyer has found liquidity. Likewise, when a seller finds a buyer who will trade at mutually acceptable terms, the seller has found liquidity."(Harris, L. 2003) Market participants can further be classified as active or passive. Passive participants display their orders so that those in search of liquidity can easily find them. In contrast, active participants are those who are in search of liquidity and take liquidity when they find opportunities that meet their terms. The search described here by the active participant is termed a unilateral search for liquidity. The search for liquidity can be describe as being the process of utilizing inputs to produce outputs. The primary input to the search for liquidity is the time expended. The outputs to the search for liquidity are good prices and the execution of desirable sizes. The bilateral search for liquidity is related to three dimensions of liquidity, immediacy, width, and depth. "Immediacy refers to how quickly trades of a given size can be arranged at a given cost. Traders generally use market orders to demand immediate trades. Width refers to the cost of doing a trade of a given size. For small trades, traders usually identify width with the bid/ask spread. It also includes brokerage commissions. Width is the cost per unit of liquidity. Traders often refer to market width by the term market breadth. Depth refers to the size of a trade that can be arranged at a given cost. Depth is measured in units available at a given price of liquidity."(Harris, L. 2003) Harris 2003 categorized liquidity providers into dealers, bid/ask spreads, block traders, value traders, and arbitrageurs.

Dealers trade passively. They seek to profit buy trading with their clients. Dealers may be identified by other names in various markets. They are known at stock exchanges as specialist or market makers. At futures exchanges they may be called scalpers, day traders, locals, or market makers. The bid/ask spread is set by dealers as a means of maximizing profits. The idea is to offer spreads that are wide enough to allow for the recovery of costs. This creates a balancing act for the dealer. If spreads are excessive, traders will refuse to trade with the dealer. "The transaction cost spread component is the part of the bid/ask spread that compensates dealers for their normal costs of doing business…This component also funds any monopoly profits that the dealer may make and any risk premium that dealers may require for bearing inventory risk."(Harris, L. 2003)

Block traders must cope with the issue of finding liquidity. Due to the sheer size of their orders, they must seek liquidity more strategically to avoid market impact. Block initiators are those traders looking for liquidity, while passive block traders are those willing to offer it. Both parties cannot publically display their intents. Block liquidity providers may question the level or degree of information asymmetry from block trade initiators. Price discrimination is another concern of block liquidity providers. To resolve these issues, block trade initiators must reveal their intents and convince block liquidity providers that they are uninformed.

Value traders access all known information about a security a then determine a fair price for that security. "Value trading is profitable only when price differs from fundamental value. Price can differ from fundamental value two ways: 1) when new information causes fundamental value to change, and thereby deviate from price, or 2) when uninformed traders push price away from fundamental value."(Harris, L. 2003) Value traders serve as liquidity providers when prices do not reflect fundamental value. They are a last resort for liquidity in a sense.

Arbitrageurs seek to profit from the relative value of securities. They assume market neutral positions. These traders focus on securities that tend to be correlated over time. There are a variety of types of arbitrage strategies. Inclusive to these are shipping arbitrages, delivery arbitrages, conversion arbitrages, spreads, pairs trading, statistical arbitrage, and risk arbitrage. Arbitrageurs are informed traders and their effect on liquidity is based on the informational advantage. "A large basis tells arbitrageurs to move liquidity from one market to another or to convert risk from one form to the other."(Harris, L. 2003)

Each liquidity provider specializes within a specific area of providing liquidity. These liquidity providers leverage their information advantage. The information that they contain relative to market participants and the fundamental value of securities with a market drive their decisions, and thus liquidity.

“Market makers have little information about fundamental values or their clients. They specialize in offering immediacy to small traders. Block dealers know a lot about their clients. They offer depth to uninformed clients. Value traders know more than everyone else does about fundamental values. They generally do not care much about with whom they trade. Their confidence in fundamental values allows them to be the ultimate suppliers of depth. Precommitted traders trade for reasons other than to supply liquidity. Since they already intend to trade, they do not care much about with whom they trade. They typically supply immediacy. Arbitrageurs are well informed about relative instrument values. When their arbitrages involve very low risk, they ensure that traders can access the depth in any market that trades the instruments that interest them.”(Harris, L. 2003)

“Since informed traders generally hurt dealers, and since dealers generally do not know when they trade with informed traders, dealers try to infer information about fundamental values from their order flows.” (Harris, L. 2003)

Order flows are what drive markets. Order flow is the interaction that takes place between buyers and sellers. This includes order arrival process, the depth, and breadth of the market. The markets reflect where buyers and sellers would like to conduct business. Informed traders study the order flow process and construct models for predicting price changes. Despite the plethora of technical indicators available, order flow is the only leading indicator to price change. Joel Hasbrouck (2007) offers an illustration of the order flow arrival process as a Poisson process. “In a limit order market, orders arrive randomly in time.” (Hasbrouck, J. 2007) The limit order book is also highly dynamic due to the ability of traders to cancel or modify orders at any time. Hasbrouck also modeled the distribution of buys and sells. “The probability of a buy does not depend on the direction of the last or any prior orders.”(Hasbrouck, J. 2007) Hasbrouck found that in the absence of informed trading, the distribution of buy orders is approximately normal. However, it was discovered that as informed trading increased, the component distributions became more distinct. The elevation of informed trading will limit to one-sided order flow. The limit order book displays the size and depth at prices upon which traders would like to conduct business. A variety of order types may be employed. To avoid market impact, traders seek to disguise their intentions by employing the use of hidden and reserve orders. “Hidden orders are the simpler of the two. If an order designated as hidden cannot be executed, it is added to the book but not made visible to other market participants. The hidden order is available for execution against incoming orders, the senders of which may be (happily) surprised by fills at prices that are better than or quantities that are larger than what they might have surmis4ed based on what was visible. Hidden orders usually lose priority to visible orders, a rule that encourages display.” (Hasbrouck, J. 2007) Reserve or iceberg orders are similar to hidden orders but they are not completely invisible. When executing a reserve order, an initial quantity is displayed and once that lot has been filled, another predetermined lot size is displayed on the order book.

"The limit order's execution probability is an important and difficult-to-model aspect of the problem. If the order is displayed (and in a modern electronic market, it usually will be), it changes the information and strategic incentives of other market participants. Execution probability is therefore properly modeled as an equilibrium phenomenon."(Hasbrouck, J. 2007)

Volatility is unexpected price change. The nature of volatility varies from market to market and over time. Prices tend to change as new information is disseminated and a need arises to re-price the value of securities. "You must understand market structure, and how it affects trader behavior, in order to understand the origins of market liquidity, price efficiency, volatility, and trading profits."(Harris, L. 2003) Volatility is closely correlated to market risk and thus a trading strategy's profitability. While an understanding of volatility is important across all markets and strategies, it carries a greater degree of importance for options traders. The value of the underlying contract is contingent upon volatility. This requires that options traders be able to measure and forecast volatility to achieve profitability. Larry Harris (2003) presented fundamental and transitory volatility as being two types of volatility. Fundamental volatility arises as a result of an unanticipated change in a security's price. Transitory volatility occurs as a result of the trading activity conducted by uninformed traders. Price change, fundamentally, is created by any factor(s) that alter the value of a security. These fundamental factors vary across markets. Volatility arises when these fundamental factors change unexpectedly. Volatility is closely correlated with certainty or predictability. The more predictable an event is, the lower the volatility around that event. In contrast, the greater the levels of uncertainty with respect to a specific event, the more volatile the price of securities. "Total volatility is the sum of fundamental volatility and transitory volatility. People generally measure total volatility by using variances, standard deviations, or mean absolute deviations of price changes."(Harris, L. 2003)

Statistical models are constructed as a means of measuring volatility. This is especially prevalent in the options markets. Volatility models seek to exploit the difference between the two types of volatility. Fundamental volatility differs from transitory volatility in that it constitutes random price changes that do not mean revert, while transitory volatility tends to display mean reverting behavior. "The transitory price changes are generally correlated with order flows of uniformed liquidity demanding traders."(Harris, L. 2003) This behavior causes price changes to be negatively correlated. Large increases in price tend to follow large declines in price and vice versa. This creates a larger probability of price reversals than continuations amid transitory volatility. "The presence of negative serial correlation in price series is therefore a strong indicator of transitory volatility. (Harris, L. 2003)

Regulators are tasked with enacting policies that advocate a fair and stable market. Extreme volatility is one of the elements of market microstructure that regulators monitor as means to impede financial instability. "Some policies that they consider can create markets which are more resilient. Other policies have little value, and many policies can harm the markets. Regulators therefore must carefully analyze how market structure affects volatility before adopting new policies." (Harris, L. 2003) A key policy that regulators enacted as a result of extreme volatility over time was trading halts and circuit breakers. Circuit breakers limit trading activity amid extreme volatility. Trading halts stop trading after a security's price has moved by some predetermined amount over a small interval. The halt can remain in effect for a specified period of time, or until the order imbalance is sorted. Price limits are another mechanism employed to reduce volatility. Price limits place bands on the prices in which traders must conduct business. "Proponents of trading halts and of price limits believe that these circuit breakers reduce short-term volatility by slowing price changes…Opponents of trading halts and price limits offer two arguments that suggest these circuit breakers may actually increase transitory volatility. First, if traders fear that a halt will occur before they can submit their orders, they may submit their orders earlier to increase the probability that they execute…Second, value traders may reduce their surveillance of the market if they know that the media will notify them when a trading halt occurs. If so, a trading halt rule would make the market less liquid between trading halts. Transitory volatility would increase and more trading halts would occur." (Harris, L. 2003) Other regulatory actions that seek to maintain market stability are transaction taxes, margin requirements, position limits, and collars. Transaction taxes seek to limit trading by taxing it. Margin requirements and position limits directly impact the size of trades that traders can take. Collars limit access to trading systems.

Portfolio Optimization

Modern Portfolio Theory, created by Harry Markowitz in the 1950s is a construct that suggests optimal portfolios can be created by balancing risk and reward. This methodology is also termed mean-variance analysis. Prior to Markowitz's study and further examination of portfolio construction, investors measured performance by returns only and risks, though known, were not quantified. A portfolio can be optimized on multiple layers. First, the appropriate weight or balance can be determined relative to the capital allocation to specific asset classes. Second, the assets and or strategies within each asset class can then be surveyed for the proper balance between risk and reward. The Efficient Frontier is a visual representation of the tradeoff between a portfolio's assumed risk or variance and reward or returns. The Kelly Criterion is often used to find the most optimal weights of each asset or strategy within a specific segment of the portfolio. The Kelly Criterion considers an asset's or strategy's probability of positive returns and negative returns, and payout or hit ratio at various percentages of investable capital in comparison to other assets or strategies as a means of arriving at the most optimal allocation to each asset or strategy within a given segment of a portfolio. The Kelly formula focuses on achieving sustainable portfolio growth. An alternative to the Kelly Criteria is the Martingale Method. This method is counterintuitive and suggests increasing your position size after every loss and decreasing it after every win. Portfolio optimization defies the Efficient Market Hypothesis which assumes that individual security selection is null as prices reflect all known information. "The strong form of EMH states that security prices fully reflect all knowable information. Furthermore, intensive analysis will not enable the analyst to reach judgments different from the market's prices with enough consistency to earn additional returns." (Security Analysis, 1987) Markowitz (1959) stated that a portfolio is inefficient if it is possible to obtain higher expected (or average) return with no greater variability of return, or obtain greater certainty of return with no less average or expected return.

Inherent to portfolio optimization is the measurement of risk and returns. Various methodologies have been employed to measure the risk and return of a given strategy and or portfolio. Some notable methods for assessing risk adjusted returns are the Sharpe, Sortino and Calmer ratios. The Sharpe Ratio measures the amount of alpha per unit of standard deviation. One criticism of the Sharpe Ratio is that it penalizes all volatility.

$$\text{Sharpe ratio} = \frac{R_p - R_f}{\sigma_p}$$

The Sortino Ratio seeks to address this criticism and is a modification of the Sharpe Ratio into a formula that takes the sigma of negative returns only into consideration.

$$\text{Sortino ratio} = \frac{\langle R \rangle - R_f}{\sigma_i}$$

<R> = Expected Return
R_f = Risk Free Rate
σ_f = Standard Deviation of Negative Returns

The Calmer Ratio is also known as the drawdown ratio. This ratio is formulated as being the average annual rate of return over the trailing three years divided by the maximum drawdown over the same period.

Risk measurement is a critical component of portfolio optimization. The assessment of risk can be divided into three categories: absolute risk measures, relative risk measures, and tail risk measures. Variance and maximum drawdown are two means of assessing absolute risk. Variance, an indication of volatility, is the squared deviations from the average or mean return of an asset, strategy, or portfolio. Maximum drawdown is expressed as a percentage and represents the maximum loss from a peak to a trough.

(Portfolio Variance 2 asset example)

$$\sigma_{pf}^2 = w_a^2\sigma_a^2 + w_b^2\sigma_b^2 + w_a w_b \sigma_a \sigma_b \rho_{ab}$$

Relative risk metrics seek to compare the risk of an asset to that of a benchmark or alternative security. The correlation coefficient and beta are two relative risk measures. The correlation coefficient depicts the strength of the relationship between two variables while the beta measures the volatility of an underlying asset or portfolio versus that of the market. The beta of the market is one, thus a portfolio with a beta greater than one is considered to be more volatile than the market while a portfolio with a beta of less than one is considered to be less volatile than the market.

The most popular tail risk measure is that of Value-at-Risk (VaR). Value -at-Risk measures the maximum loss probable over a predefined period for a given confidence interval. This can be calculated using a bootstrapping method of applying significance level to a return distribution over a given period of time, or by conducting Monte Carlo simulations. “The popularity of VaR is mostly related to a simple and easy to understand representation of high losses. VaR can be quite efficiently estimated and managed when underlying risk factors are normally (log-normally) distributed.”(Krokhmal, Palmquist, and Uryasev 2001) Conditional Value-at-Risk (CVaR) takes the weighted average between VaR and the losses exceeding VaR. This metric is also known as the Expected Shortfall and is considered to be superior to VaR. “For general distributions, CVaR, which is a quite similar to VaR measure of risk has more attractive properties than VaR.” (Krokhmal et al.) “For continuous distributions, CVaR is defined as the conditional expected loss under the condition that it exceeds VaR…For continuous distributions, this risk measure also is known as the Mean Excess Loss, Mean Shortfall, or Tail Value-at-Risk. However, for general distributions, including discrete distributions, CVaR is defined as the weighted average of VaR and losses strictly exceeding VaR.”(Krokmal et al.)

Parameter optimization plays an important role in the construction of individual strategies within a portfolio. This involves fine tuning input parameters that drive strategy performance. Parameter or strategy optimization can be achieved with the stated goal of maximizing profits, reducing drawdowns and drawdown duration, or by some combination of both.

Researchers have conducted a variety of studies toward the end of achieving an optimal portfolio. Krokhmal, Palmquist, and Uryasev (2001) developed a case study on the S&P 100 involving the implementation of Conditional Value-at-Risk (CVaR). “The primary purpose of the presented case study is the demonstration of the novel CVaR risk management methodology and the possibility to apply it to portfolio optimization.” (Krokhmal et al. 2001) The researchers displayed the possibility of employing different optimization methods for risk-return optimization problems with convex constraints. The efficient frontier was constructed using constraints relative to the CVaR. The group also offered a comparison to that of the standard Markowitz mean-variance framework. It was mentioned that for normally distributed loss functions, the two approaches yield an equivalent efficient frontier. Yet, in the case of non-normal and more significantly, non-symmetric distributions, the CVaR and Mean-Variance optimization approaches could yield significantly different results.

Tunchan Cura (2008), building on the work of Kennedy and Eberhart(1995), offered a particle swarm approach to portfolio optimization. Kennedy and Eberhart (1995) stated that particle swarm optimization is rooted in two main component methodologies, artificial life and evolutionary computation. In literature, it is stated that there exists five basic principles of swarm intelligence. The first is the proximity principle which deems that the population be able to complete basic space and time computations. Secondly, the quality principle, details the population's ability to respond to quality factors within its environment. Third, the population should avoid undertaking its activities via excessively narrow channels. This is defined as the diverse response principle. Fourth, the population is prohibited from altering its behavior at each interval in which that of the environment changes. This is called the stability principle. Lastly, the population must be able to alter its behavior to achieve an optimal reward. This is the principle of adaptability. Note that the fourth and fifth principles are inversions of each other. In other words, the population should not change its behavior every time the behavior of its environment changes, yet if a change in behavior is deemed to be optimal, the population must adapt to the change of its environment.

A Cardinality Constrained Mean Variance (CCMV) model was developed. This model built on Markowitz's Mean-Variance Model. A cardinality model restricts the number of assets and the upper and lower bounds of the proportion of each asset in the portfolio. The following is the equation for the CCMV: (note: in contrast to Markowitz's MV Model, the model below includes a lambda variable. This variable represents a risk aversion parameter. A lambda equal to zero maximizes the average return of the portfolio despite its variance. In contrast, one that is near unity reduces risk despite potential return. As the lambda increases so does the investor's sensitivity to risk and vice versa.)

$$\min \lambda \left[\sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \sigma_{ij} \right] - (1-\lambda) \left[\sum_{i=1}^{N} x_i \mu_i \right]$$

$$\text{subject to } \sum_{i=1}^{N} x_i = 1,$$

$$\sum_{i=1}^{N} z_i = K,$$

$$\varepsilon_i z_i \leq x_i \leq \delta_i z_i, \quad i = 1, \ldots, N$$

$$z_i \in \{0, 1\} \quad i = 1, \ldots, N.$$

The idea is to restrict K, or the number of assets in the portfolio using the decision variable $z_i \in \{0, 1\}$ to determine whether or not an asset will be included. An additional constraint is that an asset's weight remains within the upper and lower bounds of ε_i and δ_i respectively.

This model is derived from the standard Markowitz model below:

$$\min \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \sigma_{ij}$$

$$\text{subject to } \sum_{i=1}^{N} x_i \mu_i = R^*,$$

$$\sum_{i=1}^{N} x_i = 1,$$

$$0 \leq x_i \leq 1. \quad i = 1, \ldots, N$$

where N is the number of different assets, σ_{ij} is the covariance between returns of assets i and j, x_i is the proportion of asset i in the portfolio, μ_i is the mean return of asset i and $R*$ is the desired mean return of the portfolio.

The study introduced a Particle Swarm Optimization (PSO) heuristic method. "The swarm in PSO consists of a population and each member of the population is called a particle, which represents a portfolio in this study."(Cura 2008). A fitness function was also defined as being:

$$f_p = \lambda \left[\sum_{i=1}^{N} \sum_{j=1}^{N} z_{pi} x_{pi} z_{pj} x_{pj} \sigma_{ij} \right] - (1 - \lambda) \left[\sum_{i=1}^{N} z_{pi} x_{pi} \mu_i \right]$$

where f_p is the fitness value of particle p.

The fitness function is a function that evaluates the best position of a particle and that of its neighbor. At each of the iterations the best position and the best neighbor are updated to reflect the new information obtained. Simply put, this is a construct that illustrates how information is disseminated throughout a population and all members of the population profit from the sharing of knowledge. Each particle or portfolio is comprised of N dimensions or assets. Each particle includes a proportion variable X_{pi} which represents the number of particles in the population. Each particle also includes decision variablesZ_{PI}.

