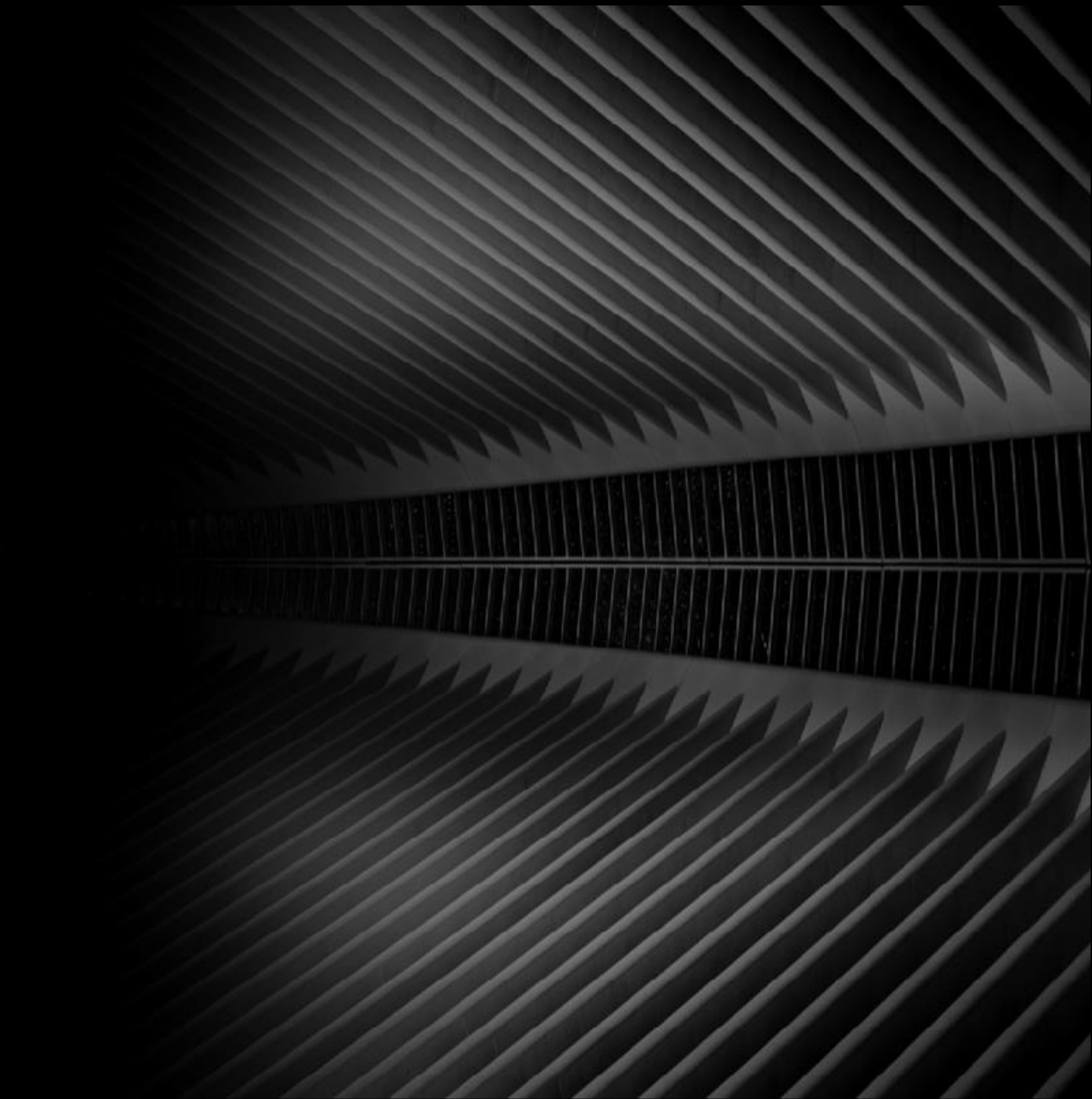
ALPHA

EVALUATION OF SYSTEMATIC TRADING PROGRAMS

by MIKHAIL MUNENZON



intalcon alpha for impact

THE EVALUATION OF SYSTEMATIC TRADING PROGRAMS

by Mikhail Munenzon, CFA, PRM, CAIA, MIA

"

All should be as simple as possible, but not simpler.

ALBERT EINSTEIN

CONTENT

INTRODUCTION

This paper is intended as a non-technical overview of the issues I found valuable in evaluation of systematic trading programs both as a systematic trader and as a large, institutional investor, having looked at numerous, diverse managers in this space on a global basis over the years. Not having been able to find similar material in one source elsewhere when I began research of systematic trading managers, I hope that other investors will become more informed about opportunities and risks of the space that is too often poorly understood. While technical education can be helpful, it is not in my view required to make a high-quality allocation decision. Rather, a process for learning about and evaluating issues relevant to systematic trading, particularly those that are unique to this investment universe, are key. Additional references are listed as necessary to allow the reader to engage in deeper research. Some of the topics discussed below apply to discretionary traders and all types of investment organizations though the focus will always remain on systematic trading. Throughout this paper, I assume that systematic trading refers to an investment program for exchange listed instruments or spot FX that generates signals, manages positions, and executes applicable orders via an automated, previously programmed process with little or no human interference. The list of topics is not by any means exhaustive, but it should be sufficient to allow an investor to start on the path towards successful allocations with systematic trading managers. As I hope will become clear over the course of this paper, systematic trading is a highly unique field with its own set of advantages over other investment approaches that investors should not overlook.

The paper is structured as follows. In Part I, I discuss the value proposition of systematic trading that ensures it a unique niche among available investment options. In Part 2, I then proceed to discuss key areas of evaluation of systematic trading programs: intellectual framework, signal generation, risk management, backtesting, evaluation of live performance, technology, and operational structure. Part 3 includes my concluding remarks and references.

PART I - THE VALUE PROPOSTION OF SYSTEMATIC TRADING

Better than average returns can theoretically come from three sources: information, behavior or analytics. First, one may have access to information that others do not, producing an edge. Second, one may behave in ways others do not that adds value over time. Third, one may analyze data in ways others do not, producing valuable insights. Except situations in which systematic (or discretionary traders) get preferential or illegal access to potentially, market moving information, systematic trading has no informational edge over humans. In fact, systematic trading relies on and prefers consistently clean data to which human traders also have access. The sources of any potential systematic trading edge over human traders must therefore lie in taking advantage of behavioral and analytical inefficiencies of human investors, poorly developed systematic traders and their organizations.

Starting with the research of Kahneman and Tversky (1982, 2000), there is by now very large literature documenting consistent irrationality of human behavior across a broad range of industries and situations, including medicine, law, politics, economics and finance. The list of cognitive biases preventing us from making the best possible investing decisions is similarly extensive, including such imperfections as recency and anchoring biases, bandwagon or herding effect, base rate fallacy, confirmation bias, gambler's fallacy, hindsight bias, loss aversion effect, normalcy and outcome biases, neglect of probability and overconfidence effects[1] . Relatedly, these biases and effects can amplify in a large group or organization (e.g., herding), which is capable of producing its own effects detrimental to investment performance such as slow decision making, inertia, preservation of status quo[2]. These biases are inherent in human nature and can at best be minimized with a very high degree of self-awareness and self – control that very few, if any, of us can reasonably be expected to demonstrate consistently. As a result, as long as people are involved in markets directly or indirectly through programs that they designed and human nature does not change, investment opportunities

[1] For a discusion of some of these biases adn their impact on investing business decisions, see Mauboussin (2012)
[2] Crowds can be smart and smarter than an individual as poularized by Surowiecki (2005) but only under a specific set of conditions that are hard to create and sutain in an organization particulary of large size. For exaample, a group must consist of a verz diverse st of members, who are indedpandent minded and not influenced by others. Some of the largest booms and busts in markets occur due to the breakdown of sch conditions-e.g., extreme consensus of positive or negaive opinions among market participants (for clssic revies of such human behavior historicall, se Kinleberger at al (2005) and MacKay (2013))

created by such inefficient behavior will exist. If with a systematic process, one can minimize as much as possible negative behavioral biases while operating in an organizational environment favorable to high quality investment performance, one should expect such a process to generate value over time.

If one analyzes data in ways others do not, value can be created. This analytical edge may range from a small but valuable incremental innovation to a trading model rivals do not possess to a broad framework that drives research and trading which is fundamentally different from what others do. For example, one may develop a way to aggregate large scale, social media data into effective trading signals. While theoretically all should eventually be copied due to the profit motive, in practice there are significant barriers to such copycat behavior, even when one knows, what needs to get done. Innovations may be hard for others to replicate due to intellectual, behavioral and organizational barriers. For example, it may be hard to reject what one has have grown up with and used professionally for years. As Keynes remarked, 'The difficulty lies not in new ideas but in abandoning the old ones".

It is normally difficult to implement even minor changes in an organization due to inertia and large career risk in case of failure. Again, as Keynes observed, "Worldly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally". As a result, opportunities for innovation and performance of already implemented innovations may persist for a significant period of time.

