

Risk-adjusted portfolio optimisation using a parallel multi-objective evolutionary algorithm

Phil Maguire, Dónal O’Sullivan, Philippe Moser and Gavin Dunne

Abstract—In this article we describe the use of a multi-objective evolutionary algorithm for portfolio optimisation based on historical data for the S&P 500. Portfolio optimisation seeks to identify manageable investments that provide a high expected return with relatively low risk. We developed a set of metrics for qualifying the risk/return characteristics of a portfolio’s historical performance and combined this with an island model genetic algorithm to identify optimised portfolios. The algorithm was successful in selecting investment strategies with high returns and relatively low volatility. However, although these solutions performed well on historical data, they were not predictive of future returns, with optimised portfolios failing to perform above chance. The implications of these findings are discussed.

I. INTRODUCTION

The goal of an investor is to allocate their assets optimally in a portfolio so that they maximise the returns on their investment. The basic theory of portfolio optimisation was set out by Markowitz in 1952 [1]. His principle argument was that investors need to balance the objective of maximising return with the risk involved, meaning that risky investments should provide higher returns. Markowitz proposed that investors should select stocks, not only based on their individual profiles, but on how they move together. Portfolios that take co-movements into account will have a lower overall risk profile, because a set of diversified uncorrelated stocks are less likely to all move in the same direction at the same time [2]. In contrast, including similar stocks in a portfolio is risky because they have the tendency to fluctuate in the same direction, thus increasing the overall volatility of the portfolio.

II. EVALUATING PORTFOLIO PERFORMANCE

Standard practice for evaluating portfolio performance and estimating future returns is to examine historical returns and adjust for interest rates [3]. Historical averages present a pertinent statistic for asset evaluation because they provide unbiased estimates of future returns [4]. As a result, portfolio testing typically involves the use of historical returns as a proxy for expected returns [5].

Assuming estimates for expected returns are available, then Markowitz’s model [1] can be used to maximise return and minimise risk for a set of investment choices. However,

All authors are with the Department of Computer Science, National University of Ireland, Maynooth, Ireland. Correspondence should be addressed to Phil Maguire (phone: 353-1-7086082; fax: 353-1-7083848; e-mail: pmaguire@cs.nuim.ie).

This work was supported in part by the Irish Research Council for Science, Engineering and Technology through an Empower Research Fellowship awarded to the first author.

this model is theoretical in focus, and impractical for dealing with large numbers of assets [6]. It uses only mean and variance to describe asset characteristics and is founded on the assumption that returns are normally distributed, a condition which is rarely satisfied in practice [7]. In addition, the Markowitz model ignores most of the constraints faced by investors [8]. Alternative techniques are required to deliver high-quality solutions for real-world investment scenarios.

In the following study, our strategy was to use risk-return characteristics to construct portfolios with the strongest investment potential. A key question which arose was how the historical performance of a portfolio should be evaluated. According to Modigliani and Modigliani [9], the financial industry continues to rely almost exclusively on total return for evaluating the historical performance of a portfolio, despite the fact that this value completely ignores risk. Consider Figure 1 below in which two time series are presented. In both cases the portfolios yield an identical return of 12%.

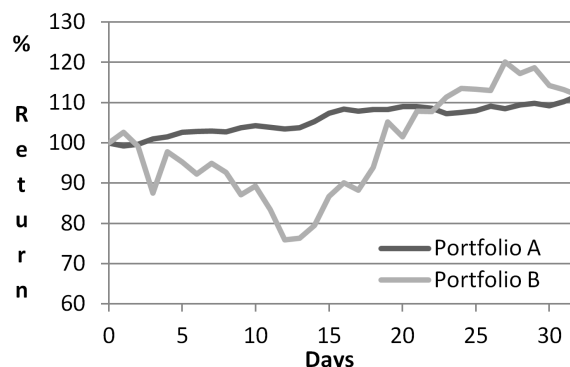


Fig. 1. Two portfolios yielding a return of 12%.