Cura (2008) provided the following as means for determining the velocities of the particles:

$$vz_{pi}^{t+1} = vz_{pi}^{t} + \omega_1 \times (Gz_{bi} - z_{pi}^{t}) + \omega_2 \times (Gz_{pi} - z_{pi}^{t}) \tag{14}$$

$$vx_{pi}^{t+1} = \begin{cases} vx_{pi}^{t} + \omega_1 \times (Gx_{bi} - x_{pi}^{t}) + \omega_2 \times (Gx_{pi} - x_{pi}^{t}) & \text{if } z_{pi}^{t+1} = 1, \\ vx_{pi}^{t} & \text{otherwise,} \end{cases} \tag{15}$$

where both ω_1 and ω_2 denote uniform random numbers between 0 and 2, t and b denote the iteration number and the best particle in the swarm respectively, vx_{pi}^{t} denotes the velocity of particle p on dimension x_i, and vz_{pi}^{t} denotes the velocity of particle p on dimension z_i. As seen in Eq. (15), vx_{pi}^{t+1} will be updated if asset i is selected by particle (or portfolio) p at iteration $t+1$, which means $z_{pi}^{t+1} = 1$, and z_{pi}^{t+1} is described in Eq. (16). Gx_{pi} denotes the best previous position of particle p on dimension x_i, and Gz_{pi} denotes the best previous position of particle p on dimension z_i. Thus, particle p moves at iteration $t+1$ as follows:

$$z_{pi}^{t+1} = round\left(\frac{1}{1+e^{-\varsigma}} - \alpha\right) \tag{16}$$

$$x_{pi}^{t+1} = \begin{cases} x_{pi}^{t} + vx_{pi}^{t+1} & \text{if } z_{pi}^{t+1} = 1, \\ x_{pi}^{t} & \text{otherwise} \end{cases} \tag{17}$$

where $\varsigma = z_{pi}^{t} + vz_{pi}^{t+1}$ and α is set to 0.06. For a given particle, if the velocity on dimension z_i^t is zero, this particle will not move in that dimension at iteration $t+1$. Suppose $vz_{pi}^{t+1} = 0$ and $z_{pi}^{t} = 0$, hence $1/(1+e^0) = 0.5$ and $round(0.5) = 1$, which means that particle p will move in dimension $z_i(z_{pi}^{t+1} = 1)$ at iteration $t+1$. In order to avoid such an unwanted move, we can use α as seen in Eq. (16).

Cura (2008) found that the efficient frontier constructed using cardinality and bounding constraints can vary from that of the standard Markowitz model. The PSO model to the PO problem was evaluated in comparison to other heuristic models. It was found that no individual model outperformed the other, but when the aim is to construct portfolios with low risk of investment, the particle swarm optimization model outperformed other heuristic optimization methods.

Ping, Yingwei, and Yong (2016) provided an analysis of applying an asymmetry robust mean absolute deviation (ARMAD) model to the problem of portfolio optimization. This approach was implemented as a means to address the symmetrical risk classification of up and down moves within the Mean-Variance model which deviates from reality. Ping et al.(2016) created an ARMAD to capture asymmetrical return distributions. An analysis of the in and out of sample models were surveyed in comparison with the Chinese market conditions. Ping et al. (2016) mentioned the work of Konno and Yamazaki (1991) in that they construct a mean absolute deviation model that avoids the calculation of the covariance matrix of assets. They also find that the optimal portfolio of the MAD and

$$\min_{x} \frac{1}{T} \sum_{t=1}^{T} \left| \sum_{j=1}^{n} (r_{jt} - r_j) x_j \right|$$

Markowitz's MV model yield similar performance. In lieu of this, it is stated that the MAD model is an acceptable alternative to the Mean-Variance model. Below is the MAD model presented in the text.

$$\text{s.t.} \sum_{j=1}^{n} r_j x_j \geq \rho M_0, \tag{1}$$

$$\sum_{j=1}^{n} x_j = M_0, \tag{2}$$

$$l_j \leq x_j \leq u_j,\ j = 1, \ldots, n, \tag{3}$$

which is equivalent to the following linear programming, denoted by (MAD):

$$\min_{x,y} \frac{1}{T} \sum_{t=1}^{T} y_t$$

$$\text{s.t. } y_t + \sum_{j=1}^{n} (r_{jt} - r_j) x_j \geq 0,\ t = 1, \ldots, T, \tag{4}$$

$$y_t - \sum_{j=1}^{n} (r_{jt} - r_j) x_j \geq 0,\ t = 1, \ldots, T, \tag{5}$$

$$(1), (2) \text{ and } (3).$$

Ping et al. (2016) proposed variations to the MAD and different means of making the model robust. The researchers compose a portfolio of 30 small cap Chinese stocks over the period of 10/27/2008-12/03/2012. This period is further divided into two segments, the overall market and the declining market. After surveying the varying performances of the ARMAD, RMAD, and MAD the researchers found that the ARMAD model can provide useful

information but simultaneously increases the complexity due to its reliance on second order cone programming in comparison to the complexity of a linear programming problem.

Chen, Yang, Abraham (2006) constructed a flexible neural trees ensemble on the Nasdaq-100 and the NIFTY indices. The FNT is coupled with the use of genetic programming and the parameters are optimized by particle swarm optimization (PSO). Flexible neural trees are employed for selecting important inputs, time delays and forecasting models. The Nasdaq-100 was analyzed from January 1995 January 2002 and the NIFTY was analyzed from January 1998 to December 2001. The training and test data were divided into approximately two equal parts.

To optimize the neural tree Chen et al. (2006) used the following mutation operators: 1) Changing one terminal node: random selection of a single terminal node and replacement with another node, 2) Changing all the terminal nodes: selecting all terminal nodes and replacing with another terminal node, 3)Growing: selection of a random leaf in the hidden layer and replacement by newly generated subtree, 4) Pruning: random selection of a function node and replacement with a terminal node.

The researchers began the implementation of the particle swarm optimization (PSO) by first generating a population randomly. Each particle represents a solution and has position represented by a vector Xi. The swarm of particles movement through particle space is represented by a velocity vector Vi. A function representing a quality measure takes in Xi as an input and is calculated at each step. Each particle tracks its own best position. This position is associated with the best fitness it has achieved and stored in a vector. The best position of the all the particles are also stored in another vector. In addition to this standard PSO, Chen et al.(2006) create another PSO that keeps track of the best position among all the topological neighbors of a particle.

To conduct feature selection for the FNT, the researchers identified the following methods: 1) the input variables initially are selected with the same probabilities, 2) variables that have a greater influence on the objective function will be enhanced and have an increased survival probability based on an evolutionary procedure and 3) the evolutionary operators provide a input selection method so that the FNT can select appropriate variables automatically. Chen et al.(2006) offered the following as means for constructing the general learning algorithm 1) create an initial population (FNT trees and its corresponding parameters), 2)Structure optimization is achieved by the neural tree variation operators, 3) if a better step is found go to step 4 else go to step 2, 4) Parameter optimization is achieved by the PSO algorithm, 5) if the maximum number of local search is reached, or no better parameter vector is found for a significantly long time then go to step 6 else go to step 4, 6) if satisfactory solution is found, the algorithm is stopped else go to step 2.

Chen et al. (2006) used the root mean squared error (RMSE), maximum absolute percentage error (MAP) and the mean absolute percentage error (MAPE) and correlation coefficient to study the performance of the trained model. Ten FNT models were implemented on the Nasdaq-100 and NIFTY respectively. Following, three ensemble methods were used to predict both indexes. The researchers found that the local weighted polynomial regression (LWPR) outperformed other models and had the highest correlation coefficient and lowest MAPE and MAP values. Seeking to predict the share prices of the following day based on the opening, closing and maximum values of the same day, the researchers found that key parameters that affect share prices are their immediate opening and closing values.

Securities portfolios are the sum of 'n' market microstructures. Technological advancement and revolutions impact and alter market microstructures. In lieu of this, it is plausible that the continual migration toward the employment of more technological resources will thus alter the means by which portfolios are constructed and managed.

Approach

Researchers have implemented a variety of approaches toward testing the significance of portfolio optimization frameworks. Chen et al.(2006) surveyed optimization methods on the Nasdaq-100 and NIFTY. Cura (2008) tested optimization methodologies on weekly prices across the Hang Seng, Dax 100, FTSE 100, and S&P 100. Ping et al.(2016) surveyed optimization techniques on Chinese small caps. Methodologies varied from Conditional VaR, Mean Absolute Deviation, Particle Swarm Optimization and Flexible Neural Trees Ensembles.

The research study conducted in this text will simulate the role of an asset manager. Key differentiating factors of this study are the assumed role of an asset manager, multiple portfolio constructions and the application of ensemble methods not to portfolio composition, but directly to components of the microstructure of which is dictated by the respective strategies. The primary microstructure component within this analysis is volatility. Four portfolios will be constructed with an initial capital value of $100 million USD. The portfolios were initially intended to be comprised over four strategy paradigms, 1) Statistical Arbitrage, 2) Microstructure, 3) Trend Following, and 4) Options Volatility Dispersion. But due to data and resource constraints, was adjusted to the Statistical Arbitrage and Trend Following strategies. A minimum cash constraint will be applied to each portfolio. Seminal optimization techniques will be implemented including the Efficient

Frontier over some portfolios while others may only feature machine learning optimization techniques. A new modeling framework that employs the use of ensemble learning will be applied to the market microstructure of each respective strategy within the machine learning based portfolios. The ensemble learning model will be configured based on the respective strategy regime and have the ability to either augment or offset strategy positions. The initial portfolio will be that of Portfolio A1 and will equally weigh capital across each strategy paradigm. Portfolio A2 will seek to employ seminal optimization methods but will not employ the use of machine learning or the ensemble framework on the microstructure of the strategy. Portfolio B will consist of the optimal portfolio created, Portfolio A2, but will add the machine learning ensemble model to the market microstructure of the respective strategies, thus creating the Stereoscopic Portfolio Optimization (SPO) framework. Lastly, Portfolio C will take Portfolio A1, or the pre-optimized or equally weighted portfolio and apply the ensemble learning market microstructure model to it. Each portfolio's performance will be assessed to determine the impact of applying machine learning ensembles directly to the microstructure of the strategies in addition to the stereoscopic implementation, impact on portfolio optimization.

Statistical Arbitrage

Statistical Arbitrage is a strategy that seeks to capitalize on the mean reverting nature of the statistical properties between two or more securities. Securities' prices are said to move in a random walk. A general idea of statistical arbitrage is that of stationarity. Stationarity is the degree to which a time series' statistical properties are stable. Most financial time series do no display stationarity. The Statistical Arbitrage methodology tests for the stationarity between two or more securities. In order to apply a mean reverting strategy, the time series must be stationary in nature. While individual financial time series tend not to display this characteristic, empirical research showed that a stationary time series can be created using securities that are fundamentally related. The stationarity of the difference between two separate securities is termed cointegration. Often the Augmented Dickey Fuller test is used to test for cointegration. If the difference or spread between two securities is stationary, a mean reverting logic can be applied by buying one security and selling the other based on the value of the z-score. The z-score is calculated by subtracting the value of the spread, ratio, or difference by the mean and dividing by the standard deviation. This in essence shows how many deviations away from the mean the time series is. When the z-score exceeds a predefined limit to the up or down side, the spread can be bought or sold in anticipation of the spread reverting back to the mean.

Driaunys, Masteika, Sakalauskas and Vaitonis (2014) applied a statistical arbitrage methodology to natural gas futures on a high frequency trading time frame. The researchers chose natural gas futures of different expirations date as they should maintain a fundamental relationship. The trading volume in natural gas futures also offered a liquid trading environment. Driaunys et al.(2014) defined the trading methodology into five steps: 1) High frequency data normalization, 2) Selection of correlated pair, 3) Defining a moving window for trading and data normalization, 4) Setting triggers for a long/short positions, and 5) Performance assessment of the trading system. "An algorithmic trading system needs HFT data to be normalized. The main problem with HFT data arises due to the discrepancies between time stamps of correlated contracts." (Driaunys et al. 2014) The researchers sought the most liquid and close to expiration contracts to test. The study covered the October 2012 and November 2012 natural gas contracts. Data from the NYNMEX for dates 2012-09-04 to 2012-09-20 were used. Timestamps were in milliseconds and approximately 350,000 normalized records were produced per day. The experiment used bid prices of one contract and ask prices of the other. Market orders were also used as a mechanism of avoiding slippage. Driaunys et al. (2014) researcher the most active trading period during the day upon which to execute the HFT statistical arbitrage strategy. The researchers calculated the average number of trades per every 15 minute window. They found that the most active time frame for conducting the HFT statistical arbitrage strategy was 11:15:00 and 18:45:00. No transaction costs assumptions were made. The researchers found that a moving window of 100, signal threshold of 5, and maximum period for keeping an open position of 1000 yielded some negative results on a daily basis, namely two days out of the research period. The researchers then shortened the moving window to 50 from 100 with the remaining parameters constant. This improvement yielded a single day of negative returns of -0.05. The researchers then altered the parameters once more to a moving window of 50, signal threshold of 6.5 and maximum holding period of 1000. This configuration yielded all positive daily results. The researchers found that the best results were obtained using lower values for the moving window and data normalization parameters along with higher values for the signal generation coefficient. The research also revealed that less profitable trading days occur due to fewer trades instead of wide fluctuations or a series of unprofitable trades.

Qazi, Rahman, and Gul (2015) applied the Engle-Granger 2-step Cointegration approach to commercial bank and financial services stocks listed on the Karachi Stock Exchange. This approach was offered as a methodology for selecting pairs of stocks to trade. The time period surveyed spanned November 2, 2009 to June 28, 2013. Daily prices of stocks retrieved from the Business Recorder were used over the sample period. The stocks were filtered by their liquidity. Pair's composition was also restricted based on fundamentals. It was believed that restricting stock pairs to those with similar fundamentals (i.e. same sector) further supported cointegration. Based on this filter 22 of the listed commercial banks and 19 financial services companies of the 23 listed commercial banks and 40 listed financial services companies were selected. Of the 231 pairs of commercial banks, 25 were found to be cointegrated and 40 of the 156 pairs within the financial sector were found to be cointegrated. Cointegration was tested by regressing one price series vs. another. The order of regression was accessed through the Granger Causality Test. "A uni-directional Granger Causality test describes which stock informationally leads another stock in a trading pair."(Qazi 2015) The researchers used the Vector Error Correction Model to model the residuals. "The residual series contains significant information pertaining to co-movement between the trading pairs. For instance the 'speed of adjustment' coefficients in the VECM describe how quickly the system reverts to its mean after observing a short-term deviation and also identifies which stock in a pair performs the error correction function."(Qazi et al. 2015)

In the first step of the EG approach the Augmented Dickey Fuller Test(ADF) is applied to the log prices and the lag length is determined using Aikaike's Information Criterion(AIC). If a log price series is stationary based on the results of the ADF, it is excluded from the analysis. This is done because one would be cointegrating a stationary and non-stationary price series and this could result in spurious regression and non-stationary residual series. The second step is to test the stationarity of the residual series. In this step, the ADF is used to verify the stationarity of the residuals. "According to the EG approach, the estimated residual series has to be stationary for the Xt and Yt to be cointegrated."(Qazi et al. 2015)

The Granger Causality Test is employed to address the issue of ordering of pairs. It has also been found to identify which stock leads or lags the other. The following is the Granger Causality Test presented in the study:

Granger causality test under bivariate (x, y) setting can be expressed as under,

$$y_t = \beta_o + \beta_1 y_{t-1} + \cdots + \beta_i y_{t-i} + \alpha_1 x_{t-1} + \cdots + \alpha_i x_{t-i} + e_t \quad \ldots \quad \ldots \quad (3.3)$$

$$x_t = \beta_o + \beta_1 x_{t-1} + \cdots + \beta_i x_{t-i} + \alpha_1 y_{t-1} + \cdots + \alpha_i y_{t-i} + e_t \quad \ldots \quad \ldots \quad (3.4)$$

The first test whether or not the x causes the y and the second test determines whether the y causes the x. "if the first null hypothesis is rejected and the second is accepted, it can be inferred that x granger causes y indicating uni-directional causality from x to y. This also depicts that x informationally leads y."(Qazi et al. 2015[Green (2002)]). If both are negated there exists a bi-directional causality. If both are accepted there is no evidence in support of causality between the two variables.

"According to the Granger Representation Theorem, when the two time series are cointgrated, the Vector Autoregressive model (VAR) is mis-specified. The mis-specification problem can be treated through incorporating the previous disequlibrium term in the VAR model as an explanatory variable and thus the model becomes well-specified and is termed as Vector Error Corrections model (VECM). (Qazl et al. 2015) This model allows the researcher to model a time series as a function of its own lags, lags of the cointegrated pair, and the error correction component.

Huang, C.F.et al. (2015) applied the use of genetic algorithms to the problems of selecting pairs for trading. Traditional approaches to statistical arbitrage employ the use of cointegration, Kalman filter, and or principle component analysis. Genetic algorithms, within the realm of computational intelligence, are used to address two sets of key problems. First, those involving stock selection, optimization, and portfolio management. Secondly, those encompassing time series prediction. Huang, C. F. et al.(2015) proposed a generalized approach to statistical arbitrage that used more than two stocks as a trading group toward the end of improving the performance of the model. Genetic algorithms were employed to optimize various parameters of the statistical arbitrage strategy. The goal of the study was to show that the GA models outperform traditional statistical models for stock selection and that the strategy itself would outperform the benchmark. Historically, pair selection has been a process of matching pairs with similar characteristics, usually fundamentals such as inclusion within the same sector. Huang, C.F. et al.(2015) showed that their model is capable of matching pairs with different traits. The researchers used the Bollinger Bands to determine if the spread of a pair of stocks deviates from its moving average. Signals were generated when the pair moved to x standard deviations from the mean and performance was based on compound returns. The researchers composed to models, market timing and pairs trading. The parameters for the market timing models included the period for the moving average, x and y for the Bollinger Bands of which determines the sigma or deviation of the MA for entries and exits. The parameters for the pairs trading model were the weights of the terms or stocks. Genetic algorithms were used to optimize all parameters across both models. The weighting terms determine which stocks to be long or short. The composition of the chromosome was made up of four parts that encoded the period parameter for the MA, the x and y values for the Bollinger Bands, and the weights for the pairs trading model. “The final output is a set of models parameters (optimized by the GA) that prescribes the pair trading and timing models.”(Huang, C.F. et al.(2015)). Two sets of Taiwanese stocks were used. The first were 10 stocks from the semiconductor industry with similar characteristics. This industry has been the most important for Taiwan over the last 20 years. The second set of stocks was ten stocks with the largest market caps across various sectors. The testing period was from 2003 to 2012. The benchmark was composed of a buy-and-hold strategy of the stocks that distributed capital equally across all ten. “In order to study the quality of solutions over time, a traditional performance metric for the GA is the “best-so-far” curve that plots the fitness of the best individual that has been seen thus far by generation n.”(Huang C. F. et al.(2015)). The GA proactively searches for best long and short components to build the most optimal spread. It also searches for the optimal timing for buying and selling the spread dynamically based on the Bollinger Bands. Significance was placed on temporal validation versus standard cross validation. In the training phase of each temporal validation, 50 runs were conducted with the best examined in the testing phase. The annualized returns of the best 50 models in each temporal validation were averaged. The annualized benchmark return was also computed for comparison with that of the GA. Of the 39 temporal values of which occurred over 39 quarters, the GA's annualized return outperformed in all in the training phase and in 30 of 39 during the testing phase. Model robustness was also tested as the percentage of true positives to total positives. A true positive consisted of the model outperforming during both the training and testing phases. A false positive occurred when the results of the training and testing phases were not consistent. Based on this precision metric, the set of 10 semiconductor stocks achieved a precision of 77% and the precision of the 10 largest stocks by market cap was 72%.