Finally, one should keep in mind that while the overall hedge fund industry is now very large, systematic trading accounts for only a small portion of the industry's assets, though in many cases systematic managers trade very large markets. For example, HFR estimates that as of 1Q '14, overall hedge fund assets reached $2.7 trillion. Preqin estimates that the top 100 hedge funds account for over 60% of hedge fund assets. Barclays Hedge estimates the size of managed futures (managers who trade futures markets) sector to be a little over $300bn as of 1Q '14 (by comparison, equity long/short universe is about 50% larger and macro universe is twice the size of managed futures managers). Of that amount, well over $100 billion is for Bridgewater, which Barclays includes within managed futures. Once one also eliminates a few, very large CTAs like Winton that alone manages around $20 billion, one is left with a relatively small number of assets spread over a large number of futures managers. As a result, due to a very large concentration of assets among the largest hedge funds and systematic traders and the relatively small asset size of the systematic manager universe, opportunities for innovation and high-quality profitability should be available for nimble, independent - minded players and their investors.

PART II - AREAS OF EVALUATION

I turn to discussing important drivers of value creation in systematic trading that any process must address.

INTELLECTUAL FRAMEWORK OF THE WORLD

It is important to understand how a person thinks about the world in general and markets in particular as such explicit or implicit views have a very significant effect on the decisions a manager makes about key parts of his investment process (to be discussed below). The traditional paradigm prevalent in financial and economic academia and in the financial industry that relies on academic models for investment decisions involves some or all of the following features: linearity, stationarity, parts are consistent with the whole, equilibrium, normality and continuous behavior. Linearity means that increasing the amount of an input will proportionately increase the amount of an output. Stationarity in the data implies that it is not crucial where one starts to evaluate a sample of data as the underlying structural characteristics of the process generating that data always remain the same as do analytical results based on various samples of data. Markets are also assumed to move towards an equilibrium and that equilibrium is the preferred state for markets. Additionally, markets are assumed to follow Gaussian (alternatively called bell – shaped or normal) distribution where extreme events are rare and events above the mean are as likely as events below the mean[3]. Relatedly, market prices are assumed to progress gradually without gaps (continuity). The appeal of the above assumptions is that they allow the deployment of relatively simple mathematical and statistical tools (e.g., regression models, mean and volatility metrics, probability analysis) with a long history in physical sciences in evaluation of phenomena with similar features.

Unfortunately, there is very little empirical evidence that markets exhibit such features. Markets are best categorized as a complex, adaptive system, which invalidates the assumption of the above features[4] . Such system include not just markets but many other important physical phenomena

---

[3] For the technically minded reader, for the Gaussian distribution skewness is 0 and kurtosis is 3.

[4] For an excellent, non – technical introduction on complexitz, the reader is referred to Johnson (2010). For the application of these ideas to some financial situations, see Johnson et al (2003). For more empirical evidence on actual market behaviour, see Mandelbrot et al (2006), Munenzon (2010a,b), Mantegna et al (2007), Sewell (2011), Sornette (2004), Voit (2005). For an entertaining and philosophical perspective on the ideas of complexitz inlife and finance, see Taleb (2005, 2010, 2014).

such as weather, traffic patterns, the spread of infectious diseases and war. Complexity in this context does not imply something that is very hard to do. Rather, it implies that the behavior of the whole is different from its parts and will be non – linear and continuously evolving[5] . As a result, studying parts is not going to be sufficient in understanding the whole. For example, a group of people may behave differently from what an individual member may have decided to do, or individually weak elements can become extremely strong in an integrated system. As importantly, such systems adapt to their environment and shape their environment continuously in self - reinforcing loops with potentially unlimited number of states of uncertain duration, making the classic concept of an equilibrium irrelevant and dangerously misleading. Therefore, boom and bust cycles are an inherent part of complexity and often lead to gaps in a system's behavior (e.g., a market will open much higher or much lower than the prior close). Since complex adaptive systems constantly change due to their interaction with their environments, they are non-stationary: based on different samples of data, one may reach drastically different conclusions as the underlying process generating that data may now be structurally different.

Additionally, complex, adaptive systems are highly sensitive to small differences in initial conditions that may lead to very different results and do not follow Gaussian distribution as extreme events (fat tails) are a regular feature of such systems.

As a result, traditional economic and statistical tools whose assumptions are drastically different from reality are powerless and dangerous when dealing with such systems.

For example, most regression models assume linearity and stationarity of the process being analyzed. The mean and volatility are assumed to be sufficient descriptors of a system when the fact the mean and volatility of a complex, adaptive system is non-stationary and can easily be overwhelmed by a large event[6], making it mislead if a large event has not occurred for a period of time. Moreover, for power law distributions that can describe systems with regular extreme events reasonably well, volatility is theoretically unlimited[7] . Complex, adaptive systems make forecasting hugely problematic except in a narrow set of cases, which reveal the difficulty of considering seriously any market forecasts or even

[5] For example, an increase in the amount of an input results in a more than proportionate increase of an output. For example, if one keeps adding cards (sand) to a large structure of cards (sand), eventually the addition of a single card (grain of sand) will result in the collapse of the whole structure (for example, see Bak (1999)).
[6] Imagine calculating the average income of a population with and without people like Bill Gates and Warren Buffett.
[7] The practical implication of this idea is that other than 100% loss, there is no pre-defined level where a market or an instrument should bottom and there is no pre-defined limit on a market's appreciation. As Keynes observed, "A market may remain irrational longer than one can stay solvent".

traditional probabilistic analysis[8]. For example, weather, whose mechanisms are well – understood as compared to those for markets, can generally be forecast reliably no more than several days ahead with significant accuracy differences across climate zones and local areas (see Palmer et al (2006)). Moreover, our understanding of such a system based on observing the result is highly likely to be faulty and at best incomplete as it is extremely challenging to understand the set of initial conditions that produce this result (e.g., non – linearity and sensitivity to initial conditions)[9]. For example, imagine seeing spilled water in some shape on a table. Did it come from a cup or a bottle or some other container? Even if we know the shape of the container and spill water from it multiple times, the final shape may vary significantly due to small differences between initial conditions. How was its final shape affected by the material of the table and other outside influences?

Complexity is a relatively new, interdisciplinary field that is typically more complicated and less mature (at least quantitatively) than those available for traditional or classical systems. However, research is growing rapidly with multiple applications. One of the key lessons of the field in my view is that in many cases, heuristic, 'rule of thumb' approaches that attempt to address reality even if imperfectly are superior to an extensive new or traditional mathematical model that is out of touch with the world, contributing to very large model risk (to be discussed below) that will eventually impair performance.

SIGNAL GENERATION

When evaluating signal generation, I recommend focusing on the following four topics:

First, what types of signals are being generated and relatedly how are entry and exit signals generated? Are signals assuming stability or convergence to the norm or typical behavior (e.g., relative value or arbitrage strategies, option writing) or are they assuming potentially large market movements and divergence from the norm (e.g., trend following, option buying)? These two types of signals have very different risk profiles. Stability – loving strategies typically target small but consistent payoffs

[8] Imagine playing a game of poker during which a deck of cards is being changed without your knowledge all the time, including combinations that you do not even think or know to be possible. As a result, any probability analysis in such a game is incomplete, naïve, and potentially misleading. That is why in my view (as Taleb similarly observed), games of chance like poker or roulette are poor analogies for markets as in such games, classical statistical/probabilistic tools and the assumption of normality apply. For a philosophical perspective on various probabilistic approaches with their strengths and weaknesses, see Gillies (2000) and Eagle (2011).

[9] In general, understanding of a complex, adaptive system and its events is extremely difficult to achieve as compared to traditional phenomena with our preferred mental models (for applications within social sciences, see Watts (2012)).

(e.g., the spreads between instruments will converge to a typical range or a stable market will allow sold put options to expire worthless as they typically do) at the cost of catastrophic loss periods that can eliminate years of profits or even all the capital[10] as any market stress is extremely unfavorable for such strategies. Relatedly, as these payoffs are typically too small to generate attractive returns due to the size of inefficiencies and large amounts of capital chasing similar opportunities, extensive leverage is utilized, which further destabilizes performance during stressful periods[11]. For example, in a relative value arbitrage strategy, an entry signal may be generated when a spread between two related instruments crosses a particularly high threshold (e.g., two standard deviations away from some historical average amount) and an exit signal may be generated when that spread returns to its typical range (e.g., under one standard deviation away from the mean). By contrast, movement – seeking strategies (e.g., a market is expected to move meaningfully from the norm) expect the status quo to change. They will typically collect relatively small losses on most signals, which may lead extended, poor performance periods. However, such strategies will benefit from a few, relatively large gains on a small portion of signals, leading to an attractive performance and risk profile when viewed from a multi-year perspective (for a more detailed discussion of these issues, see Munenzon (2010 a,b))[12]. Finally, one should consider if a short signal program is a profit – driver by itself or just an expensive capacity enhancer or an expensive hedge for a long signal program that might potentially result in an inefficient use of capital and hide risk management problems.