However, portfolio B has a higher level of volatility, which means that the returns earned in this case are more likely to represent random variance, as opposed to an underlying performance signal: the higher the standard deviation in daily returns, the more likely it is that a portfolio will deliver a given return by chance alone. In the case of Portfolio B, the return is insignificant, because it does not strongly challenge the random walk hypothesis. In the case of Portfolio A, the lower volatility means that the return is less likely to have been produced by chance, and thus more likely to persist into the future.

Anybody walking into a casino has the opportunity to trade capital for potential returns. For example, a \$1 note can be exchanged for a 10% chance of holding \$10 or a

1% chance of holding \$100 (assuming the casino offers fair odds). Accordingly, the question that should be asked when a portfolio delivers a particular return is whether that return is over and above that which could have been achieved by simply gambling: that is, whether the returns exceed the cost of holding risk [9]. In sum, reliance on simple returns can be misleading. What is needed is a rigorous means of adjusting portfolio returns for the risk involved, so that different portfolios can be directly compared.

A. Sharpe Ratio

The most commonly used measure of risk-adjusted return is the Sharpe Ratio (SR), which takes into account the ratio between the reward and the price variability to which the investor was exposed. Here, the risk free rate (the amount that could have been earned without holding any risk) is subtracted from the total returns, and the remainder is divided by the portfolio's standard deviation, thus effectively providing a measure of reward per unit of risk [10].

Given a portfolio yielding an overall return R and a risk-free return R_f , with a sampled standard deviation s , then

$$SR = \frac{R - R_f}{s}.$$

The SR always refers to the differential between two portfolios, in this case that between the risky investment and the risk-free investment. The subtracted return reflects the short position which must be taken to finance the acquisition. To yield a positive SR, an investment must provide a greater return than the cash or loan which is used to fund it [11].

B. Limitations of Sharpe Ratio

Exclusive reliance on the SR to differentiate between the historical performance of different portfolios is problematic. Over a given time period, some stocks will produce exceptional growth which is unsustainable in the longer run. These exceptional increases may be due to fortuitous events, such as a sudden technological breakthrough or the growth of a speculative bubble. In particular, the more stocks that are considered, the more likely it is that some will exhibit unsustainable growth.

The problem with the SR metric is that strong growth hides strong volatility, meaning that when short-term growth subsides, an investment which previously appeared desirable will have high risk without the compensating returns. For example, for the period analysed in our study (July 2008 to June 2011), the S&P 500 listed company Netflix registered an exceptional 1500% increase, eclipsing all other potential investments. This kind of growth is clearly unsustainable. Indeed, the Netflix stock suffered a catastrophic 65% plunge in its value in the three months that followed.

The SR fails to discriminate between fortuitous short-term growth and robust growth founded on sustainable, persistent conditions. The danger of relying on this metric to optimise portfolios to historical data is that it tends to favour a small set of high growth, high volatility stocks, whose strong performance is unlikely to be sustained (see

[12]). Accordingly, the SR should not be relied on as a sole measure of risk.

Indeed, there is no single metric that is the absolute correct metric to use for assessing portfolio risk. All metrics involve simplifications. Constructing an optimised portfolio based on any single metric is not advisable, because the portfolios that emerge are likely to be those that exploit weaknesses and simplifications in the metric, as opposed to exhibiting a genuine underlying performance signal. If the environment changes, the context for narrowly optimised solutions will be altered, and low risk solutions may become high risk [13].

Ideally, to enhance robustness and avoid overfitting, portfolios should be optimised to satisfy the requirements of multiple diverse risk measures. In light of this, we developed a multi-objective evolutionary algorithm with parallel island populations fitted to a selection of complementary metrics. In the following section we describe the metrics used, and, subsequently, the structure of the evolutionary algorithm.

III. METRICS

We developed four different metrics, each responding to different aspects of a portfolio's risk to return profile (see Table II). The first metric is the SR, which provides an overall risk to reward ratio.