Chaudhuri, Ghosh, and Singh(2017) employed the use of three different models, namely Adaptive Neuro Fuzzy Inference System (ANFIS), Random Forests (RF), and Support Vector Machines (SVM) to predict the value of the ratio of share prices of pairs of companies. Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) were used as a means of performance evaluation. The parameters that were used to construct the model were lagged values of the price ratio, ratio momentum of the two share prices, deviation of the ratio of prices from the mean and lagged ceiling and floor values of the price ratio. The study evaluated 3 pairs of Indian equities on a daily time frame over the period of 2012 to 2015. The performance was assessed on a year-wise basis. The researchers also surveyed the degree of influence each input variable had on the model. A separate framework adopting Hall's (1998) correlation based feature selection process. "We have applied Genetic Algorithm (GA) in Hall's (1998) framework for intelligent search space transversal in feature selection process to identify the factors which are most important for forecasting. This exercise has been carried out, year-wise and pair-wise, to ascertain the robustness of the input variables in prediction." (Chaudhuri et al. (2017)) The choice of Support Vector Regression, Artificial Neural Network, and Adaptive Neuro Fuzzy Inference System for prediction, non-parametric models, affords the researcher the capability of extracting non-linear patterns from the independent and dependent variables. The precise inputs used were ratio of prices in time t-1, t-2, t-3, ratio of momentum of prices with lag 1, lag 2, lag 3, Pair Prices-9day mean, and two sigma ceiling, two sigma floor on a 20day moving window. Momentum was used as an explanatory variable for the statistical arbitrage methodology. 80% of the data was used for training, while 20% was used for testing. For the Random Forest (RF) model, a minimum of 500 trees and a minimum of 3 values sampled at each split. 20 trials were conducted. The Support Vector Regression model used radial basis kernel function with the associated 3 parameters. 20 trials was conducted using this model as well. For the ANFIS, there were 9 input nodes. Layer 1 had 27 member functions, Layer 2 had 81 nodes, Layer 3 had 81 nodes, Layer 4 had

81 rules, Layer 5 was the output layer with one output. A Gaussian technique was used for the member function. The membership function parameters were varied to get 20 experimental trials. The study showed that each method is efficient in predicting the ratio of pairs but it was also noted that the low MAPE and MSE were made possible by the selection of inputs. “A good feature is identified by its correlation with dependent and independent variables”(Chaudhuri et al.(2017)) Hall’s 1998 correlation based feature subset selection approach was used. A feature that’s highly correlated with prediction but lowly correlated with each other is regarded as a good feature subset.

The following are the results for the SVR model:

The average MAPE of the first pair for the training set was 0.0003675. The average MSE over the training set for the first pair was 0.0002525. The average MAPE over the testing set for the first pair was 0.0006115. The average MSE for the testing set for the first pair was 0.000355. The average MAPE over the training set for the second pair was 0.0008. The average MSE for the training set for the second pair was 0.0005325. The average MAPE of the second pair over the testing set was 0.0009275. The average MSE of the second pair over the testing set was 0.0007925. The average MAPE over the training set for the third pair was 0.0004725. The average MSE over the training set for the third pair was 0.00049. The average MAPE of the third pair over the testing set was 0.000655. The average MSE of third pair over the testing set was 0.0007025.

The following are the results from the RF model:

The study only included the results from the first pair but results for the remaining pairs were made available upon request. The training set of the first pair had an average MAPE of 0.0003225. The first pair over the training set had an average MSE of 0.0002675. The first pair over the testing set had an average MAPE of 0.0008075. The average MSE of the first pair over the testing set was 0.00037.

The following are the results from the ANFIS model:

The study also made the results of the first pair of the ANFIS model readily available but the results of the remaining pairs were made available upon request. The average MAPE of the first pair of the training set was 0.00053. The average MSE of the training set of the first pair was 0.00034. The average MAPE of the testing set for the first pair was 0.0007825. The average MSE of the testing set for the first pair was 0.00044.

“Instead of enumerating all possible combinations using exhaustive search technique, this study uses GA to perform intelligent searching operation.” (Chaudhuri et al. (2017)) It was found that the distance from mean, the ceilings and floors were the most important features in predicting the pair ratio. The momentum feature was also important for predicting the pair ratio of one pair.

In this study, Statistical Arbitrage is conducted on US equities. The object of my research is to test the optimality of employing the use of machine learning algorithms within the context of each individual strategy as a means of achieving portfolio optimization. In lieu of this, a Statistical Arbitrage Strategy will be developed for separate portfolios. K-Means clustering is employed across all portfolios to ensure that each portfolio only differs in the optimization approach employed. The SPO model will have the ability to augment the signal generator or prevent signals from being transmitted within the machine learning based portfolios. The greatest risk to a statistical arbitrage strategy is the breakdown of cointegration, thus this framework featured in the machine learning based portfolios will have the ability to detect the likelihood of a failure in cointegration through the analysis and training on historical strategy regimes

Trend Following

Fong, Si, and Tai (2012) proposed five alternative approaches to implementing a trend following strategy. It was stated that determining the correct strategy at the beginning is crucial to trend following. A key focus of the work of Fong et al. (2012) was the presentation of a new version of the standard trend following methodology, namely the trend recalling model. The trend recalling model matches a portion of the current trend with a proven pattern from the past. The study showed that the trend recalling model had edge profitability wise over the other methods. The study was categorized into prediction and reactionary approaches. Trend following in general does not predict market movements or forecast prices. It was mentioned that a key concerns of implementing trend following are 1) indentifying the trend and 2) accounting for volatility.

The researchers experimented on the Hang Seng Index futures. Trade was performed on a daily basis to limit overnight risk and cost of carry. The profit and loss of each trade was aggregated and reported as ROI. The five methodologies implemented were the 1) static trend following approach, 2) the dynamic trend following approach, 3) the fuzzy trend following algorithm, 4) the fuzzy trend following algorithm with volatility approach and 5) the trend following algorithm with trend recalling approach. "The basic practice of trend following is to find the trend, identify trade signals and trade along with them."(Fong et al. (2012))

The static trend following algorithm has two constant parameters, the buy and sell thresholds. The EMA is used to smooth out the frequency of the trend's fluctuation. The buy and sell threshold parameters are found by studying historical data, and seeking the optimal parameters. In the dynamic trend following algorithm the buy and sell thresholds are variables instead of constants. The values are based on market trends. The RSI and EMA were used. The following are the conditions used in the study for the dynamic trend following algorithm: 1) price advancing, 2) RSI(t) is greater than RSI(EMA(t)) and RSI(EMA(t)) is less than 40 or greater than low for longs. For shorts the inverse was true. The third model offered was the fuzzy trend following algorithm. Three member functions were used. The RSI and MTM indicators were inputs and the output was the position. Twenty-seven inference rules can be implemented from the membership functions. Five of the possible membership functions were selected. The following are the five which were selected: 1) If the RSI is whipsaw or MTM is whipsaw, then POS (i.e. position) is do nothing, 2) if RSI is oversold and MTM is long then POS is go long, 3) if RSI is overbought and MTM is short then POS is go short, 4)if RSI is oversold and MTM is short then POS is go short and 5) if RSI is overbought and MTM is long then POS is go long. Whipsaw is defined as a price advance followed by a sudden reversal. The fourth model, the fuzzy trend following algorithm with volatility parameter, was implemented as a means to understand how volatility affects the trend following strategy. "To find out how market fluctuation can affect our algorithm, we attempt to simulate the market with a controllable fluctuation. We then apply our trading algorithms on it, and observe how it performs under certain degree of fluctuation."(Fong et al. 2012) The following formula was provided:

$$\textbf{MarketData(t)=(COS(e) x C x R(t)) + B;}$$

where COS is the geometry cosine which creates a basic oscillating wave structure, e is the angle that controls the fluctuation frequency, C is a constant that controls the fluctuation depth, R is a random number that creates the Whipsaw sharp, and B is a fixed base price that the generated prices will fluctuate along with. The idea is that this method will allow for the creation of market data that gradually becomes more volatile over time. This data is then fed into a simulated engine. "The central concept of volatility is finding the amount of change in the price of an asset over a period of time. Therefore we can measure it simply by the following formula:

$$\textbf{Volatility(t)=}SMA_n\ \textbf{((ln(price(t))-ln(price(t-1))) + C)}$$

Where ln is the natural logarithm, n is the number of periods, t is current time, C is a constant that enlarge the digit to significant figure." (Fong et al. 2012)

It was found that when the volatility of this simulated data rose over 50% losses occurred. Thus, being cognizant of the volatility environment can aid in validating when and when not to trade.

The final approach to the trend following strategy was to implement a trend recalling functionality. The researchers found that while cycles could not be predicted, patterns of such cycles were predictable. This approach finds trading opportunities by searching for similar patterns in the present to that of the past. The following five questions were posed as being the rules for designing the trend recalling algorithm: 1) How does the system determine what market to buy or sell at any time? 2)How does the system determine how much of a market to buy or sell at any time?, 3) How does the system determine when you buy or sell a market?, 4)How does the system determine when you get out of a losing position?, and 5) How does the system determine when you get out of a winning position? The trend recalling approach addresses the key concern of when to buy and sell. The four major processes of this approach are Preprocessing, Selection, Verification, and Analysis. It can also be used to match patterns from a once profitable trading system to the current market environment to determine its current viability.

The five approaches were tested on 2.5 years of data prior to 2010 in the Hang Seng Futures market. Researchers found that the trend recalling algorithm achieved the best performance.

Clare, Seaton, Smith and Thomas (2013) posed some interesting questions in their study on trend following. Key to these was: 1) How frequent should investment decisions be made? , 2) How useful are stop losses? , and 3) How do simple moving average rules compare to fundamental value metrics? Their analysis was conducted on the S&P500. They found that 1) there is no advantage in trading daily rather than monthly, 2) there is no value in stop-loss rules, 3) whipsawing is not a problem provided the technical signals are of reasonable length (i.e. not too short), 4) there is no advantage in complicated trend-following rules versus simple rules, and 5) trend-following rules give superior risk-adjusted returns relative to using fundamental financial metrics.

Clare et al. (2013) considered three types of trend following rules: 1) simple daily moving averages, where the buy signal occurs when the S&P 500's index value moves above the average; moving averages ranged from 10 to 450days, 2) moving average crossovers where the buy signal occurs when the shorter duration MA crosses the longer duration MA; this ranged from 25/50 days through 150/350 days, and 3) breakout rules, which indicate signals based on the index trading at an 'x' day high; x ranged from 10 to 450 days. If there were no trading signals, the portfolio was invested in T-Bills over the relevant period. Thus, a return was either earned on the S&P500 or T-Bills. A comparison was made between daily and monthly period performances. The simple daily MA rule, assumed a 20 basis point transaction cost; the highest Sharpe of this approach, 0.54, was produced by the 400-day MA, with a return of 10.5 percent a year. A buy and hold approach over the same period yielded a return of 9.49% and a Sharpe of 0.31. The most profitable monthly MA rule was the 200-day rule which yielded a returns of 10.66 and Sharpe of 0.58. The daily and monthly trend following methods were implemented on the S&P 500 from July 1988 to June 2011. Relative to the MA crossover approach, the strategy consistently outperformed the buy and hold approach, and yielded a return of 10.88 percent with a Sharpe of 0.56 when applying the 150/300-day crossover. The researchers compared this performance to a monthly MA crossover approach using the 100/250 day-crossover and found it to be better. The best breakout implementations were the 200 and 250 days breakouts which yield 11.38 and 11.59 percent and Sharps of 0.61 and 0.62 respectively. The study also showed that the simple 200-day moving average applied at the end of month was successful as any other trading rule in terms of the average return and Sharpe ratio.

The study measured the success of stop-loss rules by assessing how they impacted returns. It was stated that if a strategy's returns follows a random walk that stop-loss rules could reduce the expected return, however, if a strategy was found to exhibit momentum in returns, stop-loss rules could add value. The researchers stated that if a returns' process was mean reverting, stop losses may fail; this occurs when the stop-loss criteria are triggered; stopping the trader out, and then the asset recovers and performs as initially expected. "It is clear, and indeed intuitively appealing, that the premium from applying a stop-loss rule is closely related to the stochastic process underlying the portfolio's return and in fact is directly proportional to the magnitude of return persistence."(Clare et al. 2013) Research has shown that stop-losses are effective for assets with high volatility. It was stated that investors may misconstrue stop-loss strategies as improving investment returns when in reality their true value lies in reducing investment risk. The researchers assessed the use of stop-losses on different strategies. The first was a breakout strategy that generated entry signals when the price exceeded some threshold and triggered a stop –loss when the price declined by some threshold. Another approach was to identify the effect of trailing stop-losses. The 200-day moving average was used as a breakout entry signal and stops were placed from that entry over a range of 3-15percent. Another stop-loss rule that was mentioned was one that exited positions if the returns fell below five sigma of the initial purchase price. As a result of the research, it was stated that the best stop-loss rule is a change in trend.

Two significant concerns with implementing a trend following strategy is the identification of the trend and coping with the volatility environment. Volatility directly influences the strategy because it could increase the probability of being stopped out of a multiplicity of trades in hopes of catching a much larger move. It is my belief that this strategy could be optimized by developing a methodology to predict the probability of being stopped out. Based on this model, the trader could then deleverage his/her position, of which would allow the stop to be temporarily extended (i.e. because a smaller position size with a wider stop would equate to a larger position with a narrower stop) or depending upon the response of the model, a trade could be exited before being stopped out. It seems practical that over time this could, through the cost savings of exiting losing trades prematurely in comparison to the traditional implementation of the strategy, losses could be impeded and thus the breakeven point for profitability could be reduced. This, over time, should increase the overall profitability of the strategy and also have a positive effect on the optimization of the portfolio. To test this notion, a trend following strategy will be developed for the traditional portfolios that will lack the notion of a machine learning ensemble based model on the microstructure. In contrast, within the machine learning focused portfolios, this strategy will feature the model mentioned with the aim of uncovering regimes in which the strategy has performed poorly and augmenting the signal generator at time t based on regimes predicted at some lag of t.

Microstructure: Volatility

Kumar, Das, and Govil (2015) analyzed stock volatility using an ANN as multilayer perceptron with back propagation model and radial basis function. These model's utilities was compared to that of the GARCH(1,1) model. The study applied these models to the S&P 500, Nikkei, BSE 30, and Hang Seng. The research focused on weekly prices over varying periods with respect to the indices. The dates used were as follows: 1) S&P 500 from January 3, 1950, 2) Nikkei 225 from January 4, 1984, 3) Hang Seng from August 8, 2000, and 4)BSE 30 from July 10, 2000 to August 31, 2010. All data was retrieved from Yahoo Finance. The compounded weekly returns were calculated as the logs of relative weekly prices. After calculating returns and standard deviations, an attribute selection algorithm was applied that removed unnecessary attributes from the data and the improved model was applied for forecasting. The following errors were calculated from the forecasts over a period of 30 weeks: 1) Mean Absolute Error (MAE), 2) Root Mean Square Error (RMSE), 3)Mean Absolute Percentage Error (MAPE), and 4) Mean Square Error (MSE). The errors showed the relationship of volatility to mature versus emerging markets. Volatility tends to be more pronounced in emerging markets in comparison to developed markets.

Santamaria-Bonfil, Frausto-Solis, and Vazquez-Rodarte (2015) introduced the SVR_{GBC} model to the volatility forecasting problem. The model selects the proper kernel and its respective parameters, a problem faced when implementing standard Support Vector Regression. "The SVR_{GBC} is a hybrid genetic algorithm which uses several genetic operators to enhance the exploration of solutions space: it introduces a new genetic operator called Boltzmann selection, and the used of several random number generators."(Santamaria-Bonfil et al. 2015)The results of the SVR_{GBC} were compared against the GARCH(1,1) tuned by maximum likelihood estimation (MLE) and SVR_{GS} (i.e. SVR tuned by grid search method) models. The mean absolute percentage error (MAPE) and directional accuracy functions were used for measuring the quality of results. The models were applied across four stock market indexes from ASEAN region and Latin

American countries, over the 2007 and 2008 periods to test the methods under heightened stress. The GA operators used by the SVR_{GBC} yielded 36 different configurations. The SVR models in general outperformed the GARCH model in all but one case. It was also found that the Boltzmann selection and chaotic random number generator led to better results of the SVR in most cases compared to the classical operators. A comparison was also made between the GS and GA algorithms. The Wilcoxon test, a non-parametric test, was used to evaluate the two algorithms. First the difference between the GS and GA results are computed, then using the Wilcoxon test, a comparison of whether the median of the differences differs from zero. The null hypothesis is that the GS and GA yielded statistically similar results. The alternative is that results for the MAPE are lower and the results for the DA are greater with respect to the GA. The study offered the GA as the most optimal of the two methods.

---

2. Data

End of day data was retrieved from Quandl. The sample period ranged from 1st January 2012 to 1st January 2017. Relative to the selection of pairs for the Statistical Arbitrage strategy, the Augmented Dickey-Fuller test was conducted over the period of 1st January 2011 to 31st January 2011. The Statistical Arbitrage Strategy encompassed the use of stocks from the US S&P 500. The Efficient Frontier for Portfolios A2 and B was derived over the 2015-2016 period. The Trend Following Strategy was employed on the Australian Dollar, the Canadian Dollar, the Japanese Yen, the British Pound, and the Euro currency futures.

The decision to use end of day data for this study was congruent with the theme of simulating the role of a portfolio manager. Scaling in and out of positions is often conducted over a period of several days if not longer, thus the aforementioned strategies were designed with this in mind.