Second, what types of methods are used in signal generation? As long as some data has to be analyzed, any trading approach must use statistical methods to some degree, though the complexity of models varies greatly. One of the simpler statistical approaches is technical analysis, which attempts to identify potentially profitable patterns in market data (for an overview, see Kirkpatrick et al (2010)). Fundamental factor statistical models that utilize various economic and other non – price factors are also popular (for an overview, see Fabozzi et al (2010)). More computationally complex methods may involve machine learning (or data mining) that

[10] Imagine playing a game of poker during which a deck of cards is being changed without your knowledge all thetime, including combinations that you do not even think or know to be possible. As a result, any probability analysis in such a game is incomplete, naïve, and potentially misleading. That is why in my view (as Taleb similarly observed), games of chance like poker or roulette are poor analogies for markets as in such games, classical statistical/probabilistic tools and the assumption of normality apply. For a philosophical perspective on various probabilistic approaches with their strengths and weaknesses, see Gillies (2000) and Eagle (2011).

[11] Long Term Capital Management is a classic example of these ideas. It engaged in fixed income relative value trading (expecting spreads between various fixed income instruments to stabilize) with extremely high leverage approaching 100 times equity capital to produce attractive returns. When those spreads continued to worsen after the Russian default in the summer of 1998, highly leveraged equity capital could not support such large losses
(see Lowenstein (2001)). Consistent with the ideas I present in this paper on the applicability of the academic framework and its models to real life, two of the partners responsible for trading models (Scholes and Merton) are famous finance academics who won Nobel Prizes.

[12] For the technically minded reader, such strategies exhibit positive skew and small to high kurtosis.

apply various statistical learning models (e.g., neural networks) to data to produce trading signals (for an overview of machine learning, see Hastie et al (2011)). Finally, a trading approach may involve a combination of the above methods. For example, a machine learning algorithm may involve optimized selection of signals from among technical analysis and fundamental models applied to market data.

In my experience, the vast majority of statistical tools from simple regression models to data mining tools deployed in trading make classical assumptions that are inconsistent with a complex, adaptive system like markets. For example, as discussed above, if markets constantly evolve and are non-stationary, then one cannot reliably learn with purely statistical methods like machine learning without understanding the underlying process generating that data (currently unknown to us). In fact, by the time, one's data mining model 'learned' the behavior of a market, the moment of its demise is likely near as the market adapts to what its participants are doing. Relatedly, even some simple, technical analysis indicators suffer from the similar inconsistency between such indicators' assumptions and reality[13]. One should also remember that many technical analysis indicators (e.g., a moving average) generate entry and exit signals with at least some delay as instead of dealing with primary data for signal generation, such indicators by definition produce derived data that can then be used in signal generation. As a result, flexible and timely risk management process is particularly crucial for such signal generation systems. Particular attention should be paid to the fact that models used in research and live trading reflect actual market behavior and that the risk of using inappropriate models is explicitly addressed in risk management, which I will discuss in more detail below. One should also keep in mind that all statistically driven research, including technical analysis models, may suffer from significant data snooping bias[14] that may make their historical results irrelevant for live trading. For example, many popular technical analysis rules that appear attractive in historical data

In my experience, the vast majority of statistical tools from simple regression models to data mining tools deployed in trading make classical assumptions that are inconsistent with a complex, adaptive system like markets.

[13] For example, a moving average indicator in some form is widely used by trend following traders to minimize noise in signals. Such a tool (finite impulse response filter) is very well known and widely used in digital signal processing to deal with noise. However, from a technical perspective, this tool is only valid if the data comes from a linear, stationary process (for example, see Kronenburger et al (2008)), which does not apply to a complex, adaptive system like financial markets.

[14] For example, a moving average indicator in some form is widely used by trend following traders to minimize noise in signals. Such a tool (finite impulse response filter) is very well known and widely used in digital signal processing to deal with noise. However, from a technical perspective, this tool is only valid if the data comes from a linear, stationary process (for example, see Kronenburger et al (2008)), which does not apply to a complex, adaptive system like financial markets.

may in fact indistinguishable from chance once adjusted for data snooping (see Marshall et al (2010), Park et al (2004)).

Third, what is the source of return from signals that a manager's program will be relying on to produce above average, risk adjusted profitability? Are those returns available due to some, relatively clear behavioral or structural inefficiencies that initially stops working either permanently due to a structural change in a market or cyclically due to its normal lifecycle[15]? In my experience, managers start with just one idea and do not necessarily have the level of interest or preparation required to create immediately or over time a multi model approach to deal with the complexity of markets. Once live trading begins, time is spent on improving what is already completed and managing new and existing client relationships. As a result, if that single idea enters a drawdown early in the life of an organization, business risk becomes very significant as the research pipeline is too immature to allow for deployment of additional models.

Fourth, what is the duration of signals that a manager is focused on – short term, medium term, long term or some combination of those? In my experience, there has been over the years an increasing focus on shorter term signals, including intra-day, due to their perceived lower risk profile as compared to longer term signals[16]. Such crowding should lead to lower expected profitability of short-term programs as a large number of traders drive down returns for each other, especially if such short-term programs do not offer any unique features.

More importantly, market opportunities regularly evolve responding to internal and external influences and one cannot ex ante anticipate which duration one should focus on during the upcoming investment holding period. Finally, one should keep in mind that as one shrinks the potential duration of positions, one may also significantly diminish the potential profit of each position since execution costs per trade remain fixed while the number of transactions may increase rapidly as one lowers the duration of trading. This shift in profitability relative to execution costs may not be full appreciated in its overall impact on the program if execution costs are not properly tracked in backtesting (to be discussed below).

[15] Andrew Lo (2004) suggests that a particular trading idea may have its own lifecycle through the adaptive market hypothesis. As news of profits from a particular methodology spread, more capital targets this opportunity, which in turn lowers profitability, leading to an exit of poorly capitalized players, which in turn allows profitability to recover.

[16] The potential for losses might be larger for models generating longer term signals but it is not theoretically clear that this is always or even predominantly the case as many factors in addition to signal duration will affect the overall level of profitability and losses of a systematic trading program.

RISK MANAGEMENT

What is more important - signal generation or risk management? In my view, risk management is the key driver of value creation or destruction of any investment process, taking priority over signal generation, though in my experience, both managers and investors focus most of their time on signal generation. If every signal makes money, one does not of course care about risk management. However, this is a highly unrealistic scenario. As long as one's accuracy on profit generating trades is below 100%, even a single, improperly sized and managed position can overwhelm positive effects of prior positions by risking too much capital at the wrong time[17] . I now turn to the discussion of specific risk management areas investors should focus on.

RISK DEFINITION

This is a crucial area relating to the framework topic discussed above as the selected risk metric will determine the potential for mismeasurement and mismanagement of risk. In most cases, consistent with what is taught in academia and used in the industry, risk is defined as volatility, which has a number of serious theoretical and empirical deficiencies (for an overview, see Munenzon (2010c)). For example, it can be artificially managed to low levels by holding large amounts of illiquid instruments. Also, it does not in any way attempt to measure tail risks[18] and it treats volatility when making money the same when losing money. Value at Risk (VaR) is widely used in the financial community (and even required by regulators for banks) as an alternative to or as a supplement to volatility and similarly has a number of serious theoretical and empirical deficiencies, the most important of which (obvious to any observer of the banking industry history) is that it is not designed to measure potential extreme losses under stress[19]. Some advanced, improved versions of volatility focus on measuring average deviations below some level (e.g., 0).

[17] Imagine that 9 positions make 1% and 1 position loses 10%. Now, imagine that as part of his process, a manager adds capital to losing positions by liquidating winning positions. Finally, imagine that leverage is involved. All these conditions, contributing to ever increasing losses though most trades are actually profitable, are quite realistic and I will address them in more detail in the following sections below.
[18] Imagine that 9 positions make 1% and 1 position loses 10%. Now, imagine that as part of his process, a manager adds capital to losing positions by liquidating winning positions. Finally, imagine that leverage is involved. All these conditions, contributing to ever increasing losses though most trades are actually profitable, are quite realistic and I will address them in more detail in the following sections below.
[19] VaR is designed to measure the expected loss that will not be exceeded with some confidence level (e.g., 95%). However, it is these last few percentage points that contain extreme events that are responsible for potentially catastrophic losses, which this metric ignores.

However, the concept of the average is not what one should be worried about when dealing with complex, adaptive systems as it is misleading or may not even be defined for such a system, as noted above – one should be worried about surviving extreme events, not average events, which provide a limiting view of reality[20] . Additionally, one should remember that the dollar value of an instrument in no way reflects its risk and one should always think and measure in percentage terms to allow for valid comparison across various instruments. As a result, in my opinion, the focus should be on direct management and estimation of losses from extreme (or tail) events. Tail – focused metrics such as Conditional Value at Risk or CVaR (see Artzner et al (1997, 1999) and Embrechts et al (1999)) are relatively new in finance and bring their own set of challenges that are beyond the scope of this paper. However, though no metric will be perfect, tail – focused metrics at least concentrate on the appropriate goal[21] .