To compensate for the SR's bias towards unsustainable high-growth investments, the second metric is more risk averse. In this case, the portfolio's risk-free adjusted return is divided by its variance, that is, the square of the sampled standard deviation. This metric penalises volatility more heavily, thus favouring lower growth portfolios which are more likely to exhibit sustainable growth.

One problem associated with both of these adjusted-risk metrics is that they ignore any deviation in returns over the time period. If we consider the two portfolios in Figure 2 we see that both have identical returns over the period. Portfolio B actually has the higher standard deviation ($s = 1.36$ versus 1.33), meaning that portfolio A has a higher SR. However, to the human eye portfolio A seems more risky, because it involves a strong upwards trend followed by a strong downwards trend. These trends indicate that portfolio B is responding to changeable short-term conditions, limiting the accuracy with which its future performance can be predicted. On the other hand, the growth evident in portfolio B is more robust, suggesting that the set of supporting conditions is more likely to persist into the future.

Because the SR is unaffected by any variability in the consistency of returns, it overlooks an important element of risk. To compensate for this we included two additional metrics which take into account the distribution of returns.

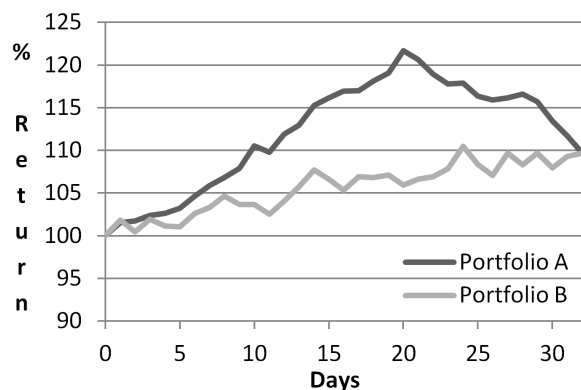


Fig. 2. Two portfolios yielding a return of 9%.

Metric three measures the greatest fall between any two successive points in a portfolio time series, adjusted for risk-free return. This measure is expressed as the maximum percentage that an investor would have lost if investing in a portfolio at the worst possible moment during the time period. In contrast to the first two metrics, the objective is to minimise this value. Portfolios with high SRs derived from lucky high-risk investments are more likely to incur sustained falls than those exhibiting consistent growth.

Metric four is similar, but instead of measuring the maximum fall, it measures the maximum amount of time for which a fall has been recorded, again adjusted for risk-free return. This value represents the maximum amount of time (expressed as a percentage of the total time period) that an investor could have waited before seeing a positive return on their investment, if they had invested at the worst possible moment. Again, the objective is to minimise this value. Portfolios with steadily increasing value will contain shorter time periods where investors could be down on their investment, whereas high-risk investments result in much longer periods of negative return (see Figure 2).

The formulas for these metrics are given in Table II where R is the return on the portfolio investment, R_f is the return on a risk-free investment over the same time period, x is the set of daily portfolio returns expressed relative to the risk-free rate, and s_x is the sampled standard deviation for that time series. To put these metrics into context we computed their values for portfolios A and B as shown in Figure 2.

	Portfolio A	Portfolio B
Metric 1	7.230	7.108
Metric 2	5.401	5.221
Metric 3	0.098	0.031
Metric 4	0.352	0.235

TABLE I
RISK METRIC VALUES FOR PORTFOLIOS IN FIGURE 2.

The values for metrics 1 and 2 indicate that the overall returns are about seven times greater than the standard deviation of the daily returns, and five times greater than the variance. The values for metric 3 indicate that the biggest fall

between any two points is 9.8% for portfolio A and 3.1% for portfolio B. The values for metric 4 indicate that the longest period without a new record high is 35.2% of the time period for portfolio A and 23.5% for portfolio B.

In sum, while metrics 1 and 2 reveal that both portfolios have experienced strong growth, they fail to distinguish that portfolio A's value has exhibited greater overall fluctuation. And while metrics 3 and 4 reveal that the value of portfolio B has grown more consistently, they do not communicate anything about the overall level of growth. A portfolio that can simultaneously satisfy all of these metrics will exhibit strong, steady growth and consistently low volatility.