3. Methodology

Statistical Arbitrage

K-Means was employed for the selection of tradeable relationships across all portfolios. Twelve features were used for stocks within the S&P 500.The equation used for the clusters can be seen via the function below:

$$f(x) = \begin{cases} \frac{x}{2} & if\, x\%2=0 \\ \frac{x-1}{2} & \text{if } x\%2!=0 \end{cases}$$

Where x is the number of observations. The above equation ensures that k is always an integer and never a float, because the latter intuitively cannot exist and would throw an error in the algorithm. The length of the S&P dataframe was 505. Some possible further research may be warranted regarding the affect of integers in comparison to float values for the twelve features. Completing K-Means using integers values for the features yielded more clusters than that of the floating point features. Whether this was merely coincidental is possible but is worth looking deeper into. The pairs selected resulted from the use of float values as the features. 70 clusters were created that contained more than a single symbol. Clusters containing at least two symbols were then sliced from the data for further analysis. Of the 70 clusters, the symbols per cluster ranged from 2 to 7. A random number generator was then built that randomly selected 5 clusters from the 70. No cross cluster testing was conducted and the selection of 5 clusters per iteration was merely a means to expedite the process of identifying tradeable relationships. Each combination of the symbols, within each cluster, over each iteration of the random selection process, was then tested for cointegration using the Augmented Dickey-Fuller test. This random selection process was repeated until arriving at 8 tradeable relationships. The Augmented Dickey Fuller test was conducted on dates from 1st January 2011 to 31st January 2011. These dates, one year prior to the sample period dates, were chosen to avoid look-ahead bias of whether or not the pairs would remain cointegrated over the sample period. Of the 8 pairs selected from this process, 4 pairs were cointegrated at the 99% confidence interval, 2 pairs were cointegrated at the 95% confidence interval, and 2 pairs were cointegrated at the 90% confidence interval. The R-Squared values for the pairs ranged from .53 to .81. Figure 1 shows the regression plots for the pairs.

The implementations of the strategy for Portfolios A1 (i.e equally weighted) and A2 (i.e. used Efficient Frontier) were the same and followed normal implementation conventions. A lookback period was selected and optimized for each instance of the strategy. The hedge ratio was calculated using the beta of the securities and the price spread was then created. However, within Portfolios B and C, the signal generator was augmented by the machine learning ensemble. The trading strategy was not changed in these implementations, but given a specific regime prediction by the ensemble, the signal generator was then augmented.

A Gaussian Mixture Model was employed to find the regimes of each strategy instance using an 80/20 train-test split. Thus, 1st January 2012 to 1st January 2016 was the training period and the 1st January 2016 to 1st January 2017 was the testing period. The GMM was then trained on the 80% of the original train split and used to make predictions over the 20% of the original train split. These predictions were then used as labels for the Random Forests. Instead of training the ensemble on the original training data and predicting the testing set, another 80/20 train-test split was conducted on the original training data. Thus, the GMM was used to predict the regimes of an internal testing split of original training period of 1st January 2012 to 1st January 2016 and the ensemble, taking this same period, used another 80/20 train-test split. This was done primarily for two reasons, 1) in this study, the test period of 1st January 2016 to 1st January 2017 is figuratively our real market data, or that in which we are actually going to trade. Thus, in an actual trading context we would not have access to this data while training, 2) Once we build the ensemble, we are then able to compare its precision across each regime as well as its total precision. This information can be compared to our analysis of each regime relative to historical returns and volatility, and if necessary we can recalibrate the model before actually using it to augment the signal generator. In short, we do not want to run the risk of potentially contaminating our theoretical live market data.

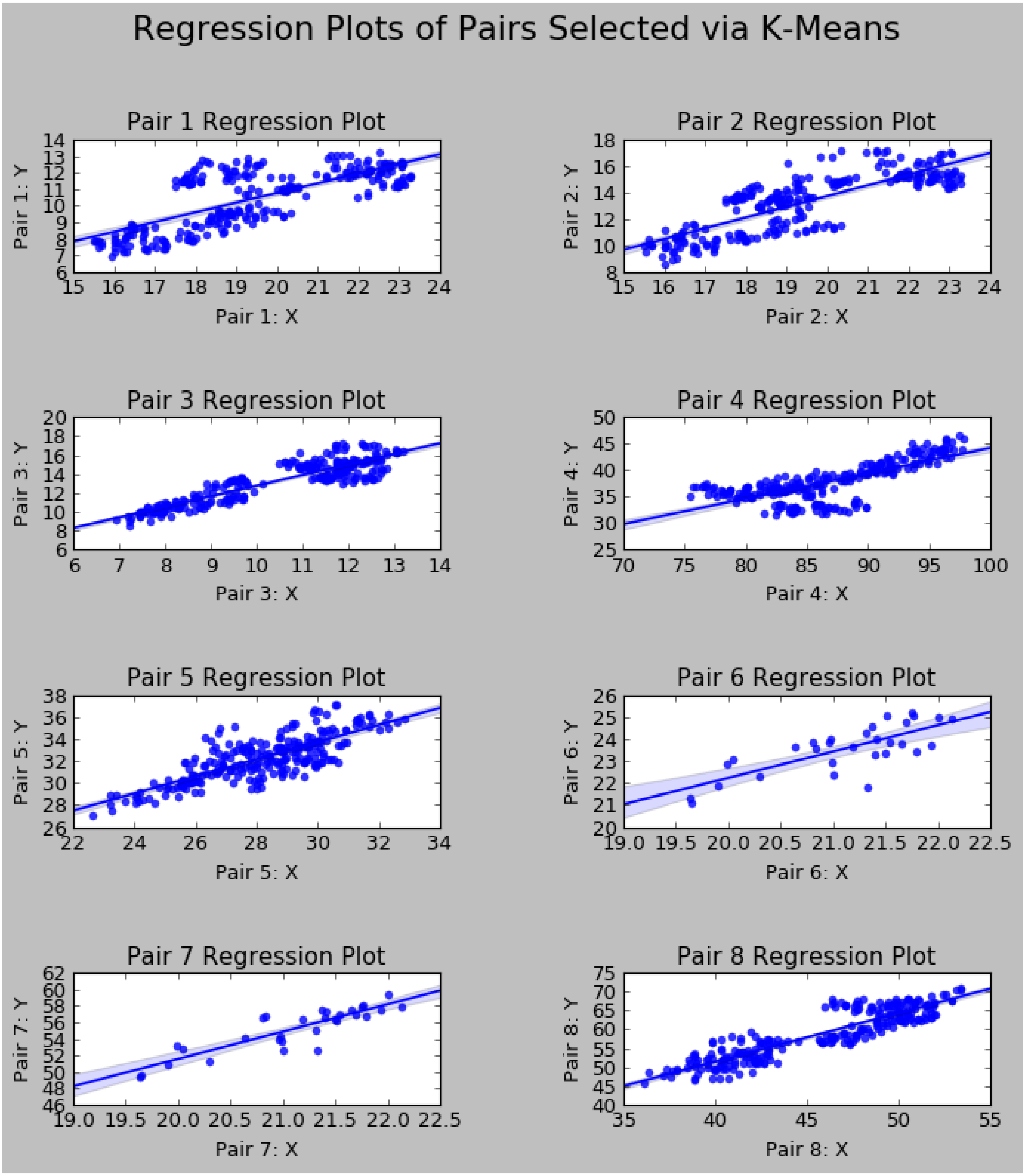
Gaussian Mixture Models are similar to K-Means in that they are both clustering techniques. Yet, they differ in that GMMs implement soft clustering while K-Means implements hard clustering. Identifying K or the number of components is a critical step in the model development process. For the K-Means implementation, due to the aim of pair selection, K was set as $\frac{1}{2} \sum n$. I could have used the “elbow” technique here but felt this value of K was sufficient given the sample desired. However, for the GMMs, a more scientific methodology was necessary. To determine N or the number of components, the

BIC, or Bayesian Information Criterion was calculated and the value that yielded the lowest BIC was used for model development.

After the regimes had been predicted, a comparative analysis was conducted between the regimes and the returns. The idea was to get a better understanding of how each respective regime influenced the historical performance of the strategy. This also allowed for the analysis of the ensemble's precision with respect to a specific regime. It was noted that while the overall precision could be relatively low, the precision of an individual regime could be extremely high. This analysis could then be furthered by looking at the ratio of observations across regimes. A key advantage of this methodology is that a high overall precision is not necessary to positively affect the strategy. Ideally, one would be interested in a subset of the regimes and seek to predict those, not every regime.

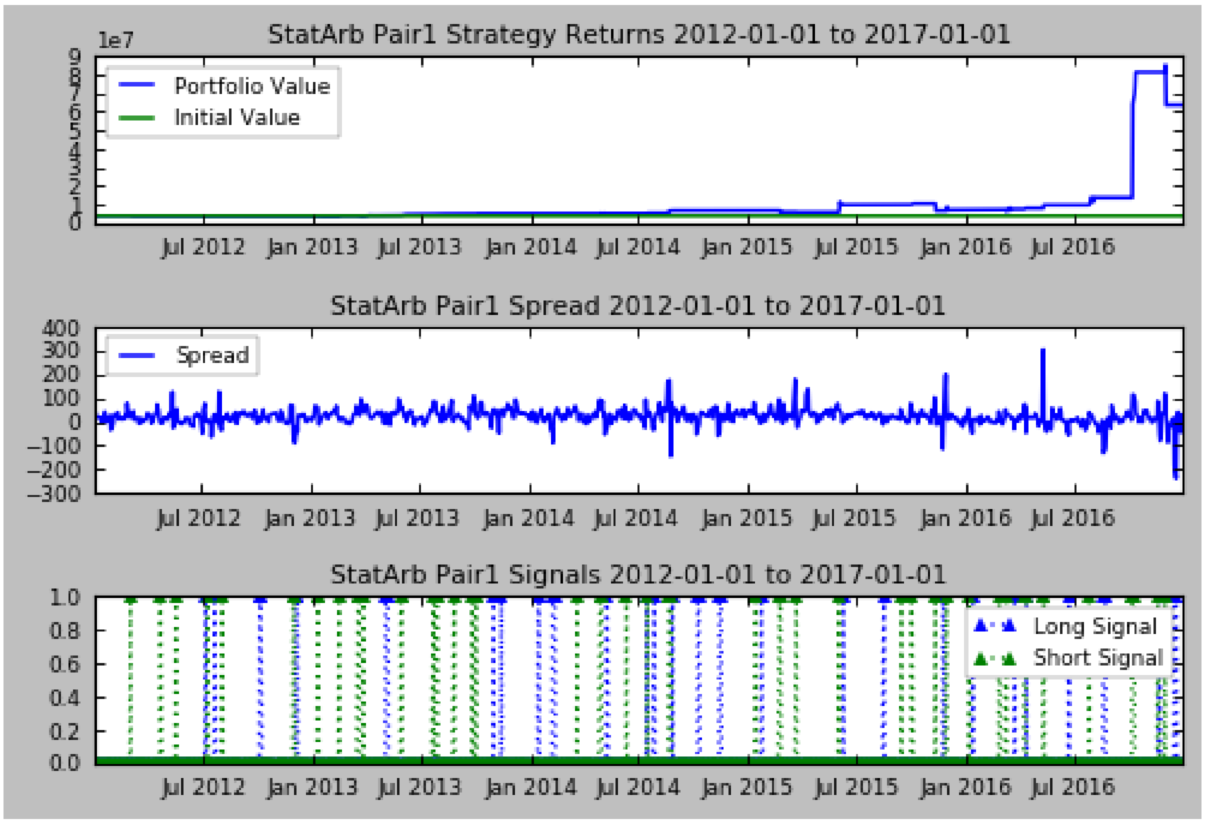
Once the relationship between regimes and that of the strategy was understood, the regimes could then be used as labels for the machine learning ensemble. The Random Forest took multiple features as inputs, several of which were different lags of volatility, in addition to the VIX. These volatility metrics were then used, via the ensemble, to fit a model over the training period of the instance. The model was then used to predict the regimes. Using the inherent relationship between the regime and performance of the strategy, the signal generator was augmented based on the regime predicted by the ensemble.

Figure 1


Regression Plots of Pairs Selected via K-Means
Pair 1 Regression Plot
Pair 1: Y
Pair 1: X
Pair 2 Regression Plot
Pair 2: Y
Pair 2: X
Pair 3 Regression Plot
Pair 3: Y
Pair 3: X
Pair 4 Regression Plot
Pair 4: Y
Pair 4: X
Pair 5 Regression Plot
Pair 5: Y
Pair 5: X
Pair 6 Regression Plot
Pair 6: Y
Pair 6: X
Pair 7 Regression Plot
Pair 7: Y
Pair 7: X
Pair 8 Regression Plot
Pair 8: Y
Pair 8: X

To get a better idea of the strength of the bottom-up optimization process we will briefly review its implementation over one of the pairs created by K-Means clustering. For this analysis we will look at Pair 1 of the Statistical Arbitrage regime. Figure 2 displays the performance of Pair 1 within Portfolio A1 over the entire sample period.

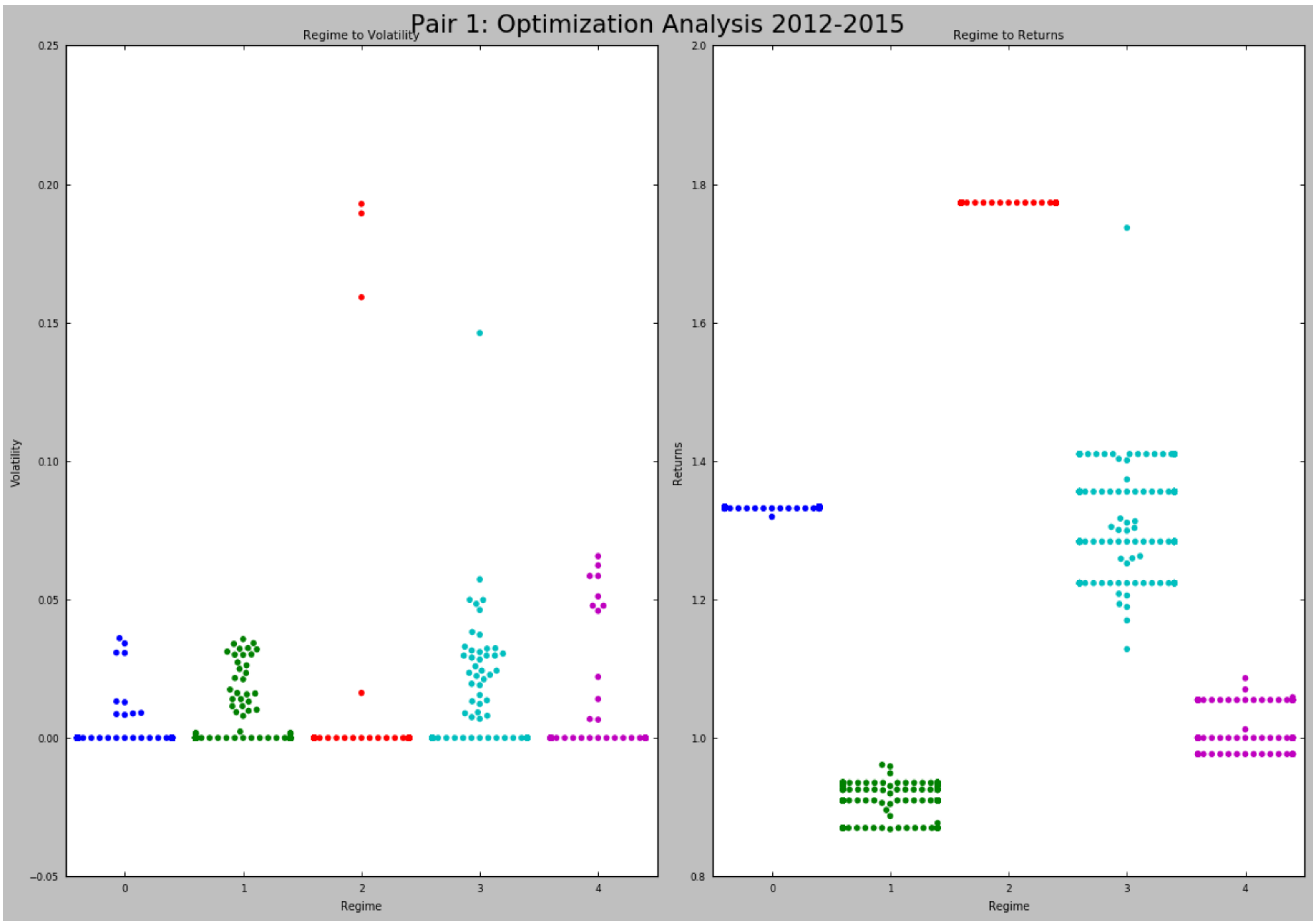
Figure 2



There was a significant spike in volatility in November 2016 which can be seen in the portfolio value of the pair over the same time frame. This was due to the U.S. election. This presents a good analysis period because in a real world trading context, events that have not been priced in by the market can cause volatility to spike, which in turn will affect our portfolio.

After viewing the performance of the strategy over the entire sample period, we then seek to understand the historical regimes. A Gaussian Mixture Model is used for this analysis. The GMM was trained over the 1st Jan 2012 to 1st Jan 2015 period, predicted regimes over the same period and then was used to predict regimes over the 1st Jan 2015 to 1st Jan 2016 period. This internal train-test split was done as a precursor to our Random Forests. Before we can use the Random Forests to predict our internal testing period, or the 2015-2016 period, we must train it over the 2012-2015 period. In order to do this, we must have regime predictions for both 2012-2015, which will be used as y_train, or the training labels in the Random Forests, as well as 2015-2016 regime predictions, or y_test, or the testing labels for the Random Forests. Once we have the predictions, we now must analyze the performance of the strategy historically within each regime as well as the volatility within each regime. The figure below displays the results of this analysis.

Figure 3



We can see how Pair 1 has performed historically within each regime as well as the volatility associated with each regime. This gives us significant insight into augmenting the complex event processing engine to optimize our strategy and thus our portfolio. Before we can train the ensemble, we must now predict the regimes for the 2015-2016 period. Those regimes can be seen in Figure 4.