RISK BUDGET AND POSITION SIZING

Once risk is defined, one should set an appropriate risk budget for some horizon for the overall portfolio and strategies/markets traded based on the risk definition – e.g., a very common practice is to assume that risk is volatility and then select a particular annual volatility target. One can then determine how much capital can be traded in the portfolio without exceeding the previously set risk budget – e.g., a portfolio's volatility target is 10%, a single strategy in the portfolio produces an entry signal for a stock that volatility of 20% and therefore 50% of the portfolio capital can be used to purchase that stock[22] . An investor should carefully consider how frequently a risk budget target, regardless of how it is defined, is exceeded as frequent and meaningful exceedances suggest inappropriate measurement of risk. Too frequently, historical data is used to set a risk budget for some future period based on convention or anticipated investor preferences (e.g., 10% volatility or 2x ATR (average true range)) and without any understanding of the effectiveness of initial estimates as compared to actual risk behavior. This is particularly unfortunate because to be of value, risk management must be forward looking, rather than backward looking and because risk, regardless of how it is defined, is much easier to forecast than returns due to a well – documented volatility clustering effect across markets, initially noted by Mandelbrot (1963) and Fama (1965) (for a review, see Granger et al (2002)) – high volatility period

[20] Imagine a strategy that in 99 cases loses 5% but in the worst case, loses 25%. The downside average relative to 0% is only 5.2, which is completely unrepresentative of the actual loss potential of such a strategy.
[21] As keynes noted,"It is better to be roughly richt than precisely wrong".
[22] For simplicity, in this example, I assume that the portfolio allows no leverage and has only one position. Numer of shares to purchase is then easily determined as 50 % of the portfolio capital/ price per share.

is likely lead to a high volatility period and low volatility period is likely to lead to another low volatility period. Finally, one should pay particular attention to periods of market stress, especially if a market stress is also likely to be an unfavorable environment for the strategy being traded as during such environments, large price fluctuations and large execution costs can lead to large risk budget exceedances.

Typically, an annual risk budget target is established and remains fixed throughout the year, regardless of the magnitude of gains and losses. An investor should carefully consider two issues about this practice. First, how is risk of current positions updated based on current market data to ensure consistency with the risk budget? For example, a multi - month market data used in risk estimation and update may not properly reflect current market conditions, leading to a significant over - (under -) estimation of risk in the portfolio. On the other hand, a very short data period may capture noise, leading to unnecessary transactions in the portfolio based on that noise. Second, should a strategy be risking the same amount of capital at differing magnitudes of gains and losses? Due to convenience in code implementation and live management, the typical answer in systematic trading is yes. However, I suggest that the answer should be no. For example, risk budget for the portfolio should be much lower when it already lost 20% as compared to a loss of only 5%. Similar logic applies to risk budget sizing when a portfolio is profitable as a program should become more aggressive when the environment is favorable. More generally, for any adaptive system like markets, a flexible risk management approach should perform better over time than a fixed approach. Such non – fixed approach to risk budgeting and position sizing[23] , which is more common in discretionary than systematic trading, is more complicated to implement but should be able to preserve capital better in times of stress and grow capital faster in favorable periods as compared to a conventional, fixed approach.

STOP LOSSES

It is important to note that stop losses are different from exit signals, though the two terms are sometimes mistakenly used interchangeably. Exit signals result in liquidation of a position because a model rates conditions to be unfavorable for the position (e.g., in a traditional trend following model, price falls below its moving average for a long position,

[23] It is ironic that in my experience such a more complicated, non – fixed approach is much more common in discretionary trading (cutting positions at a faster rate as losses increase and growing positions at a faster rate as profits grow) than in systematic trading as the consistency of implementation is a key edge of a computer trader as compared to a human trader.

producing a sell order). Stop loss levels if crossed result in liquidation of a position because some pre-defined amount of capital was lost[24] . As a result, there is no theoretical relationship between exit signals and loss levels of a position or a portfolio. In fact, the same exit signal may lead to a drastically different impact on a portfolio at different times (e.g., in a traditional trend following model, a price move from the entry level to the typical exit price level equal to the moving average of price based on half the length of the lookback window used to calculate the moving average of price used for entry signals can have significantly different performance effects).

Should there be any limits on how much capital is lost in the portfolio before all positions are liquidated or at least meaningfully reduced? Most managers (discretionary or systematic) will answer no for a variety of reasons. Losses are only permanent if one closes a position and markets will eventually reverse and see validity of the premise behind those losing positions – I call it the 'hope' argument also common among many retail investors due to the psychological difficulty of achieving closure on a losing position. In fact, this psychological difficulty is what often leads investors to sell their strongest investments to finance margin calls due to their weakest investments. This argument ignores an empirical fact that due to the volatility clustering phenomenon noted above, not only good but also bad periods have a tendency to last. The risk with such an approach is that losses continue to increase from a manageable level to a very large level, at which point they are effectively permanent as any market recovery may be insufficient to allow for the meaningful recovery and future compounding of capital at a sensible rate[25] . With another argument, an investment that one liked before is now even more attractive at a lower price – this 'attraction' argument is common with many non-systematic equity and debt managers and also stability – loving strategies such as relative value arbitrage. Therefore, one should actually be adding capital to positions as they decline – e.g., as the spread between instruments continues to increase rather than return to its normal range, one should buy more of that spread. The risk is similar to the one above but now at a potentially, far greater magnitude, especially if leverage is employed in the portfolio – if markets do not reverse shortly, one will be losing capital at an accelerated rate. One may eventually be proven right as even a broken clock is right twice a day but not until all capital is exhausted. With the 'history' argument, a manager will avoid using stop losses because based on historical data, his approach always manages to recover, perhaps even from large losses. Unfortunately, for a complex, adaptive system, the future will sooner or later be different from historical

[24] This pre-defined amount may be calculated at the position, model or portfolio level or all such levels, as I will discuss below.
[25] If a portfolio is down 50 %, it needs to rise 100% just to get back to the same level before the decline

data used in the backtest and such structural or cyclical change in market behavior can lead to potentially large losses. Such losses can be particularly severe if a trading approach uses many parameters optimized on historical data that may be significantly sensitive to even small deviations from historically observed ranges (also see below on optimization). One should keep in mind that for many markets (especially for non – US markets), long term, quality historical data may not be available and even when it is available, continuous market adaptation and change may make it of little use for future trading. I strongly recommend that investors consider the absence of a clear, thoughtful stop loss policy a clear deficiency of a manager's process. Stop loss mechanisms allow good positions and ideas to shine without being overwhelmed by bad ones. They allow a program to survive extended drawdown periods, which are in some cases due to the volatility clustering effect. They are also key to controlling a potentially large model risk and dealing with unexpected and unfavorable market developments. Stop loss trades will by definition result in losses though of manageable size and one should make sure that a stop loss policy is not so conservative that it does not give even a slightly losing trade a chance to recover. However, one is far more likely to destroy value by the absence of a stop loss policy than by its presence.

If stop losses are to be used, then at what level should they be applied - portfolio only, individual markets or strategies traded, etc.? All of the above points in favor of a stop loss process when a portfolio consists of a single market or strategy traded still apply even if multiple markets and strategies are traded. If one adopts a top down, centralized approach by setting stop loss levels only at the portfolio level, a single market or just a few markets can overwhelm results of winning positions. As a result, a more flexible, bottom-up approach may be more appropriate where each component traded has each own stop loss level consistent with the implied portfolio stop loss level, which will be achieved only if all portfolio components cross their individual stop loss levels[26] . However, this point is less crucial as long as at least a portfolio level stop loss policy is in effect.

An investor should also ensure that the expected stop loss process is realistic. For example, many option selling programs assume a continuous capability to hedge via related futures contracts or ETFs. This assumption is unrealistic as in stress, market prices gap and execution costs due to slippage will be very significant. If a manager deploys limit orders to control his hedging costs during such times, his orders may be unfilled due to price gaps, leaving him without a hedge precisely when it is needed most.

[26] Computationally, this is of course harder to implement, especially with a multi – marketing/strategy trading process, as multiple data points need to be tracked for stop loss levels at each level instead of a single stop loss level for the overall portfolio.

Finally, clarity and consistency in advance between the definition of risk and the metric used in risk budgeting and stop loss control is crucial as when a portfolio is in a drawdown, one is not likely to be able to think clearly and rationally. For example, if risk is defined as volatility and risk budget is set to some annualized volatility level, it does not make sense to shut down trading (or even to reduce position size) after some, pre – defined drawdown level. Volatility has no direct relationship with drawdown and utilization of multiple, inconsistent risk metrics in risk management (especially if some of them have significant theoretical and empirical limitation as noted above) should meaningfully and negatively affect portfolio performance over time.

MODEL RISK

Model is closely related to the topic of the framework used in trading and investing discussed above. It addresses areas of model uncertainty, model instability, parameter uncertainty, parameter instability and data snooping - all of which can be significantly affected by the degree of relevance of the applied framework in the real world.

Model uncertainty involves the risk of using an inappropriate model or models that do not reflect the behavior of the real world or using models that were not appropriately tested, leading to live trading based on a misleadingly favorable backtest. Given the complexity of markets as noted above and our current limited understanding of the underlying mechanism of such a system, all models will be wrong by definition. However, the degree of model uncertainty will be closely related to the assumptions each model makes. As a result, the fewer are assumptions about model structure and parameters and the more realistic are a model's assumptions about market features, the lower is model uncertainty even if it is highly unlikely to be fully eliminated.