IV. EVOLUTIONARY ALGORITHM

Evolutionary algorithms are a computational technique inspired by Darwin's theory of evolution and first popularised by Holland [14]. The idea is that if the fittest individuals in a pool of candidate solutions are continually selected to produce the next generation, then the population is guided towards regions of the search space with good solutions.

Each iteration of the evolutionary algorithm involves a competitive selection process that gradually weeds out poorer solutions. A small amount of mutation is also included to mitigate against premature loss of diversity and stagnation on restricted regions of the search-space.

	Concept	Formula	Range
Metric 1	Return-to-standard deviation ratio	$\frac{R - R_f}{s_x}$	> 0 indicates positive return, < 0 indicates loss
Metric 2	Return-to-variance ratio	$\frac{R - R_f}{s_x^2}$	> 0 indicates positive return, < 0 indicates loss
Metric 3	Maximum fall	$\max_{i < j} (x_i - x_j)$	0% indicates no daily falls, 100% indicates total loss
Metric 4	Maximum duration of any fall	$\max_i (\max_j \{x_{i+1} < x_i, \dots, x_{i+j} \leq x_i\})$	0% indicates no daily falls, 100% indicates investment never recovers first day value

TABLE II
DESCRIPTION OF RISK METRICS.

Evolutionary algorithms can quickly identify good solutions to optimisation problems and have been successfully applied to a number of areas in finance. For example, Allen and Karjalainen [15] used genetic programming to derive trading rules for the S&P 500 while Neely, Weller and Dittmar [16] used evolutionary algorithms to develop trading rules for foreign exchange markets (see [17] for a review of evolutionary algorithms used in financial modelling).

A. Multi-objective optimisation

Of the four metrics outlined in the previous section, none provides a perfect assessment of portfolio performance. All

involve simplifications, and, as a result, solutions fitted to one metric in particular are likely to exploit weaknesses in the metric, as opposed to exhibiting genuine performance. Each of the metrics relate to different aspects of risk, from standard deviation, to variance, to maximum falls, to frequency of record highs. The mark of a robust solution is that it should rank consistently highly on all of these related measures. Multi-objective optimisation algorithms attempt to solve problems that have multiple overlapping or possibly competing objectives [13]. Such algorithms have been applied in several areas of finance including, risk management, and portfolio management (see [18]). In particular, multi-objective algorithms have been found to produce well-balanced portfolio performance, outperforming any single fitness algorithm [19]. An intuitive strategy for solving multi-objective problems is to construct a single aggregate objective function which combines the different objectives into a single formula, though the use of weightings [20]. However, doing so simply creates a more elaborate unitary metric, which is again susceptible to over-fitting. Our solution for maintaining genetic diversity was to employ an island population paradigm, with multiple genetic algorithms operating in parallel and a migration process operating between islands.

B. Parallel evolutionary algorithms

Parallel genetic algorithms have been reported to yield better performance than genetic algorithms which rely on a single panmictic population [21]. Having multiple sub-populations on separate ‘islands’ helps to preserve genetic diversity, as it allows each island to explore a different area in the solution space, mimicking the role of locality on biodiversity. In our evolutionary algorithm, each island operates a different fitness function corresponding to one of the risk metrics in Table II. After a certain number of generations, the fittest solutions are taken from each island, and this new elite population is used to repopulate all of the islands. The advantage of this system is that it preserves genetic diversity while at the same time favouring robust solutions that satisfy the requirements of a range of risk metrics.

C. Algorithm Structure

Solutions are modelled as a series of weights, reflecting the percentage allocation of investment for each of the stocks in the S&P 500. Weights can either be positive or negative, allowing the portfolio to be long or short on any stock. The solution string is a series of 500 weights, the absolute values of which are normalised to total 100%. To initialise the population, solutions are created for each island using randomised weights.