Figure 4

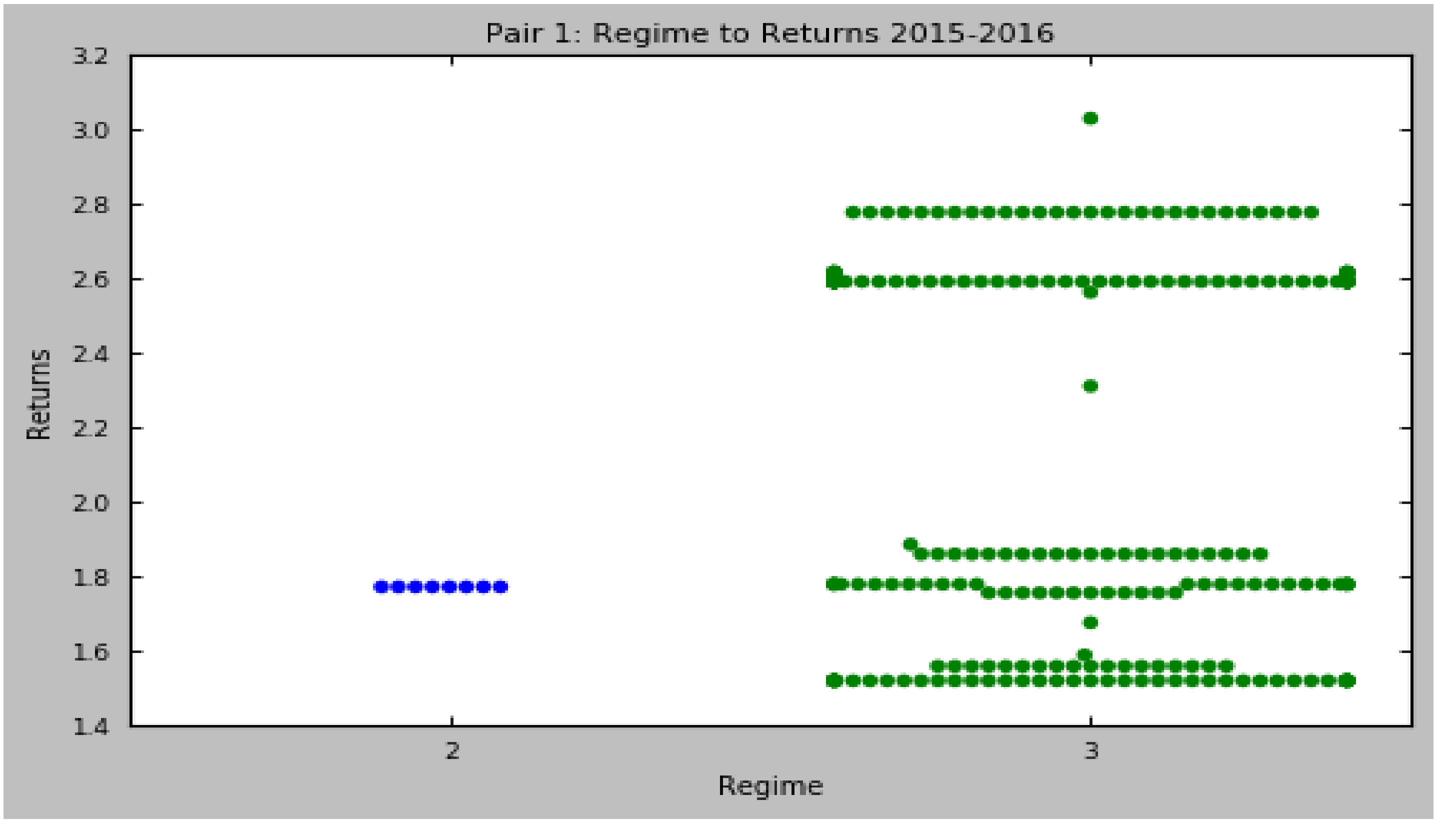


Figure 4 shows us that Pair 1 only traded in Regimes 2 and 3 over 2015-2016. This is plausible because our historical analysis covered 2012 to 2015. The regimes predicted over that period were a composite of each year's regime. Thus, the way that we would interpret our historical analysis is that over the next n periods, we could experience one or more of the historical regimes. It is not likely that we would experience each regime within one interval. Note that our historical period covered four years. The 2015-2016 period, represents 1/4th of our historical period and thus we might expect to experience 1 or 2 of the historical regimes. We can see that Regime 3 dominated the 2015 to 2016 period. We can cross reference this with our analysis on the historical volatility of this regime.

Now that we have our y_test period regimes, we designate the features that will be used in the ensemble to predict the regimes. We first train our ensemble on the 2012-2015 features and labels then use it to predict the 2015-2016 regimes. Afterwards, we create our classification report which compares our model's predictions against the actual regimes over the period. In essence, within this step, we are surveying, given new features, within what regime would our model classify those data points and we compare the model's predictions to what really happened over the period, of which can be seen in the above figure. Figure 5 displays the classification report generated by this Pair's model.

Figure 5

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	0.00	0.00	0.00	8
3	0.99	0.56	0.72	244
avg / total	0.96	0.54	0.69	252

The classification report shows that our model did a good job identifying the regime. To recap, we started by using a GMM to predict regimes over 2012-2015, or our training set, and then 2015-2016, or our testing set. These regimes were then used as labels for our Random Forests ensemble. We selected features that included various lags of volatility as well as the VIX and fitted our model over the 2012-2015 period, then used it to predict the 2015-2016 period. Note, this was completed in one revolution, or in other words without any model recalibration. The classification report showed that our ensemble does a good job at predicting the regimes that Pair 1 trades in based on historical analysis. Had our model been deficient, we then would need to recalibrate it.

Another consideration for the Pair 1 analysis was the presence of negative volatility within the regimes. This was modeled as the standard deviations of negative returns. However, because this implementation had no negative returns, the analysis was null.

At this point we still have no idea of what regime lies ahead in our figurative live market environment over the period of 2016-2017. But, we do have an idea of what the possible regimes and volatilities might be. We use our prior analysis to augment the complex events processing engine. We do not change our strategy in any way with the exception of adding in the ensemble's regime predictions. Stated another way, our strategy is the same within Portfolios B and C as it is in Portfolios A1 and A2 but Portfolios B and C applies the analysis we just covered to the signal generator.

This process was repeated across each pair in the Statistical Arbitrage strategy regime. Below are examples of the Statistical Arbitrage strategy from Portfolio A2(i.e. used Efficient Frontier only) and Portfolio B (i.e SPO framework) over the testing period.

Figure 6

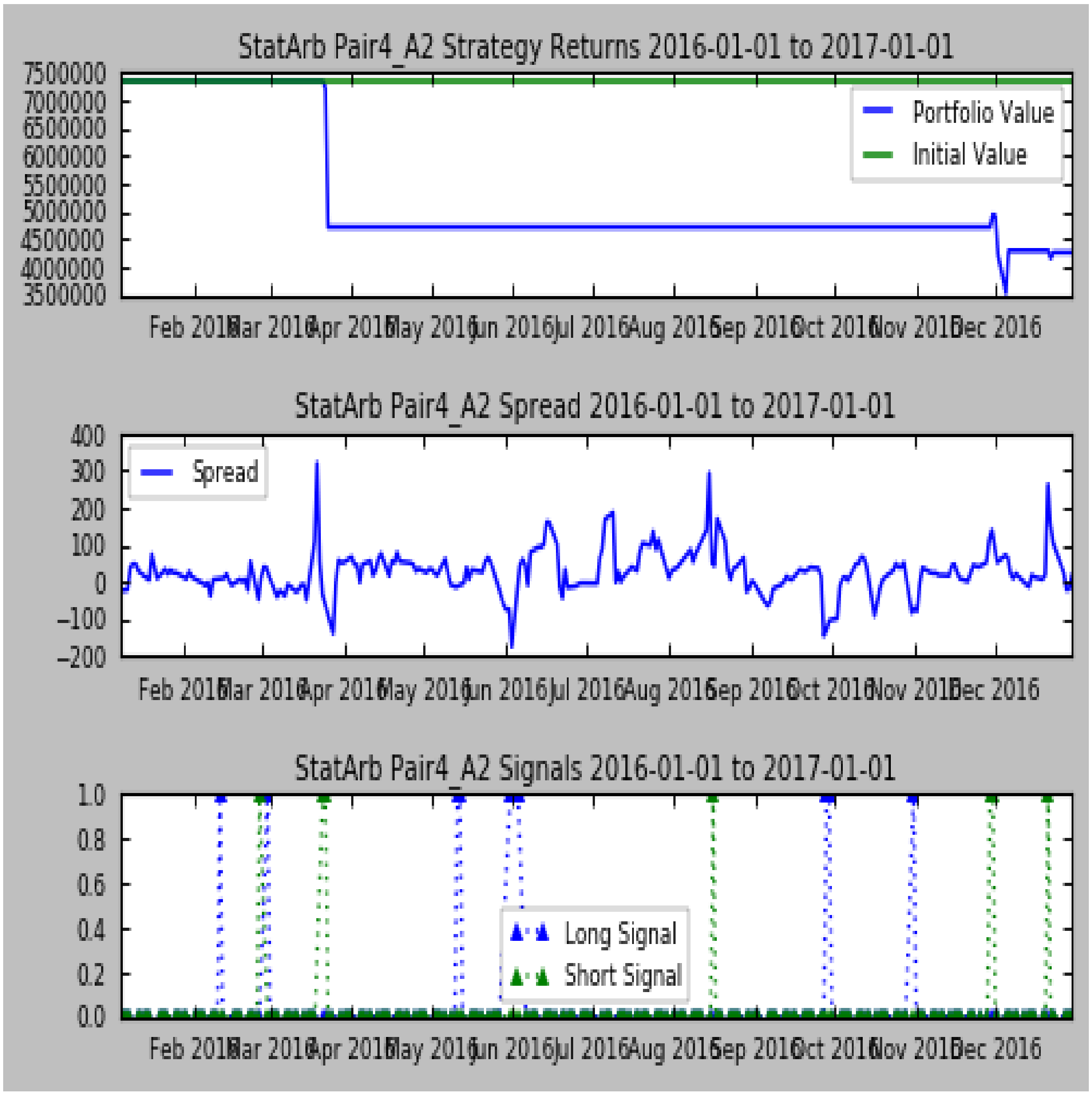
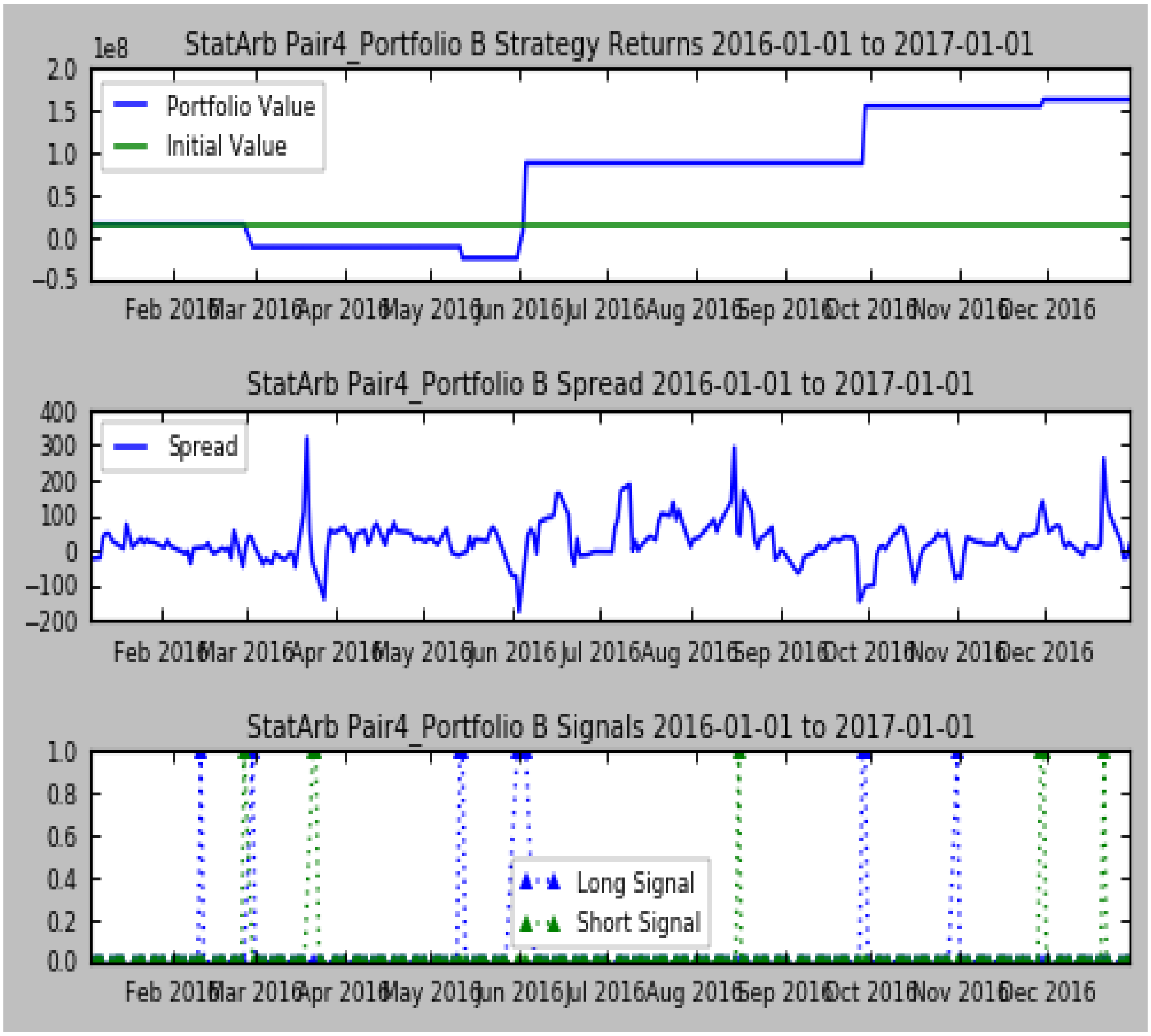


Figure 7



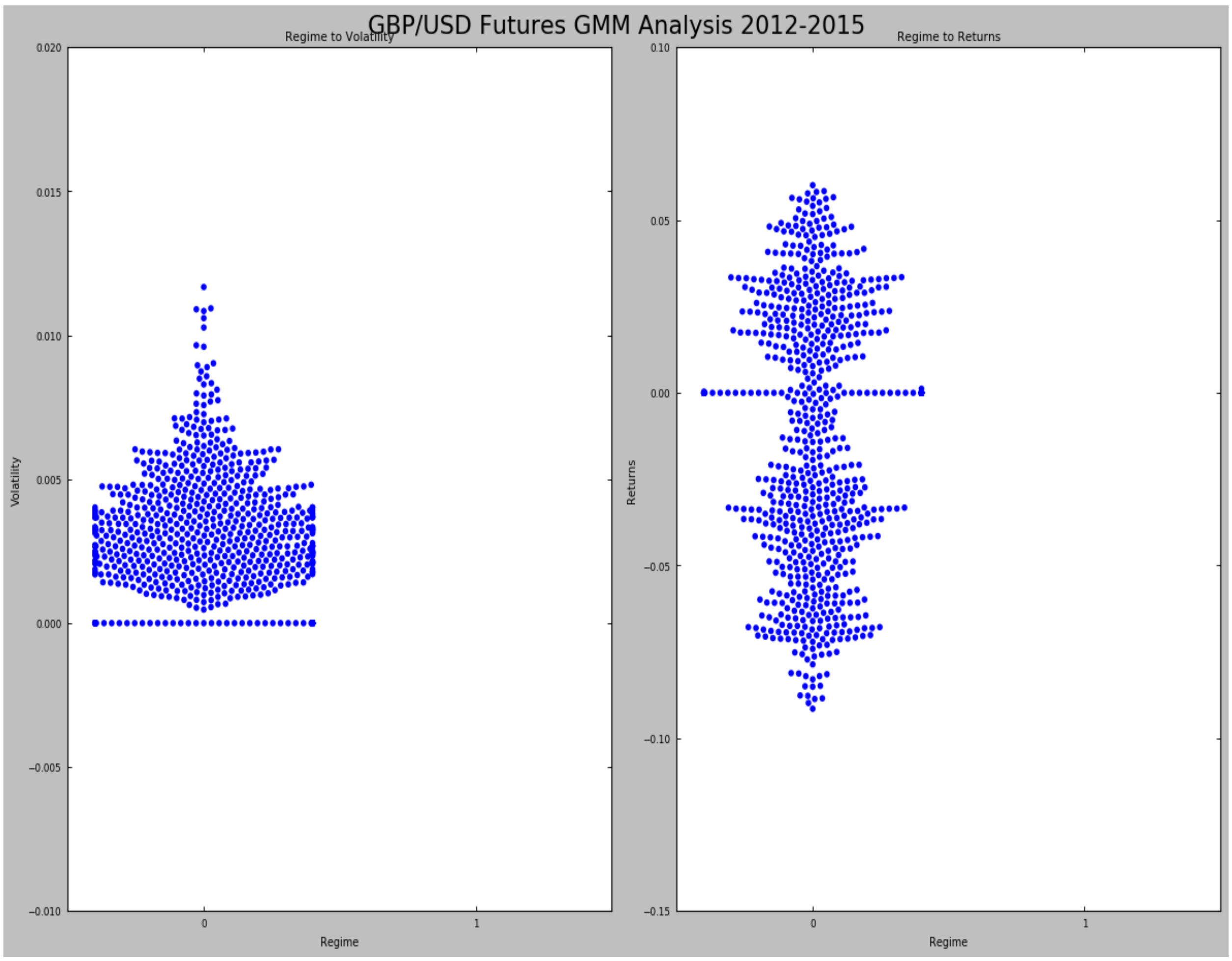
Trend Following

The Trend-Following implementations followed conventions supported by empirical evidence. The signal generator was composed in a way congruent with the literature. Once the strategy implementations were configured, the parameters were then individually optimized across each currency future. This process was consistent across Portfolio A1 and A2.

Portfolios B and C employed the use of the machine learning ensemble and thus deviated from the exact implementation of the aforementioned portfolios. Neither the strategies nor parameters were changed. A Gaussian Mixture Model was used to identify the regimes inherent to each currency future.

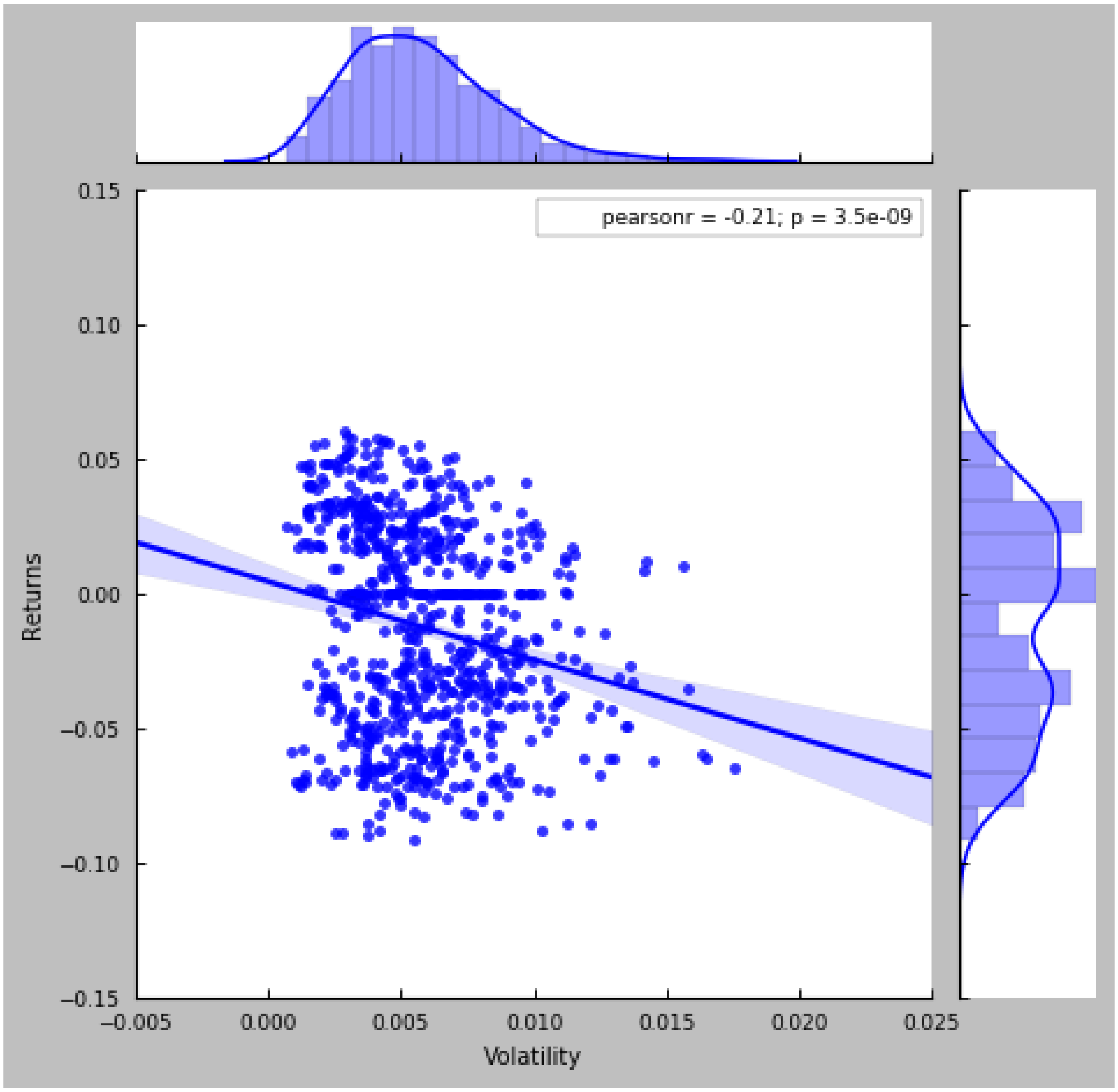
The GMM and RF methods within the Trend-Following strategy regime presented an unforeseen outcome. After completing the GMM on each implementation within the Trend Following regime, it was discovered that each only fell into a single regime. The figure below shows the GMM results for the Pound Futures.