Related to the risk of model uncertainty is the risk of model instability. As complex, adaptive systems continuously evolve, any model must be able to adjust to remain relevant. As a result, a single model especially with static or relatively static set of parameters is likely to experience significant periods of underperformance for cyclical or structural reasons as the phenomenon a model tries to take advantage of evolves.

Finally, even if we could precisely establish an appropriate model structure with the correct set of parameters to fully capture market behavior, the risks of parameter uncertainty and instability remain. We will never know exact values that parameters of a model need to have due to limitations of data and errors in data measurement. This is true even for scientific and engineering fields that can run controlled experiments, which are of course not an option for large scale, social phenomena. Just as a model can experience cyclical or structural change, so can its parameters, resulting in parameter instability. As a result, consistent with the above recommendation, the fewer parameters a model is expected to have and the less sensitive are a model's results to different parameter values, the lower are the risks of parameter uncertainty and instability[27] .

A very significant area of model risk relates to data snooping (or data dredging, data fishing) – the practice of using statistical techniques (or data mining) to find relationships in data. The risk exists whenever the same data set is used multiple times to test different models until the one that works is found or when a researcher decides to test a model after first looking at data or when research is purely or primarily statistically driven[28]. One must be extremely careful to make sure that any profitable relationship found in historical data is not just a lucky accident that has no relevance for future trading.

As importantly, one must ensure that the model is not incorporating and optimizing based on noise in historical data[29] . While such an approach may produce spectacular results in historical data, it is very likely to produce poor results in live trading as such a model cannot then 'recognize' any meaning in live data. Finally, with any statistical method used in research, one should check that the assumptions of such a method are consistent with empirical features of the system that is being analyzed.

There are two ways in which data snooping can be controlled. The first, qualitative approach focuses on developing an idea to trade based on a hypothesis about some market phenomenon and then testing that idea with the simplest possible model. If results are unsatisfactory, one should revise that idea or develop a new one, not shuffle or optimize parameters of the model until one finds a profitable result. It is important to note that data mining techniques are inherently incompatible with this approach – a machine may select parameters for a profitable model that are impossible to understand fundamentally for a human trader even though they are statistically sound. Relatedly, ideally, any profitable

A machine may select parameters for a profitable model that are impossible to understand fundamentally for a human trader.

[27] There are a few, complex econometric tools that attempt to detect structural change in data (for example, see Paye et al (2005) and Perron (2005)). However, they typically require a lot of data and may have significant errors regarding their identification of a structural break and its timing. Most importantly, their information is not likely to be timely for real time trading decisions as structural breaks will be obvious only after the fact, if at all.

[28] For an excellent, non/technical overview of th dangers when statistical methods play a primary, rather than supplementary role in research, the rader is referred to a book written by one of the best statisticians of his generation – Freedman (2009).

[29] Models with numerous parameters, especially when developed via optimization and other data mining techniques, are particularly vulnerable to such an issue.

relationship adjusted for any cyclicality found this way should be valid across time, instrument types and regions. For example, a manager whose program trades a supposedly powerful, profitable relationship that only applies when trading a single market with many parameters and with a precisely specified, single duration is much riskier than a manager whose program trades a supposedly profitable relationship with few parameters and that is valid for markets globally across multiple durations[30]. The second, quantitative approach involves explicit testing for data snooping, which is a relatively new area for statistics. There are a number of methods of varying degree of complexity that developed over the past fifteen years. Each method has its own benefits and deficiencies, the discussion of which is beyond the scope of this paper[31]. However, regardless of any potential imperfections, the goal of all these methods is sound and invaluable – they are trying to establish if the results developed during model testing on data are indistinguishable from luck and therefore should be ignored. As noted above, adjusted for data snooping bias, many popular trading methods produce at best mixed results, indistinguishable from luck. In my experience, the first, qualitative approach to deal with data snooping is rarely followed and I am yet to meet a manager who is aware of the second approach and its methods and applies these ideas in his research.

DIVERSIFICATION

In modern portfolio management, the purpose of diversification is to improve volatility adjusted return of a portfolio by holding a group of assets that fluctuate in different ways[32]. Simplistically, the more markets traded in a portfolio, the better is the risk adjusted result as long as those markets are at least somewhat different. Capital allocations among markets/strategies traded may be determined based on a manager's judgment or more formally via various optimization routines that incorporate return, volatility and correlation assumptions for the portfolio's components (optimization will be further discussed below).

There are potentially several theoretical and empirical problems affecting performance that an investor should consider for the above approach. First, one must be aware that the traditional diversification approach is not intended to minimize losses during stress or optimize return for each

[30] This second, broader program is preferable even if its profitability is lower than that of the first program as such a program is likely to produce far more sustainable profitability in live trading.
[31] The earliest formal test is White's Realty Check (White (2000)). Addtional proposals that attempt to improve or advance White's ideas are Hansen (2005), Burns (2006), Romano et al (2005), and Hsu et al (2009).
[32] In more academic terms, a portfolio of risky assets will be exposed only to systematic risk that is non - diversifiable, while instrument specific (idiosyncratic) risk will get eliminated (e.g., for a stock portfolio)

unit of tail risk. It is intended to eliminate instrument specific risk in a portfolio of instruments, leaving an investor only with broad macro – based exposures to such risk factors as market, credit, and interest rate risk. Relatedly, the measure of risk is almost always volatility, which, as discussed above, is not designed to measure tail risk. Second, when allocating via optimization methods, one encounters very large model risk due to inconsistent assumptions of such methods with actual market behavior (e.g., stationarity, linearity; see more below) and significant sensitivity of optimization results to small changes in input parameter values. Model risk from the use of such methods can easily overwhelm any potential benefits. Third, when evaluating a potential portfolio of markets traded, one should keep in mind actual capital allocation, instead of focusing on the number of markets that may have a position. For example, it is common for many futures managers to advertise that they may trade fifty or more markets. However, once one looks more closely at the actual capital allocation, one discovers that most of the capital is typically concentrated in just a few of the largest markets among equities, fixed income and foreign exchange (e.g., SP500, Nasdaq, Dax, 10-year German and US bonds, Eurodollars, Euro and British pound, gold, oil), accounting for 50% or more of the overall allocated capital. Such a situation is normally driven by a manager's capacity goals that I will discuss more below. However, as a result of this behavior, the effective capital allocation of a portfolio is exposed to just a few unique return drivers that cannot provide much diversification theoretically or empirically. Finally, empirically, relationships among markets that drive diversification gains within such an industry/academic framework do not behave as expected by this framework (for an introduction, see Munenzon 2010 (a,b)). There is a meaningful clustering of markets during stable market environments and particularly, during stressful periods. For example, during a bull market, it is not crucial which stocks one has in a portfolio as long as there is some meaningful exposure to equities. During bear markets, there are very few places to preserve and enhance capital except with some government bond and rate markets and gold. As a result, with the traditional diversification approach that normally results in relatively stable market allocations across time due to a long historical data set, one is likely to be significantly under-invested in the best performing markets during both bull and bear markets. However, if one applies the traditional diversification method with a relatively short term, rolling window of data, one may be creating a portfolio driven by short term market effects far removed from the longer-term relationships (e.g., traditional allocation to equities in a portfolio will be very significant if based on the five continuous years of a bull market) and further increasing model risk (see walk forward optimization below).

EVENT RISK AND LEVERAGE

For certain types of instruments and strategies (e.g., individual company stocks), one should be mindful of how event risk is addressed as such an event may lead to large price gaps that will not allow stop losses to be enforced even if they are in place. For example, a positive quarterly earnings announcement on a heavily shorted stock that a manager has a short position in may open at a much higher price level, resulting in large losses. Relatedly, how reliable are a manager’s stock loan arrangements so that his borrowed stock is not recalled at the most unfavorable time?

What is the level of leverage in a manager’s strategy and how likely that he will be forced to reduce it by his broker at the worst possible time? This issue is particularly important for many relative value or market neutral strategies in equities or fixed income whose return profile without large amounts of leverage is unattractive for most investors. Therefore, a manager may rely in leverage provided by his brokers to improve returns to meet an investor’s target.