We experimented with several techniques for selection, including fitness proportionate selection and tournament selection. Tournament selection was identified as being the least likely to lead to stagnation. Using this technique, a fixed-size random sample is selected from the current population and the fittest pair of solutions in the sample is recombined to produce an individual in the next generation. The process

is repeated until a full population has been created from the current generation (see [22]). The selection pressure of tournament selection is directly linked to the size of the tournament: the larger the size, the less likely it is that unfit solutions will contribute to the next generation.

For the recombination process, we experimented with averaging the weights between the two parents and random point crossover. Random point crossover was found to be less likely to lead to stagnation. Using this technique, two random positions are chosen along which the parent solution strings are divided. One portion of the string from one parent is then joined with the portion of the string from the other parent, creating a new individual with features of both parents.

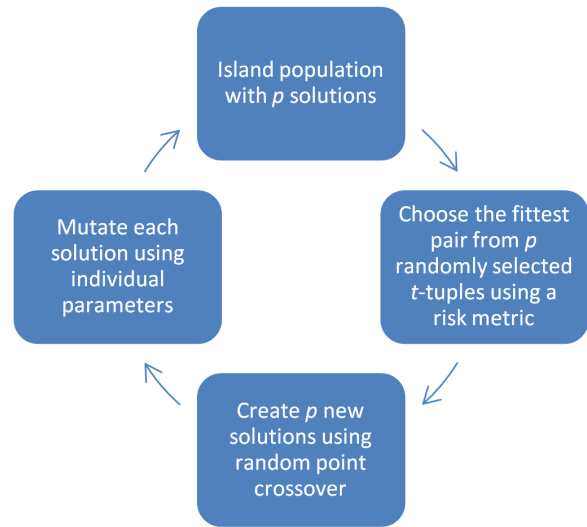


Fig. 3. Selection and recombination processes.

The mutation rate of a genetic algorithm is often handled as a global, external parameter which remains constant over time. However, it may be desirable to reduce the number of external parameters of a genetic algorithm, so that the range of the search space and the manner in which it is searched are not artificially constrained (see [23]).

Accordingly, three mutation parameters are included as part of the solution string, which are themselves subject to the mutation process. The first parameter encodes the percentage probability of a solution being modified. The second parameter encodes the percentage probability of each individual weight in the solution being modified, while the third encodes the absolute magnitude of the random mutation.

A final component included as part of the genetic algorithm is a process for eliminating insignificant weights. Each additional investment included in a portfolio incurs a cost to actively trade it. To make the portfolio simpler and more manageable, the algorithm zeroes any investment weightings of less than 0.2% (i.e. one five hundredth).

D. Evolutionary Process

A set of genetic algorithms are run in parallel, with each island population being fitted to one of the four risk

metrics. After a set number of generations, the migration process is triggered, whereby the fittest solutions on each island are gathered together and this pool is used to re-seed the populations of all of the islands, thereby enhancing genetic diversity. Diagrams of the selection, recombination and migration processes are shown in Figures 3 and 4.

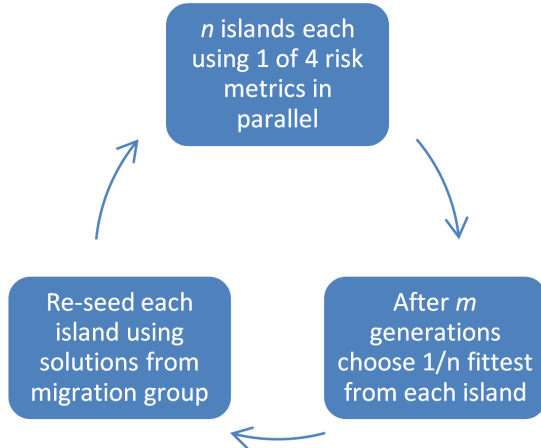


Fig. 4. Migration process.