Figure 8: Pound GMM Analysis



I began this study with the intent of using the regimes predicted by the GMM as labels for the ensemble, thus this required an alternative means to apply the analysis conducted on the Trend Following regime. I viewed the above example, as well as the other components of its regime, from the perspective of volatility thresholds. I sought to better understand what levels of volatility were pivotal in performance. The figure below displays this analysis.

Figure 9: Pound Trend-Following Analysis



The intent was to identify levels of historical volatility that could mark shifts in the return distribution of the strategy. Once this threshold was identified, it was then used as an input to the signal generator to augment the strategy. Below are some examples from the Trend-Following regime over the testing period.

Figure 10: Portfolio A1 Pound Futures

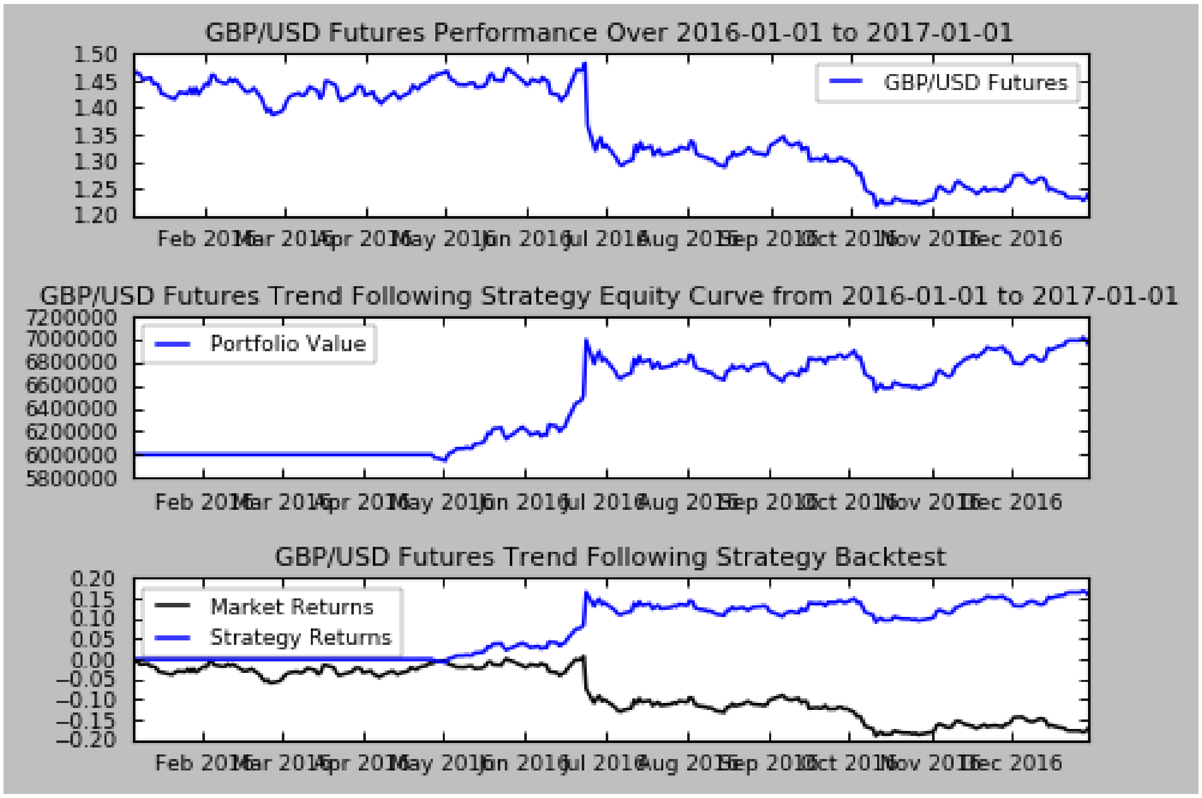
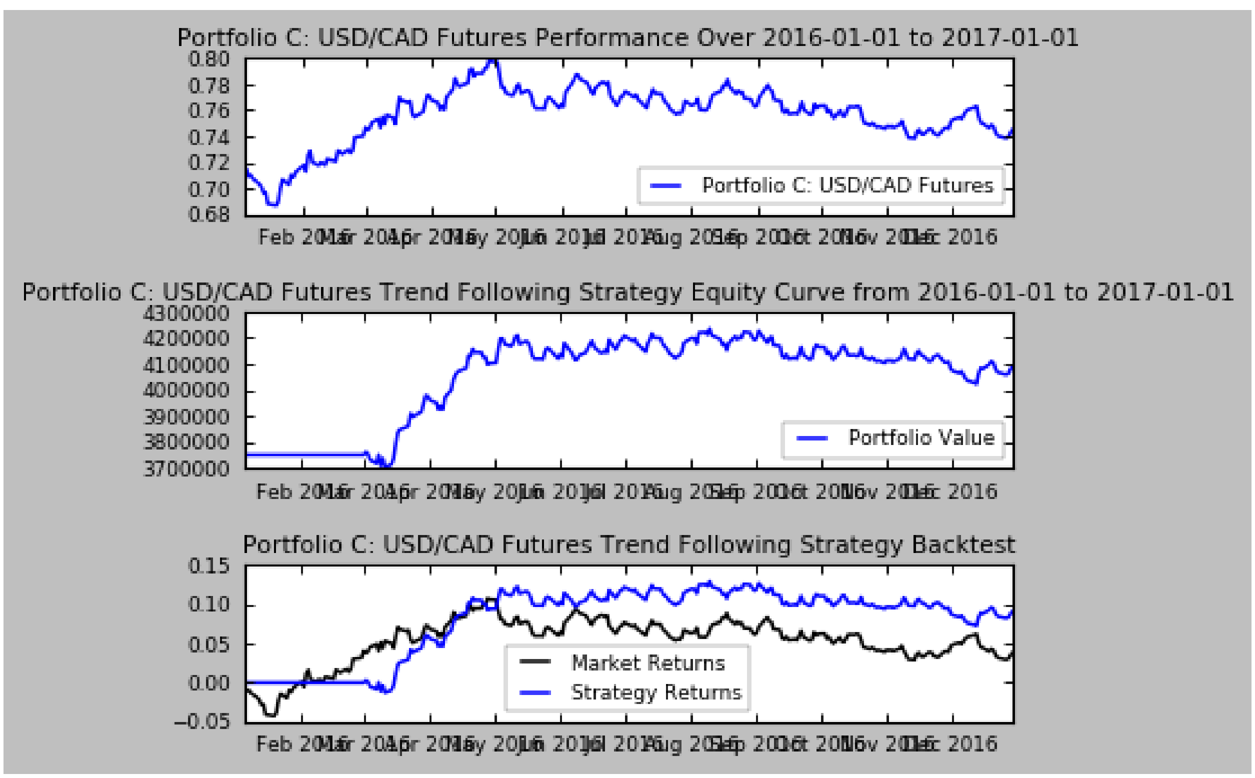


Figure 11: Portfolio C CAD Futures



Portfolio Composition

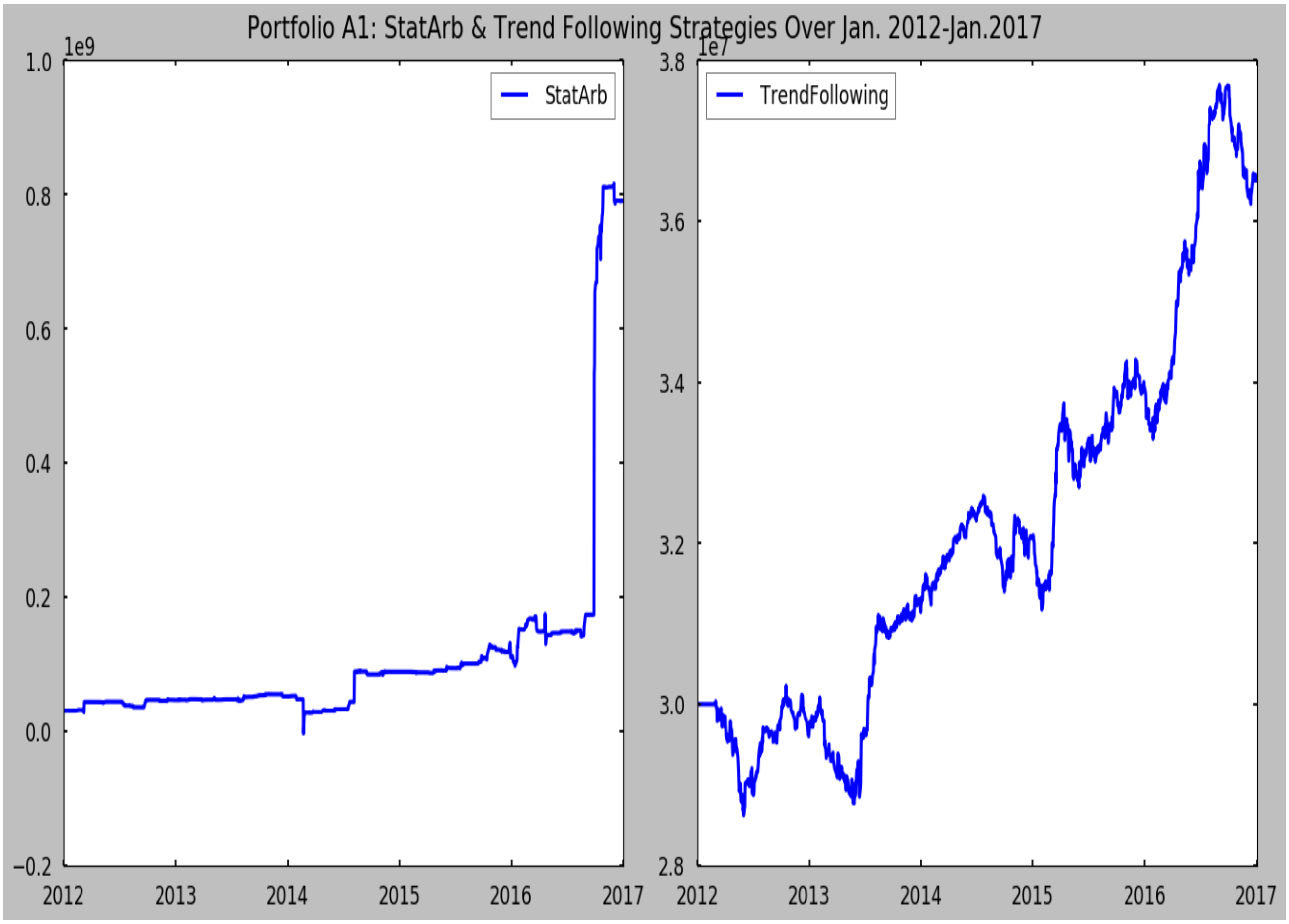
Each portfolio is composed of the Statistical Arbitrage and Trend Following strategy regimes. The Statistical Arbitrage and Trend Following strategies were initially backtested over the period of 1st January 2012 to 1st January 2017, however due to the implementation of the 80/20 train-test split needed for Portfolio B and C, the evaluation period, for comparison purposes became 1st January 2016 to 1st January 2017, representing the 20% testing period.

The performance of Portfolio A1, over the entire sample period (i.e 2012-2017), can be seen in Figure 12 below. Initially, the portfolio's initial values were set at $100 million USD. However, after experiencing issues and limitations surrounding the microstructure and dispersion strategies, resolving to test the idea across the statistical arbitrage and trend following strategies, the portfolio initial value was then reset at $70 million USD. The running cash allocation was $10 million USD. Portfolio A1 (i.e. equally weighted) applied the remaining $60 million USD evenly across the statistical arbitrage and trend following strategy regimes.

To develop the portfolios each instance of each strategy regime was implemented and the portfolio value series were then aggregated to form a composite portfolio value series for the strategy regime. After completing this initial phase on Portfolio A1, the Mu and Sigma could be derived and be used to create the Efficient Frontier needed for Portfolio A2 and B implementations. The same weights arrived at during the Portfolio A2 implementation was used in the Portfolio B implementation. Next, each strategy regime composite was then aggregated to yield the total performance of the respective portfolio.

Portfolio A1 is an equally weighted portfolio across each strategy regime. I refer to this portfolio as the "Equally Weighted" or "No Optimization" portfolio. Portfolio A2 contains allocations to each strategy regime on the basis of the Efficient Frontier. I refer to this portfolio as the "Top-Down Optimization" portfolio. Portfolio B, using the weights derived in Portfolio A2, combines top down optimization with the use of machine learning ensembles on the microstructure. I refer to this portfolio as the "Stereoscopic Optimization" portfolio. Portfolio C, building on Portfolio A1, contains equal weights across each strategy regime, did not undergo the Efficient Frontier, but employs the use of machine learning ensembles on the market microstructure. I refer to this portfolio as the "Bottom-Up Optimization" portfolio.

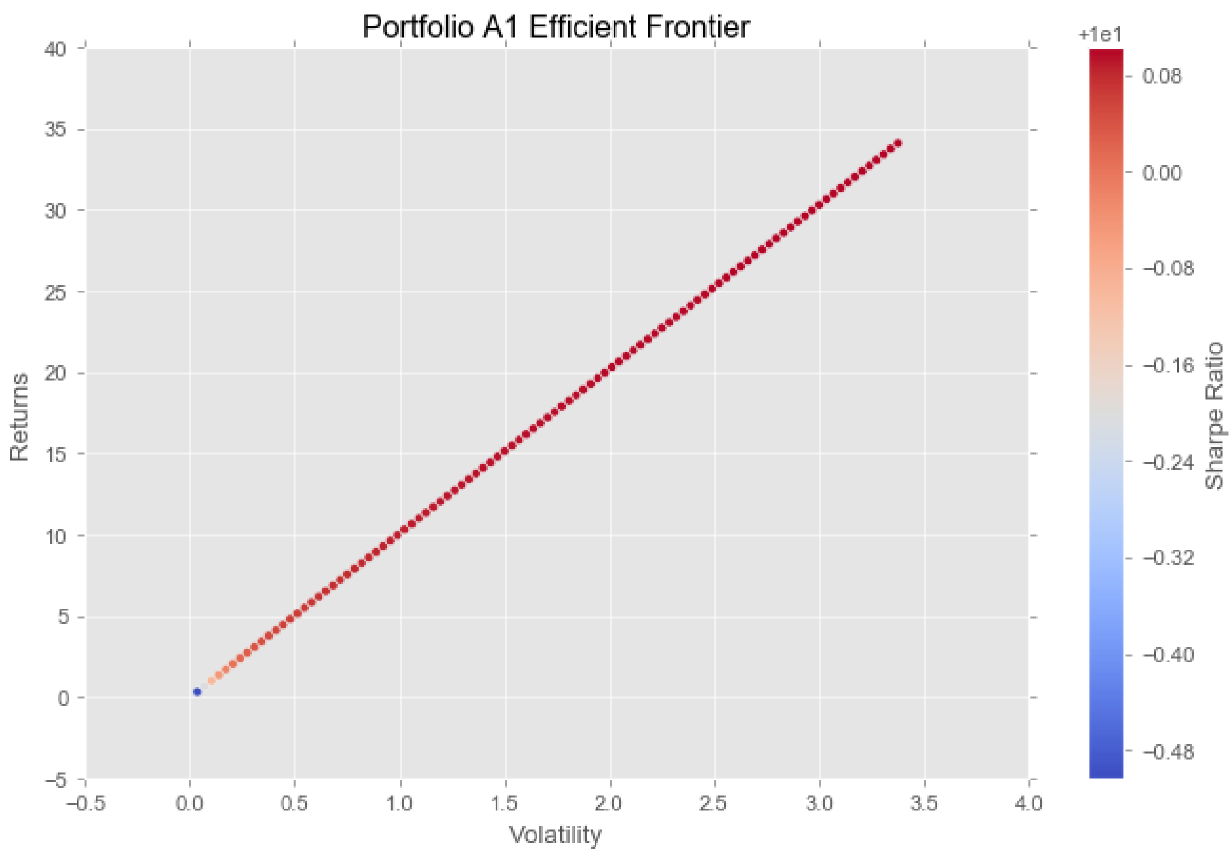
Figure 12



Portfolio A1: StatArb & Trend Following Strategies Over Jan. 2012-Jan.2017
1e9
StatArb
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
2012
2013
2014
2015
2016
2017
1e7
TrendFollowing
3.8
3.6
3.4
3.2
3.0
2.8
2012
2013
2014
2015
2016
2017

A key turning point in my analysis came when conducting the Efficient Frontier. To avoid look-ahead bias, the weights were derived from 2015-2016 performance. This data was made available as I had built the entirety of Portfolio A1 over the entire sample period (2012-2017) and thus was able to parse out the performance for 2015-2016. The percentage change in the portfolio value of each regime was used to derive the Mu and Sigma over the period. The Statistical Arbitrage regime performed exceptionally over the period achieving returns of 36% with a Sharpe of +2%. However this posed a slight issue for the Efficient Frontier configuration. As a result of the relative outperformance of the Statistical Arbitrage regime, the Efficient Frontier stated to grossly over-weight the regime. The figure below shows the Efficient Frontier derived.