OPERATIONAL RISK

Review of operational risk plays an important role for many systematic managers, especially those that generate many orders due to intra-day trading and/or numerous positions in the portfolio. For such managers, if a trading server does not function, orders will not be properly submitted, resulting not just in potentially foregone profit but also losses as stop losses cannot be enforced. Submission of phone orders may not be practical in high volume trading programs and the manager may not even have a copy of the list of orders for the upcoming day before a trading server failed. As a result, an investor should consider the following questions. Where is a trading server (the server that runs live strategies, updates them with fresh data and submits orders to brokers) located? Is the trading server fully separate from any research or other computational activities of the manager so that live trading is not accidentally affected? If at the manager’s office, what happens if a manager cannot get access to the office building? What is the backup plan if electricity or the internet is down at the building? If a trading server is hosted in a professional facility, how quickly can a backup server be turned on if the primary server is down? Ideally, that backup server is located in a different professional facility in case the original location is affected. If the primary broker connection is down, can a manager execute orders manually or electronically with that broker or another broker? What is the alert system for any technical issues – broker connectivity, internet, server failure, etc. – as one ideally becomes aware of problems immediately so that trading delays are minimized, and backup plans can be activated? Finally, one should consider the security of the code used in live trading, regardless of whether it is deployed on on-premises computers or servers in a hosted location[33]. Typically, passwords will be deployed to access the code. If hacked, the attacker will get access to a fully readable code that might meaningfully diminish the value of the business. Ideally, the code is compiled into an executable file and then deployed for live trading. If such a file is stolen, it will be useless as it is just a stream of 1s and 0s. However, in my experience, the creation and deployment of an executable file is rarely done as many off-the-shelf software programs used by traders for systematic trading do not provide such a capability.

BACKTESTING

Backtesting is the process by which a researcher tests his ideas on historical data with the assumption that successful historical results are likely to be repeated in the future. For example, a researcher may have a trading idea. He then determines the universe of markets to which this idea may apply and after implementing that idea gets historical results as if it were trading live previously. On the basis of such results, a decision is made whether to ignore that idea or to launch it live. There are many potential pitfalls in backtesting methods a manager might use that may produce attractive historical results with little or no relevance for potential live results. As a result, many investors reject all back-tests and focus only on live results[34] before making an allocation decision. I recommend a more balanced approach. A properly designed backtest should allow an investor to invest with a quality, systematic manager whose process is consistent across time much earlier than with a discretionary manager who may typically require years of live results to prove his process. Such earlier entry opportunities should allow not only for more time for profits to accumulate but also for more capital to invest while capacity is still available. Moreover, there is no theoretical or empirical evidence that live

There is no theoretical or empirical evidence that live results are - by definition - more useful than a properly implemented historical test.

[33] IT security professionals joke that there are only two types of users – those who are aware that they were hacked and those that were hacked but are not aware of it.
[34] Using industy jargon, live results are out of sample results, as compared to in-sample or historical results of backtesting.

results are by definition more useful than a properly implemented historical test (see Inoue et al (2002)). The below issues should allow an investor to evaluate the quality of a manager's backtest approach.

First, is backtesting just a confirmation tool for a manager's research ideas or in fact part of the research process? For the latter, a manager might test an idea and then adjust parameters or their values manually or with various optimization tools (more on this topic below) until more attractive results are achieved. In such a case, model risk plays a very significant role that an investor should carefully evaluate, as noted above, because live results can significantly deviate from historical results due to overfitting of models and other problems in backtest to make results look artificially attractive.

Second, one should make sure that the market universe one uses in backtesting is in fact what was possible and available to trade historically, not what is possible to trade now. For example, markets in the current market universe may have been much smaller a few years ago and therefore unsuitable for an institutional program (e.g., see VIX futures daily trading volume from 2007 to the present).

Alternatively, some markets that used to be available for trading and might have been generating signals historically may no longer be available (e.g., a company went bankrupt or was acquired). I once reviewed ten year backtest of a manager focused on systematic trading of multiple Asian futures markets. Results were very attractive. However, until a few years, most of these markets were not possible to trade in any meaningful institutional size and the manager was using historical information on markets with the retail level of liquidity to generate attractive backtest results that have no historical or future relevance for an institutional investor. This issue of a historically meaningful market universe might significantly affect any backtest results if many rapidly growing markets or markets with rapid turnover in its membership (e.g., small cap equities) play a meaningful role in the strategy being tested. High quality, accurate historical data can be expensive depending on the number of markets and time frequency chosen. However, in some cases, it will be impossible to produce a realistic backtest without such an investment.

Third, one should evaluate if realistic execution assumptions (commissions and slippage) are used in backtest. For those programs that trade irregularly and focus on longer term signals, such assumptions are not likely to play a meaningful role in historical or live results. However, for those strategies that trade frequently and focus on short term signals (e.g., intraday models), even small changes in execution assumptions may turn historical profits into losses.

Fourth, capacity analysis should be embedded within backtesting so that a manager has the ability to evaluate market liquidity available at historical periods when a model could have traded live. As noted above, some markets that are large now may have been small a few years ago, making them irrelevant to any institutional trader who traded live at that time. Alternatively, some markets that may have been an important part of the market universe that generate signals may no longer exist (e.g., bankruptcies, takeovers, etc.). These adjustments may require meaningful investment in analytics and data, but such a process will make research results more valid for live trading and capacity targets. The issue of capacity is particularly crucial for high trading volume, shorter term models as with such models, it is very easy to make a large market impact during order execution that will negatively affect performance through very high execution costs. At the very least, an investor should determine how orders are expected to be executed and what percent of daily trading volume orders typically represent (or another time frequency used to execute orders such as 1 hour that is consistent with the duration of signals). Finally, an investor should keep in mind that there is a constant tension between capacity and long-term performance. As the available market universe shrinks due to the inclusion of only the largest markets in the portfolio to maximize capacity, the available profit opportunities shrink as well, especially if a single model is deployed live.

Fifth, optimization tools are used extensively in systematic trading for various purposes, which I will discuss below shortly (for an overview of applications in finance, see Pachmanova et al (2010)). However, initially, it is useful to understand the assumptions of optimization methods that allow them to generate value in certain situations and some of the implications of those assumptions when optimization is used for investment decisions. Optimization theory and methods improved rapidly after WWII with the rise of computational power. Some of the areas that benefit greatly from such methods are logistics and production scheduling. Optimization will find the mathematically best solution in historical data based on a set of criteria a researcher provided (e.g., find the best combination of assets that optimize a portfolio's Sharpe Ratio[35] given the input data). In order for optimization results to relevant for the future, historical data and relationships must be stable or stationery and relationships among items one is optimizing should be linear. If data is stationary, it does not matter what sample of historical data is used in analysis as that historical data will be representative of the future and linearity among items in optimization allows correlation to be utilized in optimization.

[35] Please keep in mind my point above that volatility is not an appropriate metric of risk and as a result Sharpe Ratio cannot be an appropriate measure of a risk adjusted return.

While these assumptions work well when one is trying to find the shortest route between two points (e.g., San Francisco and New York) or to schedule the most cost-efficient production run on an established factory floor (e.g., car manufacturing), they fail in markets. Markets are not stationary as I noted above and as a result, markets regularly experience cyclical and structural changes, which are generally obvious too late and with perfect hindsight. Therefore, we do not know which historical data may prove relevant for the future or even if we now have some historical data with relevance for the future. Relatedly, correlation is a linear tool that cannot capture non – linear nature of relationships among various markets during various environments. There are tools that allow to model non – linear relationships that reflect empirical reality[36] though they are beyond the scope of this paper. However, these tools require sophisticated simulation algorithms that are more complicated than the typical optimization tools available in the industry. Finally, because optimization tools assume that all information is real and not noise, the best solution is highly unstable as any small changes in input data might lead to a very large change in results. In other words, optimization treats all input parameters (e.g., expected returns, volatility and correlations for a group of markets) as known with full certainty, which is far from correct, especially in light of the continuously evolving nature of markets as a complex, adaptive system. Additionally, market data is in fact typically quite noisy and that market noise is not necessarily the type of noise one encounters in physical systems, where the underlying process may be well understood. As a result, well – defined, de-noising tools are not available. As a result, our understanding of the values of input parameters for optimization is far from certain.

Market non-stationarity and noise lead to several, well – known problems of financial optimization: 'corner' solutions and poor out of sample solutions. For example, if during the past few years, a market experienced a particularly strong performance (e.g., SP500), it will be heavily overweight in optimization results at the expense of other markets as this outperformance is mathematically assumed to be permanent – a 'corner' solution. Relatedly, multiple studies demonstrate that equal weight portfolios that have far lower model risk than optimized portfolio outperform optimized portfolio during live investment periods, as one would expect based on the ideas I

Multiple studies demonstrate that equal weight portfolios that have far lower model risk than optimized portfolio outperform optimized portfolio during live investment periods.

[36] Copulas, especially when combined with fat tail distributions. For an accessible introduction, see Alexander (208, Ch.6).

present above (see Jobson et al (1981), Jorion (1985), DeMiguel et al (2009)[37]).

There are some optimization approaches, whose discussion is beyond the scope of this paper, that attempt to address the above issues, though each has its own significant set of limitations at this point[38].

In general, I recommend that one should keep in mind that it is highly unlikely that a high level of quality of results can be reached when applying optimization techniques to a system whose behavior does not reflect basic assumptions of the tools used, especially when combined with problematic criteria like volatility or Sharpe Ratio[39].

Despite the significant limitations noted above on the application of optimization tools to markets, their use is widespread due to the increasing availability of cheap computational power and the proliferation of business and finance graduates trained in these academic methods or those who switch careers to finance, bringing methods from other fields that have no relevance for markets. First, optimization tools play a large role in many data mining techniques managers (particularly with a computer science background) use in research. For example, optimization is important in fitting neural networks to historical data and genetic optimization is used to 'create' historically high performing strategies from an initial pool of candidates (e.g., various technical analysis indicators)[40] .