Once the genetic algorithm has completed a specified number of generations, the fittest individual for each of the four metrics is identified and a voting process is used to select an overall champion. Each solution receives 4 points for being ranked in first place by one of the metrics, 3 points for second, 2 points for third and a single point for being ranked in last place. The solution gaining the highest number of votes is chosen by the algorithm as the overall fittest solution.

The input parameters to the program follow the syntax $\langle\langle P \rangle\rangle, \langle\langle T \rangle\rangle, \langle\langle G \rangle\rangle, \langle\langle N \rangle\rangle, \langle\langle M \rangle\rangle$, where P is the total seed population across all of the islands, T is the tournament size, G is the total number of generations, N is the number of islands and M is the number of generations between migrations.

V. FINANCIAL DATA

The evolutionary algorithm was written in Python and processed daily changes from the S&P 500 companies, downloaded from the Yahoo! Finance website. The S&P (Standards and Poor's) 500 is an index of the stocks of the top 500 publicly held companies that trade in either the New York Stock Exchange or the NASDAQ, the two largest stock exchanges by market capitalisation in the world.

In this study we used three years of daily price changes, with the training data beginning on July 22nd 2008 and finishing on January 1st 2011, and the test data covering the period from January 1st to May 31st 2011. This sample happens to coincide with the credit crunch and global financial crisis which ensued following the collapse of Lehman Brothers, which filed for bankruptcy on September 15th 2008. The goal of the study was to investigate whether a portfolio identified based on 30 months of historical data

could be held for the subsequent 5 months and be expected to generate a profit above the risk-free investment rate. This risk-free rate was defined as the returns on short-term US Treasury bills for the period.

In conventional single period portfolio optimisation, portfolios are allocated for a single upcoming period (see [1]). In multi-period optimisation, the portfolio is rebalanced to a specified allocation at the end of each period, a strategy known as Constant Proportion (CP), or Constant Ratio Asset Allocation (CRAAL) (see [24]). This maintains the relative contribution of each asset in the portfolio. In light of the length of the training period, we applied daily re-balancing to prevent stocks with large increases from dominating the portfolios.

VI. RESULTS AND DISCUSSION

The genetic algorithms were successful in identifying portfolios which, in hindsight, delivered outstanding performance. Had these solutions been available at the start of the training period, they could have been leveraged against the risk-free borrowing rate to provide very significant returns (see Figure 5).

Evaluated By	Generated By			
	Metric 1	Metric 2	Metric 3	Metric 4
Metric 1	0.133	0.508	.064	.174
Metric 2	0.063	0.291	.069	.301
Metric 3	0.011	0.058	.018	.326
Metric 4	-0.006	-0.049	.084	.356
Parallel	0.127	0.600	.040	.135

TABLE III

AVERAGE METRIC VALUES FOR 10 RUNS OF THE GENETIC ALGORITHMS ON HISTORICAL DATA USING $\langle\langle P \rangle\rangle, \langle\langle T \rangle\rangle, \langle\langle G \rangle\rangle$ AS (1000,5,100) FOR THE INDIVIDUAL METRIC ALGORITHMS AND $\langle\langle P \rangle\rangle, \langle\langle T \rangle\rangle, \langle\langle G \rangle\rangle, \langle\langle N \rangle\rangle, \langle\langle M \rangle\rangle$ AS (1000, 5, 100, 4, 25) FOR THE MULTI-OBJECTIVE ISLAND ALGORITHM.

An overview of the performance of the various evolutionary algorithms, in isolation and in parallel is given in Table III, as averaged over 10 runs. Each row refers to the average of the top solutions produced by a specific fitnessfunction and each column involves a different metric for evaluating the performance of those solutions.

The results reveal that the parallel islands model was extremely successful in enhancing the quality of the solutions produced. The migratory system reduced the probability of premature stagnation, and the use of multiple risk metrics operating in parallel broadened the search space and facilitated the identification of more robust solutions. On average, island solutions had a SR that was nearly as good as solutions fitted specifically to this metric (0.127 versus 0.133 respectively). They also had the best average value for metrics 2 and 4 and the second best for metric 3. Interestingly, the solutions optimised to metrics 2 and 4 did not, on average, produce the fittest solutions for those metrics. This observation highlights the importance of maintaining diversity in the population pool and provides further support for the effectiveness of multi-objective optimisation [19].