Figure 13



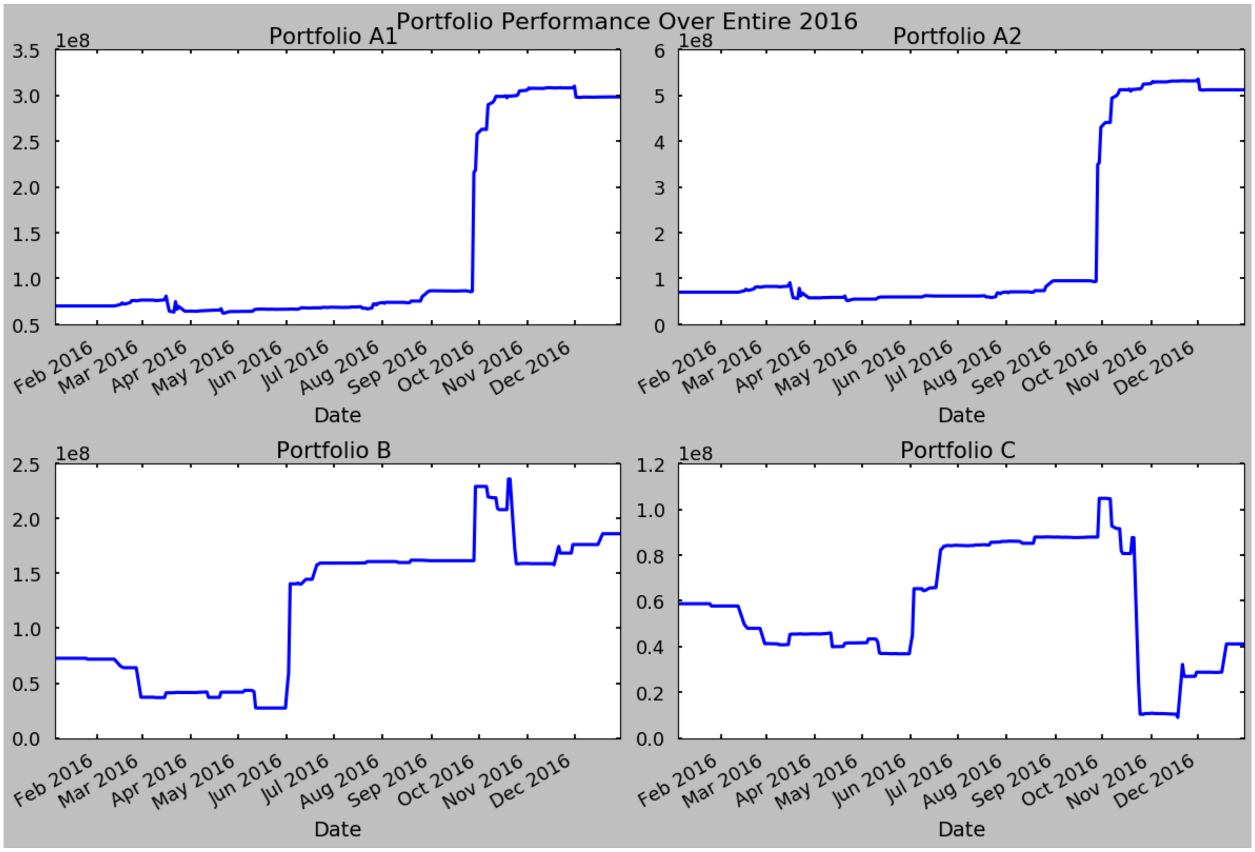
4. Conclusion

The primary metrics used for comparative analysis are the Mu, Sigma, Sharpe, and Sortino, with the Sharpe and Sortino illustrating relative risk-adjusted returns, which is the key differentiator of the portfolios. Mu and Sigma were calculated based on the percentage change in the portfolio value. Portfolio A1achieved a Mu of 0.008 and Sigma of 0.099. Portfolio A2 achieved a Mu of 0.015 and a Sigma of 0.18. Portfolio B achieved a Mu of 0.008 and a Sigma of 0.11. Portfolio C achieved a Mu 0.008 of and a Sigma of 0.18.

With a zero interest rate assumption, the Sharpe and Sortino Ratios are as follows: Portfolio A1 Sharpe of 8% and Sortino of 58%. Portfolio A2 Sharpe of 8% and Sortino of 55%. Portfolio B Sharpe of 7.2% and Sortino of 20%. Portfolio C Sharpe of 4% and Sortino of 13%.

Within the study an interest rate assumption was used based on the level of the Fed Funds futures. The Fed raised its target range from 25bps-50bps to 50bps-75bps on December 14, 2016. Thus, the majority of the testing period saw rates in the 25bps-50bps range. Applying the midpoint of this range as an interest rate assumption, the following are the Sharpe and Sortino Ratios: Portfolio A1 Sharpe of 4% and Sortino of 31%. Portfolio A2 Sharpe of 6% and Sortino of 41%. Portfolio B Sharpe of 4% and Sortino of 11%; Portfolio C Sharpe of 3% and Sortino of 7%. The figure below displays the equity curves for each portfolio over 2016.

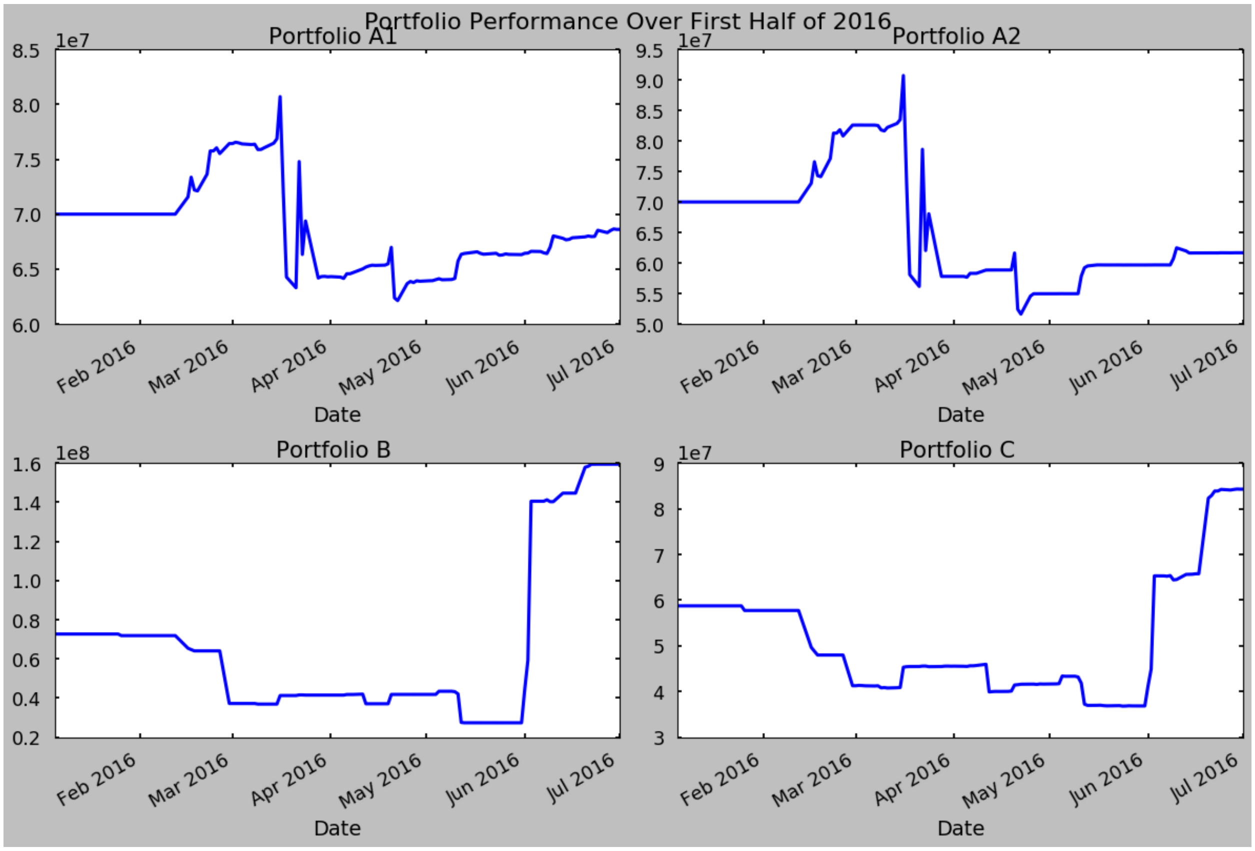
Figure 14



It is to be noted that two factors skewed the above metrics, 1) the gross weights of the Efficient Frontier and 2) the increase in volatility due to the U.S. election. The Efficient Frontier, derived over 2015-2016 period to obtain weights for Portfolio A2 and B, saw a period in which the Statistical Arbitrage regime returned 36% with a Sharpe over 2%. This, in conjunction with the relative underperformance by the Trend-Following regime, though the strategy was positive, led to the Efficient Frontier assigning weights of 98 to 2 in favor of Statistical Arbitrage. While this is not practical from a real world vantage, it was interesting as a means of testing the SPO framework as well as illustrating the importance of regimes and regime shifts to portfolio management. The correlation of the two strategy regimes over the 2015-2016 period was 72%, and 42% over the 2016-2017 period.

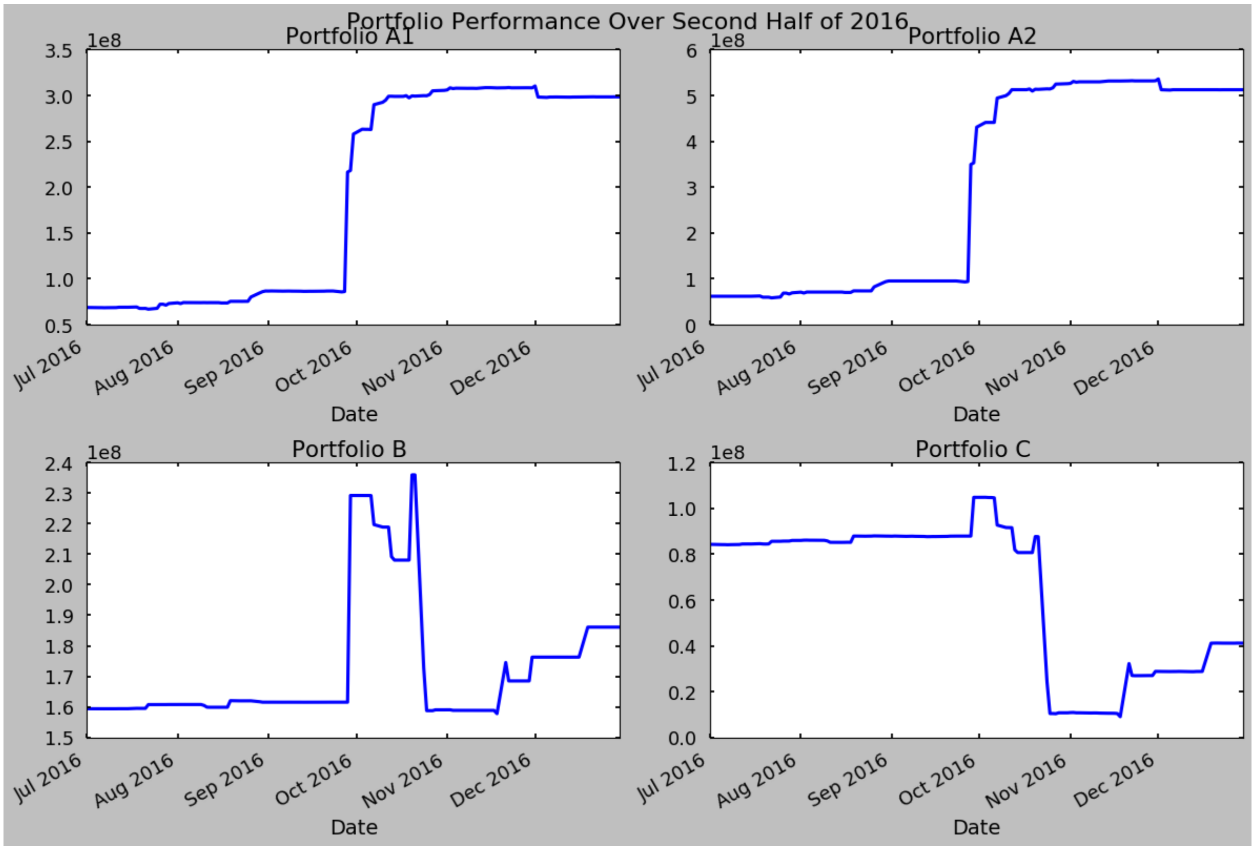
The increased volatility around the U.S. election undoubtedly skewed the Mu and Sigma which are inputs into the Sharpe and Sortino. The median serves as a better metric in the presence of outliers. The figure below shows the performance of each portfolio over the period of 1st Jan 2016 to 1st July 2016, or the first half of the year.

Figure 15



Of keen interest was the comparison between Portfolios A2 and B. Portfolio B implemented the SPO framework while A2 implemented the traditional top-down optimization (i.e. Efficient Frontier). The weights used, relative to each strategy within each regime were consistent across each portfolio as were the strategy implementations. However, Portfolio B augmented the signal generator using the predictions of the ensemble based on the regimes found by the GMM over 2012-2015. It appears that due to 1) the change in volatility due to the 2016 U.S. election and 2) the skewed weights induced by the Efficient Frontier that the SPO framework struggled to perform in an environment in which it had not been trained for, or in other words experienced a regime shift. The next to figures corroborate this and display that the sudden increase in volatility skewed the original metrics.

Figure 16



The above plot confirms my assumption, in addition to that of Figure 15 that the SPO framework portfolio, or Portfolio B's performance was impeded by an influx of volatility of which created a regime shift that the model had not been trained for. The figure below provides metrics that further support this notion. The 2016 testing period was partitioned into the first and second half of the year. The Mu, Sigma, and Sharpe metrics were derived and can be seen, in addition to the same metrics for the entire period below. The sudden change in volatility took the non-machine learning portfolios from a lagging to a leading state and induced the inverse for the machine learning portfolios. It is evident that true performance is based on consistency and not returns based on sudden bursts.

Figure 17

Portfolio Performance Metrics				
Portfolio	Mu	Sigma	Int.Rate	Sharpe
Portfolio A1				
A1 Total	0.008	0.099	0.00375	0.043
A1 1st Half	0.0002	0.027	0.00375	-0.1315
A1 2nd Half	0.0164	0.135	0.00375	0.094
Portfolio A2				
A2 Total	0.0145	0.178	0.00375	0.061
A2 1st Half	0.000432	0.055	0.00375	-0.06
A2 2nd Half	0.0282	0.2496	0.00375	0.1007
Portfolio B				
B Total	0.00799	0.101	0.00375	0.039
B 1st Half	0.01378	0.1482	0.00375	0.0676
B 2nd Half	0.00228	0.0482	0.00375	-0.0304
Portfolio C				
C Total	0.00826	0.18016	0.00375	0.025
C 1st Half	0.0042	0.0552	0.00375	0.00858
C 2nd Half	0.0121	0.2473	0.00375	0.0339

The above figures are in support of the hypothesis test in this study and confirm that the SPO framework is a viable approach to both the strategy and portfolio optimization problem. Below are some key takeaways from this effort.

- The Stereoscopic Portfolio Optimization framework, though rendered in this introduction to the idea in a primitive manner, yet proved to be a robust solution to the portfolio optimization problem
- The study also showed that the framework has some application over medium time horizons
- The implementation of the SPO is sensitive to sudden changes in volatility which in this study may be due to the primitive design
- Despite the inefficient weights derived by the Efficient Frontier from a period of volatility, the SPO still managed to outperform all other portfolios over the most relevant period of the study, the first half of 2016.
- The performance of the SPO, of which was achieved without recalibrating the models shows continued room for further advancement.

An interesting use case of the framework may be over shorter intervals. One could alter the interval based on foreseen shocks. From a trading perspective, based on the fact that this study showed predicting current regimes based on prior elongated periods may lead to regime shifts during implementation; it appears that a viable approach for the SPO framework would be within a shorter term trading perspective. The framework could be used on a fractal basis in which the model was applied to a higher time frame and the implementation was conducted on a lower time frame; this should address the issue of regime shifts. Or, if one would like to apply the model outside of an intraday trading time frame, it appears that an effective means to do so would be by developing the model over a look back of 2-3months and forecasting regimes over the next month for trading, recalibrating the model at month end. In both cases, the model should be recalibrated at the end of the period to account for market shifts.

The SPO can be applied to portfolios with longer holding periods, but it would seem that, to avoid regime shifts, this framework would need to be recalibrated in some shorter term roll (i.e. 3 months).

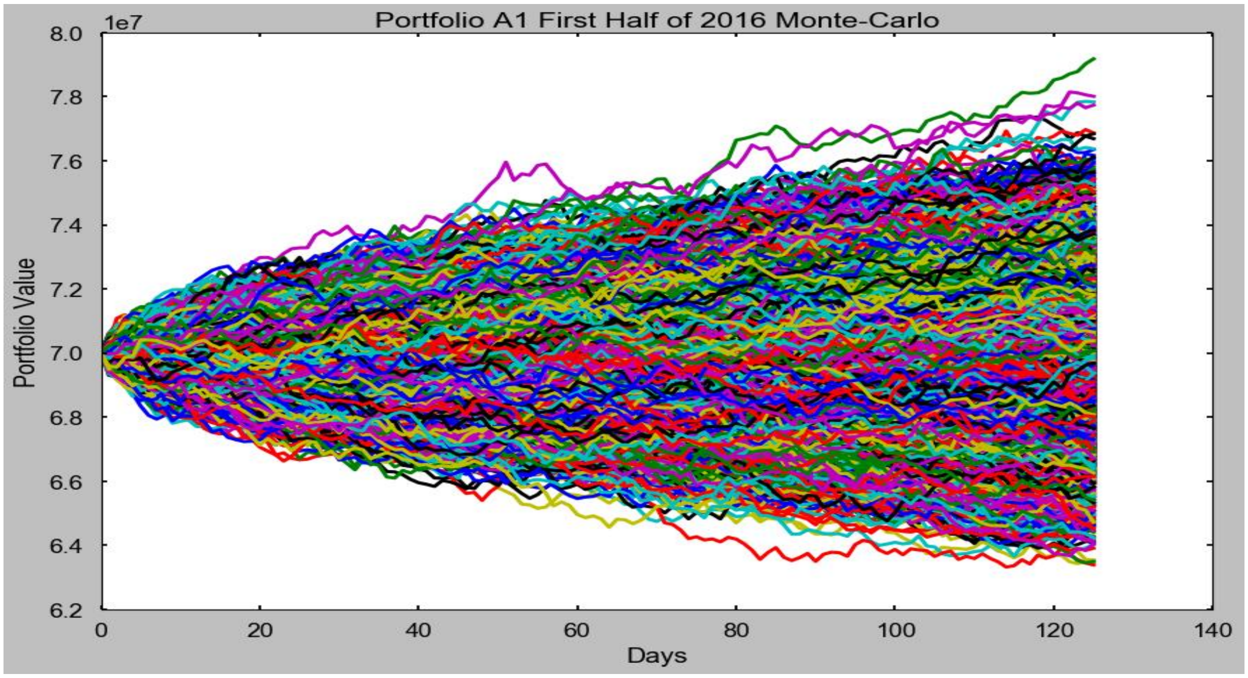
The Stereoscopic Portfolio Optimization (SPO) framework consists of the combination of a traditional top-down approach in addition to a bottom-up approach similar to the one outlined in this text. The fundamental assumption is that a portfolio is the sum of 'n' market microstructures and thus portfolio optimization can be achieved by supplementing well known top down methods with this new idea of bottom-up portfolio optimization via ensembles on the microstructure. To implement the SPO, one need not remain in the confines of the Efficient Frontier and combination of GMM and RFs. These were chosen as introductory methods. The SPO can be implemented via any combination of a top-down and bottom-up approach in which an existing strategy is present upon which the bottom-up approach augments the signal generator based on some microstructure factors. The microstructure factor need not be volatility in isolation, but can also consist of any combination of volatility, order arrival rates, liquidity, etc.