Once parameters of a model or models are optimized, live trading can begin. Values of parameters may remain fixed or re-optimized at some interval (e.g., daily or monthly) via walk forward optimization (WFO, see below). Second, even without formal data mining techniques, optimization may be used to find optimal values of a parameter or parameters that produce the most favorable historical performance. For example, a manager looking to trade a moving average model[41] will search

[37] Interestingly, the superiority of equal weight allocation seems to apply to indies as well (see Plyakha (2012)).
[38] Bayesian optimization models such as the Black – Litterman optimization allow to incorporate the user's uncertainty of inputs. However, they still rely on correlation metrics, equilibrium concepts and errors in the user's perspective on the quality of his knowledge. Robust optimization techniques adapted from robust statistics attempt to eliminate noise from input parameters before optimization is applied, though uncertainty about the noise process and the linearity assumption of relationships still remain and the overall process is quite computationally complex. Stochastic optimization appropriately attempts to embed uncertainty about parameter values within optimization, modeling multiple probabilistic, future scenarios. However, there is uncertainty about probabilities used in optimization and even now, the computational power required for this method is so significant that it is not practically useful beyond just a few possible markets and future periods. For an overview, the reader is referred to Fabozzi et al (2007).
[39] See above my reservations about volatility and Sharpe Ratio. Also, both metrics can be 'improved' through investments in illiquid instruments - many strategies that invest in illiquid instruments have misleadingly low volatility and high Sharpe Ratio and will receive an extremely high allocation with typical optimization techniques and criteria.
[40] Please note that in both cases, the models a computer chooses to trade live will most likely have no obvious fundamental meaning that a human trader can recognize and will be based on some combination of multiple parameters that proved successful historically without any attention paid to their business logic.
[41] Buy a market if its current price is above its moving average over some period and sell if its current price is below its moving average

over all possible lookback windows until finding the one that optimizes the model's performance criterion such as Sharpe Ratio. Even with a basic computer, such a search will not take a lot of time today. Even more dangerously, he may choose to search multiple parameters combinations. A relatively new technique popular in the trading community is walk forward optimization (WFO) (for example, see Pardo (2008)). With WFO, the key idea is that one should select a window of time (e.g., past five years), optimize the chosen parameters of a strategy within that window, shift that window one day forward with each new, completed trading day and repeat as each new, trading day is completed. The premise behind this method is that instead of fixed parameter values that may prove out of date during live trading, a model adjusts as new data enters optimization. In my view, this approach creates more problems that it attempts to solve. All the previously discussed pitfalls of optimization and data mining remain virtually unaddressed while a new problem of the arbitrarily chosen, rolling window is added to the list of factors affecting model risk. Should we choose past one year, three years or five years for the length of a rolling window? What if the past five years were particularly favorable to risky assets a model favors? Why should we shift the rolling window forward by just one day, instead of multiple days to minimize the impact of market noise on optimization and trading costs? I am yet to find any thoughtful answers to the above questions among traders to use WFO. A more subtle problem is that with WFO, there is no clear separation between historical, in – sample research and live, out – of – sample performance as a model is effectively never completed. As a result, when live performance fails to match historical results for a period of time, it is very hard to determine whether a trading idea is no longer cyclically or structurally valid or a manager's implementation of WFO is invalid in some way (more on the topic of live vs backtest performance below). Finally, optimization may be used to determine parameter values used not just in signal generation but also exit levels. For example, a long trade should be exited if the current price drops 2.2% below the entry price. From this perspective, it is sensible to think that stop losses at the strategy or portfolio levels are unnecessary since both entry and exit signals are optimized on historical data to achieve desired profitability – the perspective I argued above to be inappropriate for market.

EVALUATION OF LIVE PERFORMANCE

Once live trading begins, it is key that the backtest model is not turned off and continues to run indefinitely in parallel with the live model to confirm that there are no inconsistencies between the two. This is particularly crucial if the code used for historical research is not the same code that is used to trade live. For example, high frequency strategies require highly specialized coding skills in C++, a time-inefficient language, researchers may not be well familiar with. As a result, a researcher may initially test ideas in a language that he finds more familiar and convenient such as R and then work with developers to convert his research code into live trading code. Therefore, inconsistencies between live and backtest performance can arise for any number of reasons, such as errors in code migration from research to live trading, inappropriate assumptions for live execution and live data feed errors. Ideally, a manager should test his ideas with the same code that is used in live trading by adding a live data feed as a necessary instead of a historical data feed when research is completed.

However, in some cases, this may not be possible due to legacy issues or diverse preferences of researchers on the team.

Assuming live performance is consistent with backtest performance and a program makes money as expected, there is not much to do except to confirm that the drivers of live profitability are in fact consistent with what was expected from backtesting. However, sooner or later, live performance enters a drawdown phase. If there are no pre – defined limits on losses, it is particularly important to understand the conceptual drivers of losses (not actual positions that contributed to losses), especially when a current drawdown approaches or passes the prior maximum level. Relatedly, one should try to understand if the drawdown is due to a cyclical or a structural market change as with a structural market change, a model will become out – dated by definition and recovery becomes highly unlikely. Crucially, this key step is virtually impossible with strategies where data mining and optimization played meaningful roles as what is traded live may not have any sensible logic that a human being can understand. This situation is further exacerbated by the fact that with any systematic trading program, an outside investor will never have a chance to review the code used to implement a trading idea and therefore catch any mistakes in implementation or methodology. Additionally, as noted above, even if a valid idea is traded live, it may experience extended losses due to a very large model risk that optimization and other statistical methods inconsistent with empirical reality contribute.

At such times, there is generally a great temptation for a manager to make changes to his models to improve performance. An investor should resist such changes unless there is a clear, conceptual understanding of what is being corrected and the underlying need for it. Relatedly, an investor should keep in mind that it is very difficult to conduct quality research when the life of a business is potentially at stake and large losses are already in place, especially if one is trying to change a program's risk management process. Therefore, the most prudent course of action in my view is to have a pre – defined loss limit at the portfolio level, at which point all trading in an investor's account should stop, especially if a manager's program does not have such limits itself. Afterwards, one may take as much time as necessary to review reasons for losses calmly and possibly resume trading at a future point if any concerns about a program's viability are addressed.

Finally, as long as loss thresholds are respected, an investor should make sure that live performance is evaluated over an appropriate length of time to avoid overreaching to data. For example, several years of performance may be necessary for those who trade infrequently and focus on longer term signals whereas just several months of live data may be sufficient for a high volume, intraday trader.

TECHNOLOGY

Technology can be an important contributor to value creation though its potential contribution to performance is hard to quantify and may vary greatly depending on a manager's trading approach. On the one side of the technology spectrum is high frequency trading that requires highly specialized coding and hardware skills. On the other side is a trader who trades a single market like SP500 once a day using a model implemented in Excel. I recommend that an investor spends some time trying to understand that the technology a manager uses to research and implement trading ideas is consistent with the manager's trading approach and is scalable. For example, if a manager claims to have a unique approach for evaluating options across multiple markets and then trading them regularly, including active intra – day re-balancing, one should be concerned if a manager says that all of that occurs through Excel - software that has difficulties with large data sets and complex customizations. Relatedly, some platforms have their inherent limitations of what can and cannot be done in them, especially as almost all of them are closed source and therefore cannot be easily or cost effectively customized by anyone except the company owning the software.

For example, though Matlab is popular among many quantitative researchers and can support a broad range of models of almost any degree of complexity, it can track live data for only a few markets at a time as its architecture was never intended for the purpose of enterprise grade, systematic trading. Larger capacity would require a significant investment in hardware infrastructure and additional software tools. Relatedly, meaningful changes in research code created within Matlab may be required before it can go live within Matlab, which might create problems later once live trading begins, as noted above. Additionally, ideally, a manager should not use multiple platforms for various programs (e.g., different software packages are used for different types of instruments or different researchers work in different platforms and send their signals via a standard protocol to the overall portfolio). Comparability of research results and preparation for live trading may be affected and integration of signals from multiple software platforms in live trading is a non – trivial task that increases the risk of poor execution. Additionally, an investor should consider the potential impact of software ownership. If a manager uses off – the – shelf software and that company goes out of business or does not appropriately support the product in the future, how quickly can the manager replicate his ideas elsewhere? If technology plays a key role in a manager's process, one should expect that manager to gradually develop a fully customized solution to maintain his edge as by definition that edge will be hard to preserve if everyone has access to the same platform.

ORGANIZATIONAL STRUCTURE

A number of organizational issues can serve as important signals of the long-term potential of a systematic trading program as they help determine the long-term path of the business that in turn affects a manager's focus on long term performance for investors. First, how is a manager's investable capital invested? In my view, if a manager truly thinks that he is doing something special, one should expect the majority of his investable net worth to be invested in his trading program.