One of the central questions this study seeks to answer is how well portfolio solutions optimised to fit historical data perform on future data. We applied the solutions generated by the islands algorithm to the test data, which covered the subsequent five months of price changes. Performance above the risk-free rate here would indicate that portfolios optimised to historical data can be predictive of future returns. Table IV shows the average performance on the training and test data for the same 10 islands solutions used in Table III.

	Metric 1	Metric 2	Metric 3	Metric 4
Training	0.127	0.600	.040	.135
Test	-0.117	-0.712	.031	.871

TABLE IV
AVERAGE PERFORMANCE ON TEST DATA OF THE FITTEST SOLUTIONS
GENERATED BY 10 RUNS OF THE PARALLEL EVOLUTIONARY
ALGORITHM.

In total, 8 of the 10 solutions generated negative returns relative to the risk-free rate, resulting in a negative average value for metrics 1 and 2. Because of these negative returns, the maximum fall duration covered a significant portion of the test period, resulting in a high average value for metric 4. The low average value for metric 3 reflects the fact that, in general, the investments flat-lined during the test period. There was no significant correlation for any metric between training and test performance values. These results clearly reveal that portfolios optimised to historical data fail to sustain their performance into the future.

Figure 5 charts the performance of the fittest solution identified for each metric. In every case, the value of the investment can be seen to markedly flat-line at the transition point. Because these portfolios include both long and short investments in the S&P 500 companies, they are mostly dollar neutral. Solutions consist of long investments on stocks which steadily increased over the training period and short investments on stocks which steadily decreased. These trends have failed to continue into the test period, with the result that the portfolios have reverted to neutral price fluctuations, failing to match the risk-free rate and pushing the portfolios' SRs into negative territory. Interestingly, the average daily volatility decreased from 0.39 during the training period to 0.22 during the test period, reflecting the transition from strong growth to stability.

The consistency in daily volatility between training and test periods indicates that risk can be successfully reduced by diversifying asset allocation: a diversified portfolio will remain diversified because the characteristics which describe the relationships between stock movements are enduring. However, the results bring into question the assumption that portfolios can be optimised for future growth. As evidence by Figure 5, The historical performance identified by the genetic algorithms is fluky as opposed to representing a persistent characteristic of the portfolios generated.

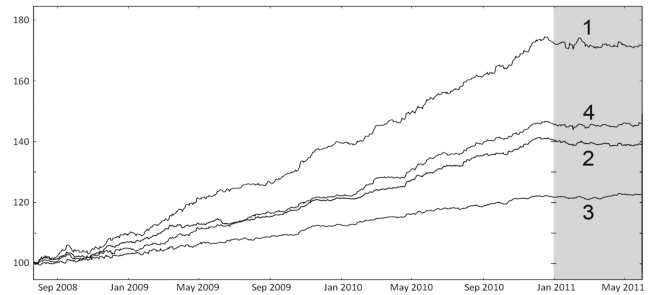


Fig. 5. Performance of the fittest solutions identified for metrics 1, 2, 3 and 4.

The principle contributor to this overfitting is the large number of free parameters that the solution strings can avail of. For instance, the portfolios evaluated in Table IV contain an average of 303 weighted investments. The more free parameters available, the greater the extent to which random noise can be manipulated to produce an apparent pattern.

Consider, for example, a situation where a set of time series is generated by a random walk signal. The more free parameters that are available, the more successfully these random signals can be combined in a portfolio to produce what appears to be a steadily increasing signal. Yet, because the constituent components are entirely random, the portfolio's performance on future data will flat-line, following the law of large numbers [25].

Effectively, it is the choice of free parameters that is doing the work of creating the pattern, as opposed an intrinsic signal in the constituent time series. The more free parameters that are available, the less significant is the identification of a particular pattern.