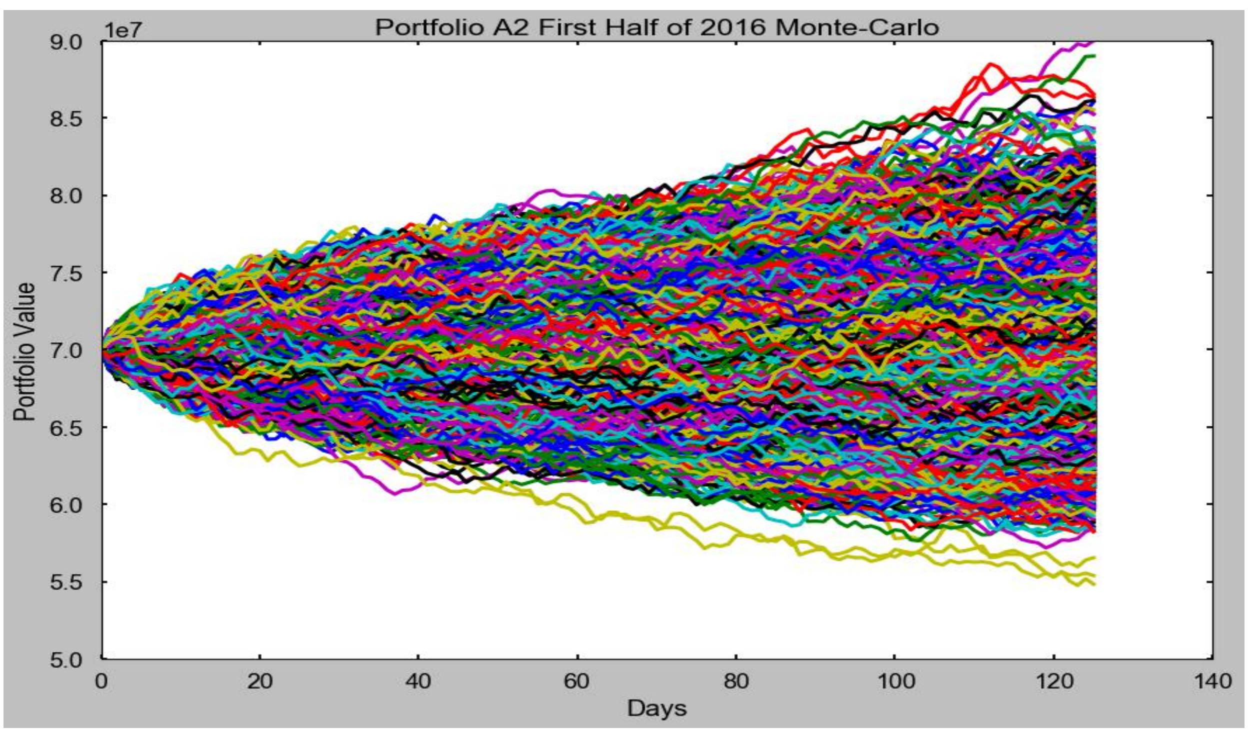
In the future, I intend to extend this idea toward a shorter term quantitative trading portfolio environment. I am also interested in different combinations of the SPO and the resulting impact on trading portfolios.

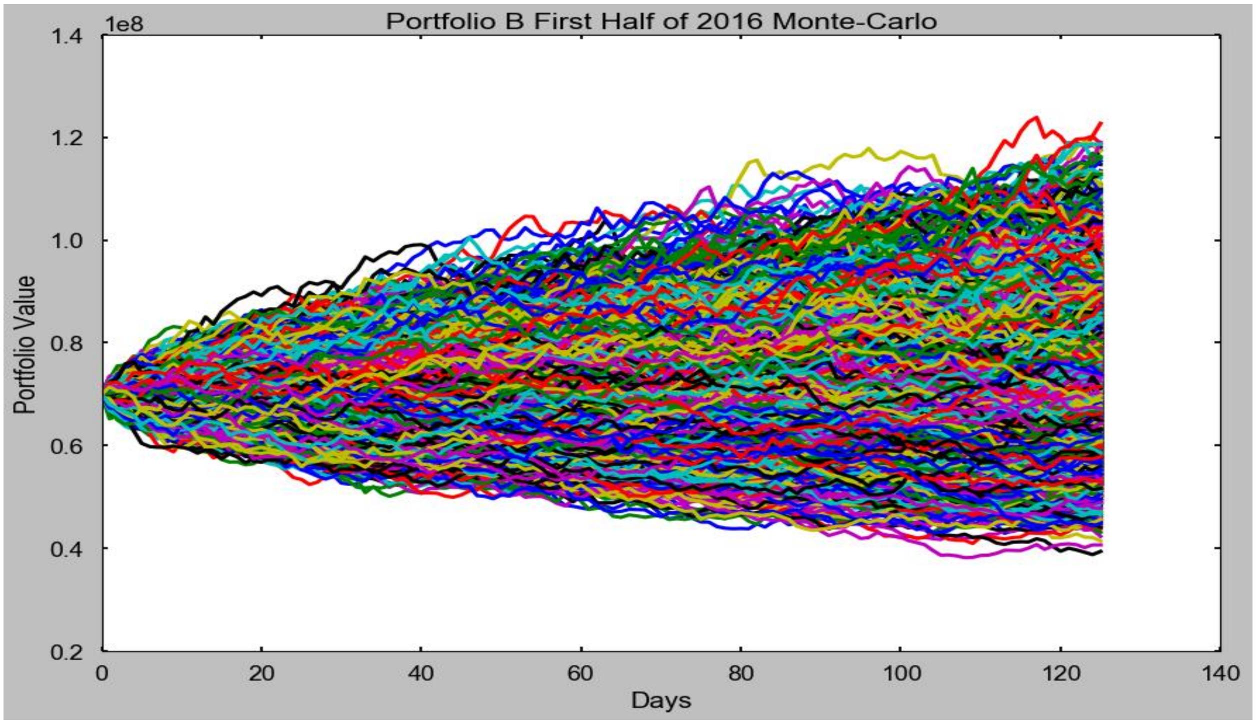
Author's Note:

I am open to feedback and additional suggestions for further research. This paper serves as an introduction to the idea of the Stereoscopic Portfolio Optimization (SPO) framework and I intend to test this idea in different settings. I am also open to general discussions relative to quantitative research and trading. I may be reached via twitter at @MarcusQuant.

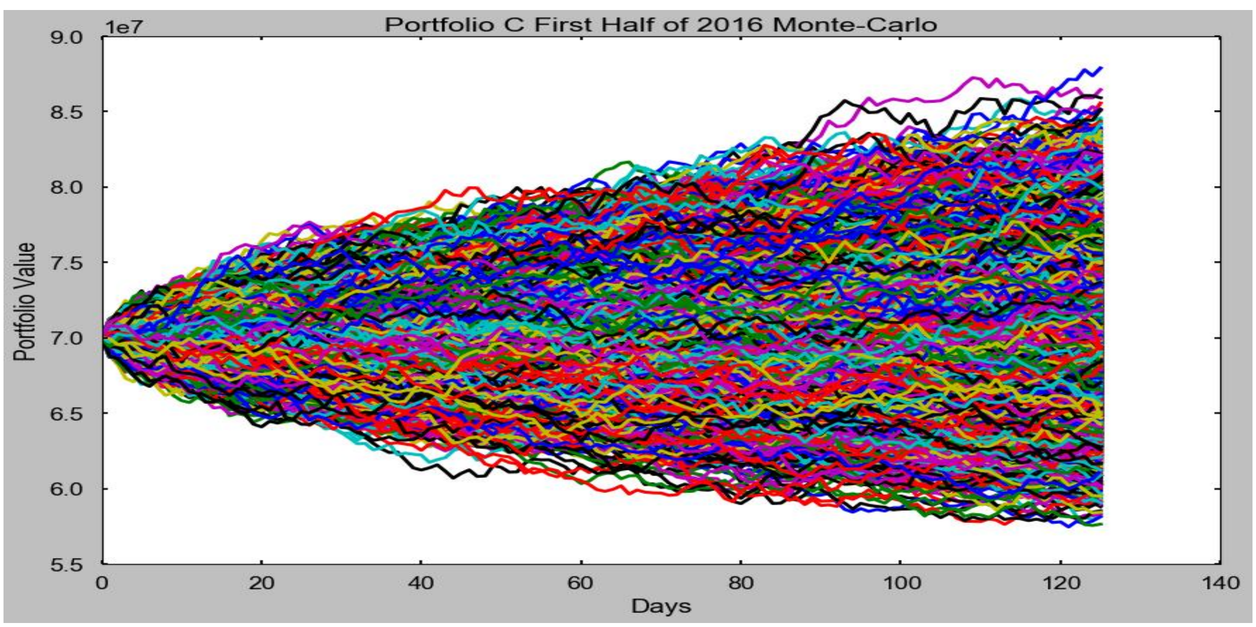
Appendix A: Monte Carlo Analysis







Portfolio B First Half of 2016 Monte-Carlo
1e8
Portfolio Value
Days



Portfolio C First Half of 2016 Monte-Carlo
1e7
Portfolio Value
Days

Appendix B: Dispersion Trading

Dispersion Trading

Dispersion trading seeks to capitalize on the difference between implied and realized volatility. While the existence of a spread between implied and realized volatility is known amongst market practioners, possibly less common is the fact that the difference between implied and realized volatility is skewed in favor of index options in comparison to single stock options. This presents an opportunity in which traders can establish short volatility positions on the index options and simultaneously take long volatility positions on the individual components of the index. It is to be noted that not every component of an index is necessary to exploit the inherent correlation of the index and components volatilities, so long as a majority percentage of the index is represented.

Dispersion trading is a methodology that seeks to profit from the correlation of implied volatilities of an index and its components.

The index can be mathematically represented as the sum of each individual stock multiplied by its weight. The variance is the sum of the products of the weights and sigmas squared plus the sums of the products of each component weights, sigmas, and correlation. See the equation below:

For Index

$$I = \sum_{i=1}^{N} w_i \, S_i$$

where W_i is the weight of the stock 'i' in the index. The variance of index can be calculated using the

$$\sigma_I^2 = \sum_{i=1}^{N} w_i^2 \, \sigma_i^2 + 2 \sum_{i=1}^{N} \sum_{j>i} w_i w_j \sigma_i \sigma_j \rho_{ij}$$

where σ_I^2 is the index variance, w_i is the weight for stock in the index. σ_i^2 is the individual stock variance, and ρ_{ij} is the correlation of stock i with stock j.

The general theme of the dispersion trading strategy is that the correlation of the implied volatilities of the index and its components mean reverts. If a trader is long correlation, the realized volatility of single stocks will be higher than that of the index. Literature presents a variety of methods for profiting from this event. Some of which include profiting from the overpricing of index options relative to its components' options as well as seeking dispersion trading as a function of analysts' forecasts concerning index components.

To ensure that the strategy's primary focus is on capturing the effect of changes in volatility delta hedging is employed. This is achieved by simply establishing a position in the index equivalent to the net delta of the options portfolio. The index position's delta cancels out the delta of the options portfolio bringing total delta to 0. It is to be noted that this process is dynamic in nature and the trader will need to adjust the index position as delta changes over time.

Researchers have implemented several variations of the options volatility dispersion trading methodology. Lisauskas (2010) applied the dispersion trading methodology to the German options market. The strategy was limited to call options on the DAX 30 and the index itself. Testing occurred in the DAX from November 3, 2008 to May 10, 2010. Positions were not held longer than one month. The volatilities retrieved from DataStream were used as the implied volatilities of one month at-the-money options. The component volatilities were computed using Markowitz Implied Volatility (MIV). Correlations were calculated for three time periods, 3-months, 6-months, and 1-year. The MIV was calculated each day by inputting the weights, implied volatilities and the correlations into the standard Markowitz formula. This value was then compared to the Index Option Implied Volatility (IOIV). Signals were generated when the MIV was one volatility point larger or smaller than the IOIV. This means, due to the use of three timeframes for the correlation calculation, the MIV of all three time periods must provide a congruent signal in order for a trade to be initiated. The rationale for this adjustment was that by doing so, the volatility of the strategy would be reduced and thus the Sharpe Ratio would improve. A profit target of 10% was used. A stop loss of 3% was also used. If the time to expiry of the options was less than two weeks, the next expiry options were traded. During this period, Lisauskas (2010) discovered that the dispersion between the Markowitz Implied Volatility (MIV) and the Index Options Implied Volatility(IOIV) was dependent upon the timeframe in which the correlation was taken. The longer the timeframe, the lower the dispersion, and thus the more closely the MIV would track the IOIV. In some instances, the researcher found that the MIV would exceed the IOIV. The MIV and IOIV would vary more so over shorter correlation horizons. Lisauskas' research also suggested that the dispersion strategy had been "arbitraged away", suggesting that due to the greater efficiency of the market, it would be difficult to produce profits after transaction costs. It was suggested that the strategy may be best employed by market makers of who were not subject to transaction costs. Before transaction costs, the strategy's average return was 10.22% with volatility, skewness, kurtosis and sharpe of 8.62%, -0.05, 018, and 1.18 respectively. Of the 80 signals generated, only 29 were traded. Of the 29 trades executed only 6 were losses. Over the same time period, the DAX index rose by 11.9% on 30.07% volatility. This produced a Sharpe Ratio of 1.18 for strategy compared to the index's 0.36. After transaction costs were modeled, the strategy's average return was 3.34% on 8.44% volatility. The Sharpe Ratio declined to 0.4 with skewness and kurtosis of

0.38 and 0.81 respectively. Lisaauskas (2010) also sought to address the assumption of the Black-Scholes model concerning a flat volatility smile. In reality, volatility tends to increase as the strike price declines and decrease as the strike price increases. The Black-Scholes model assumes that volatility is not affected by the strike or time to expiry. Given that the implied volatilities from the DataStream were plugged into the BSM to arrive at the value of the option, in this case, call options, the volatility smile could have increased or decreased the expected profits. Though the researcher is limited by access to real market data for different strike options, he references Goldman Sachs studies which state that the implied volatility of the stock index is on average increasing by 4 percentage points when the strike/spot ratio decreases by 0.1, and that it remains constant at ATM-6% level if the ratio increases above 1.15. Though this assumes a linear skew in a sense, which is not completely accurate, it provides the researcher with some information as to how the strategy will perform with the volatility smile factor introduced. Post introduction of the volatility smile, the returns of the strategy were 9.60%, a decline of 0.62%, the Sharpe ratio improved to 1.27% from 1.18%.

References

Chaudhuri, T. D., Ghosh, I, and Priyam Singh(2017) Application of Machine Learning Tools in Predictive Modeling of Pairs Trade in Indian Stock Market, IUP Journal of Applied Finance, Vol. 23, Issue 1, p5-25

Chen. Y.,Yang, B., Abraham, A. (2006) Flexible Neural Trees Ensemble for Stock Index Modeling. Neurocomputing 70 (2007) 697-703.

Clare, A., Seaton, J., Smith, P., and Thomas, S. (2013). Breaking Into the Blackbox: Trend Following, Stop Losses and the Frequency of Trading-The case of the S&P 500., Journal of Asset Management (2013) 14: 182. https://doi.org/10.107=57/jam.2013.11

Cottle, S., Murray, R. F, and Block, F.E.,(1987). Graham and Dodd's Security Analysis, New York, New York, McGraw-Hill Book Company, ISBN: 0-07-013235-6

Cura, T. (2008). Particle Swarm Optimization Approach to Portfolio Optimization. Nonlinear Analysis: Real World Applications 10(2009)2396-2406

Driaunys, K., Masteika, S. Sakalauskas, V., and Vaitonis, M. (2014). An Algorithm-Based Statistical Arbitrage High Frequency Trading System to Forecast Prices of Natural Gas Futures, *Tranformations in Business & Economics,* Vol. 13, No. 3 (33), pp.96-109

Fong, S., Si, Y. W., Tai, J. (2012). Trend Following Algorithms in Automated Derivatives Market Trading. Expert Systems with Applications, 39, 11378-11390

Harris, L. (2003). Trading and Exchanges: Market Microstructure for Practioners. New York, New York: Oxford University Press. ISBN 978-0-19-514470-3

Hasbrouck, J. (2007). Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading. New York, New York: Oxford University Press. ISBN 978-0-19-530164-9

Huang C. F., Hsu, C.J., Chen, C. C., Chang, B. R. and Lii C. A. (2015). An Intelligent Model for Pairs Trading Using Genetic Algorithms, Hindawi Publising Corporation, Computational Intelligence and Neuroscience, Vol. 2015, Article ID 939606

Kennedy, J., Ebherhart, R. (1995). Particle Swarm Optimization. Neural Networks, 1995. Proceedings., IEEE International Conference on., DOI: 10.1109/CNN.1995.488968

Krokhmal, P., Palmquist, J., Uryasev, S. (2001). *Portfolio Optimization with Conditional Value-At-Risk Objective and Constraints*

Kumar, M., Das, S., and Govil, S. (2015). Analysis of Stock Volatility Clustering Using ANN. Information Resources Management Journal, 28(2), 32-45, April-June 2015

Lisauskas, J. (2010). *Dispersion Trading in German Option Market*, Master's Thesis, Tillburg University.

Markowitz, H.M.(1959). Portfolio Selection: Efficient Diversification of Investments. New York: John Wiley & Sons. (reprinted by Yale University Press, 1970, ISBN 978-0-300-01372-6; 2nd ed. Basil Blackwell, 1991, ISBN 978-1-55786-108-5)

Ping, L., Yingwei, H., Yong, X. (2016) Portfolio Optimization Using Asymmetry Robust Mean Absolute Deviation Model. Finance Research Letters 18 (2016) 353-362

Polikar, R. (2012, January). Ensemble Learning. Zhang, C. & Ma, Y. (Eds.),Ensemble Machine Learning: Methods and Applications (pp. 1-34), Retrieved from www.springer.com/us/book/9781441993250

Qazi, L. T., Ur Rahman, A., Gul, S. (2015). Which Pairs of Stocks Should We Trade? Selection of Pairs for Statistical Arbitrage and Pairs Trading in Karachi Stock Exchange. *Pakistan Development Review;* Islamabad Vol. 54, Iss. 3, 215-244

Santamaria-Bonfil, G., Frausto-Solis, J., and Vazquez-Rodarte, I. (2015). Volatility Forecasting Using Support Vector Regression and Hybrid Genetic Algorithm. Comput Econ(2015), 45:111-133, DOI:10.1007/s10614-013-9411-x

Vishwanath,R. & Krishnamurti,C.(2009). Introduction to Market Microstructure, *Investment Management: A Modern Guide to Security Analysis and Stock Selection*