Second, what are his incentives? Do management fees play an important role? If so, sooner or later, it is highly likely that the focus of the organization will shift to asset gathering and relative rather than absolute performance (regardless of what the manager may claim as of now) as it is much easier to make money from management fees than performance fees. Relatedly, as discussed above, what are his capacity targets? Are they sensible in light of what and how he plans to trade? Is there a pre – existing commitment to close a program at some level of assets? One should note that a manager's fee structure will have a meaningful impact on his decision about capacity. There is tension not just between capacity and long-term performance I discussed above but also between capacity and a manager's fee structure as a focus on management fees tend to lead to unjustified capacity targets, impacting long term performance. Third, who are his investors? Is he open to fund of funds money that is typically unstable and may negatively impact other investors in the strategy due to its frequent moves in and out of the program? Fourth, are there large, fixed expenses outside of the value - adding activities of research, technology and data? Is there flexibility in expenditures on research, technology and data? Constant pressure to pay for high level of fixed expenses may force a manager to favor short term signals and performance as there is no financial flexibility to take a longer-term view, to focus on management fees at the cost of performance and to attract investors of any level of quality.

PART III - CONCLUSION

If the reader reaches this point, I hope it is not with the overwhelming sense of how risky systematic trading is, which may only be re – enforcing his prior conviction to stay away from this space entirely. Rather, a very large number of options to add or destroy value available to any researcher in systematic trading presents numerous opportunities to generate sustainable profitability for investors.

In my view, as with any type of a manager, the key goal of all investors researching a systematic trading program for a potential allocation should be to understand core assumptions a manager made when thinking about the world and implementing his program - ultimately, an investor is buying a process that should be thoughtful, consistent and scalable, not a specific trade. Any systematic program may have embedded biases of a developer that may negatively impact performance as much as what one would experience with a human trader. However, those systematic programs that can consistently optimize their behavioral and analytical edges over other market participants should prove highly attractive and should always have a role in an investor's portfolio.

REFERENCES

Alexander C.. Market Risk Analysis, Practical Financial Econometrics, Volume II. 2008. Wiley.

Artzner P., Delbaen F., Eber J. M., and Heath D.. Thinking Coherently. 1997. Risk 10 (11), 68 – 71.

Artzner P., Delbaen F., Eber J. M., and Heath D.. Coherent Measures Of Risk. 1999. Mathematical

Finance, 9, 203 – 228.

Bak P.. How Nature Works: The Science Of Self – Organized Criticality. 1999. Copernicus.

Burns P.. Random Portfolios For Evaluating Trading Strategies. 2006. Working Paper.

DeMiguel V., Garlappi L. and Uppal R.. Optimal Versus Naïve Diversification: How Inefficient Is 1/N Portfolio Strategy? 2009. Review Of Financial Studies, 22, 5, pp. 1915- 1953.

Eagle A.. Philosophy Of Probability: Contemporary Readings. 2011. Routledge.

Embrechts P., Resnick S. I., and Samorodnitskiy G.. Extreme Value Theory As A Risk Management Tool.

1999. North American Actuarial Journal, 3 (2), 32 – 41.

Fabozzi F., Kolm P., Pachmanova D. and Focardi S.. Robust Portfolio Optimization And Management.

2007. Wiley.

Fabozzi F., Focardi S. and Kolm P.. Quantitative Equity Investing: Techniques And Strategies. 2010. Wiley.

Fama E.. The Behavior Of Stock Market Prices. 1965. Journal Of Business, 38, 34 – 105.

Freedman D.. Statistical Models And Causal Inference: A Dialogue With The Social Sciences. 2009. Cambridge University Press.

Gillies D.. Philosophical Theories Of Probability. 2000. Routledge.

Granger C. and Poon S.-H.. Forecasting Volatility In Financial Markets: A Review. 2002. Working Paper.

Hansen, P. R.. A Test For Superior Predictive Ability. 2005. Journal Of Business And Economic Statistics, 23, pp. 365 – 380.

Hastie T., Tibshirani R. and Friedman J.. The Elements Of Statistical Learning: Data Mining, Inference, and Prediction. 2011. Springer.

Hsu P.-H., Hsu Y.-H. and Kuan C.-M.. Testing The Predictive Ability Of Technical Analysis Using A New Stepwise Test Without Data Snooping Bias. 2009. Working Paper.

Jobson J. D. and Korkie B. M.. Putting Markowitz Theory To Work. Summer 1981. Journal of Portfolio

Management 7, pp. 70 – 74.

Johnson N.. Simply Complexity. 2010. Oneworld Publications.

Jorion P.. International Portfolio Diversification With Estimation Risk. 1985. Journal Of Business, 58, pp. 259 – 277.

Johnson N., Jefferies P. and Hui P.M.. Financial Market Complexity: What Physics Can Tell Us About

Market Behavior. 2003. Oxford University Press.

Inoue A. and Kilian L.. In – Sample Or Out – Of – Sample Tests Of Predicatability: Which Should One Use? 2002. ECB Working Paper No. 195.

Kahneman D., Slovic P. and Tversky A.. Judgement Under Uncertainty. 1982. Cambridge University Press.

Kahneman D. and Tversky A.. Choices, Values And Frames. 2000. Cambridge University Press. Kindleberger C. and Aliber R.. Manias, Panic And Crashes: A History Of Financial Crises. 2005. Wiley.

Kirkpatrick C. and Dahlquist J.. Technical Analysis: The Complete Resource For Financial Market

Technicians. 2010. FT Press.

Lo A.. The Adaptive Market Hypothesis: Market Efficiency From An Evolutionary Perspective. 2004. Working Paper.

Lowenstein R.. When Genius Failed: The Rise And Fall Of Long Term Capital Management. 2001. Random House.

MacKay C.. Extraordinary Popular Delusions And The Madness Of Crowds. 2013. CreateSpace. Marshall B., Cahan R. and Cahan J.. Technical Analysis Around The World. 2010. Working Paper.

Kronenbruger J. and Sebeson J.. Analog And Digital Signal Processing: An Integrated Computational

Approach With Matlab. 2008. Thomson Delmar Learning.

Mandelbrot B.. The Variation Of Certain Speculative Prices. 1963. Journal of Business 36, 394 – 419. Mandelbrot B. and Hudson R.. The Misbehavior Of Markets: A Fractal View of Financial Turbulence.

2006. Basic Books.

Mantegna R. and Stanley E.. Introductions To Econophysics: Correlations And Complexity in Finance.

2007. Cambridge University Press.

Mauboussin M.. Think Twice: Harnessing The Power Of Counterintuition. 2012. Harvard Business

Review Press.

Munenzon M.. 20 Years Of VIX: Fear, Greed and Implications For Traditional Asset Classes. 2010a. Working Paper.

Munenzon M.. 20 Years Of VIX: Fear, Greed and Implications For Alternative Investment Strategies.

2010b. Working Paper.

Munenzon M.. Risk Management From Theory to Practice: Is Your Risk Metric Coherent And

Emprirically Justified? 2010c. Working Paper.

Pachmanova, D. and Fabozzi, F.. Simulation And Optimization In Finance. 2010. Wiley.

Palmer T. and Hagedorn R.. Predictability Of Weather and Climate. 2006. Cambridge University Press.

Pardo R.. The Evaluation And Optimization Of Trading Strategies. 2008. Wiley.

Park I. and Irwin S.. The Profitability Of Technical Analysis: A Review. 2004. Working Paper.

Paye B. and Timmermann A.. Instability Of Return Prediction Models. 2005. Working Paper.

Perron P.. Dealing With Structural Breaks. 2005. Working Paper.

Plyakha Y., Uppal R. and Vilkov G.. Why Does An Equal - Weighted Portfolio Outperform Value – and Price – Weighted Portfolios? March 2012. EDEC Working Paper.

Romano, J. P. and Wolf M.. Stepwise Multiple Testing As Formalized Data Snooping. Econometrica, 73, pp. 1237 – 1282.

Sewell M.. Characterization Of Financial Time Series. 2011. Working Paper.

Sornette D.. Why Stock Markets Crash: Critical Events In Complex Financial Systems. 2004. Princeton

University Press.

Surowiecki J.. The Wisdom of Crowds. 2005. Anchor.

Taleb N.. Fooled by Randomness: The Hidden Role of Chance In Life And In The Markets. 2005. Random House.

Taleb N.. Black Swan 2nd Edition: The Impact Of The High Improbable. 2010. Random House.

Taleb N.. Antifragile: Things That Gain From Disorder. 2014. Random House.

Voit J.. The Statistical Mechanics of Financial Markets. 2005. Springer.

Watts D.. Everything Is Obvious: How Common Sense Fails Us. 2012. Crown Business.

White, H.. A Reality Check For Data Snooping. 2000.

Econometrica, 68, pp. 1097 – 1126

ABOUT THE AUTHOR

Mikhail Munenzon is an experienced finance and investment professional. He holds qualifications as a Chartered Alternative Investment Analyst (CAIA), Chartered Financial Analyst (CFA) and as a Professional Risk Manager (PRM).

Today (2021) he works as a director at Valentiam Group in Morristown, New Jersey, USA. Valentiam is part of WTS Global, a leading global tax firm with offices in more than 100 countries.

PUBLISHER

Intalcon GmbH

Wilhelm-Herbst-Str. 7

28359 Bremen

Germany

www.intalcon.com