The problem with the genetic algorithms used in this study is that they fail to factor in how the number of free parameters available diminishes the significance of the resulting solutions' performance. As a result, it is not certain that the quality of the solutions exceeds that which could be achieved through the manipulation of random signals.

VII. GENERAL DISCUSSION

The efficient-market hypothesis [26] proposes that the instantaneous pricing of a security fully incorporates the expectations of all market participants and hence any subsequent movement in price is random; no profit can be derived.

Recently, the validity of this hypothesis has been called into question, with statistical support emerging for both short and long-term inefficiency in global markets, and clear evidence of speculative economic bubbles [27]. As a result, financial investment companies are engaged in an ongoing 'algorithmic arms race' to develop increasingly complex automated strategies for extracting profit from securities trading. It was estimated that by the end of 2010, approximately 53% of all trading in equities was carried out through automated trading [28].

As knowledge of simpler strategies for exposing inefficiencies has entered the public domain, the bar for effective trading algorithms has continued to rise. Currently, hedge funds

use complex short-term mean-reversion portfolios involving large numbers of securities and significant computational infrastructure to derive profit.

As a result of this continuously rising standard, one would expect any remaining inefficiencies in the market to be difficult to identify. It may be the case that exploiting such inefficiencies requires high-frequency trading, as opposed to holding a portfolio over a period of months, as was investigated in this study.

A. Adjusting for selection bias

The parallel multi-objective optimisation algorithm we developed succeeded in identifying highly profitable low-risk portfolios for the training period. Yet these portfolios failed to produce risk-adjusted returns for subsequent test data.

These results highlight an important observation: it is not only the performance of a portfolio that matters (i.e. risk / return) but also the manner in which it was constructed or identified. No matter how convincing the historical performance of a portfolio, this alone cannot be relied on to predict performance in the future. For example, although the solutions presented in Figure 5 might have seemed like outstanding investments on January 1st 2011, the significance of their performance is completely neutralised by the fact that they have been specifically selected for this feature. As well as being risk-adjusted, portfolio performance must be adjusted for selection bias before it can be meaningfully interpreted.

Counter-intuitively, the size of the search space that was available during a portfolio's construction has a direct bearing on the significance of its performance. Consider the situation where an acquaintance tells you about an outstanding portfolio manager who has generated substantial risk adjusted returns for their clients over the past few years. The key question that must be asked here is how the investment opportunity has been identified. If it has been selected from among a much larger group of potential investments specifically because of its performance, then this serves to lower the significance of its historical returns: the larger the search space, the more likely a spuriously strong performance signal will arise by chance.

Future work might involve adjusting fitness functions so that they take into account the effect that free parameters have on decreasing the significance of historical portfolio performance. For example, solutions could be penalised for the number of companies included in the portfolio. Each additional asset would need to justify its inclusion by providing additional performance beyond that which could be expected by chance, in light of the broadening of the search space. Observations based on the optimisation of purely random data could be used to adjust for selection-bias when optimising real-world data.

B. Adjusting S&P 500 returns for selection bias

Given our findings, it is not possible to definitely rule out the possibility that static long-term portfolios can deliver real risk-adjusted returns. Further research is required to see

if strategies which adjust for selection bias might succeed. Different time periods and stock markets should also be investigated before conclusions are drawn.

However, it is worth noting that analyses of hedge fund performance suggest that even complex investment strategies fail to deliver risk-adjusted returns. For example, Lay [29] observed that funds delivering the best performance in one period seldom perform well in the subsequent period, with the result that fund portfolios constructed using SRs produce dismal results. Hedge funds use the latest technology and advanced trading strategies to generate profits for their clients. If any financial time series were to exhibit a performance signal, one would expect hedge funds to fall into this category, given their position at the pinnacle of financial modelling and high frequency trading. Yet, the inconsistency of hedge fund returns suggests that even the performance of the most experienced fund managers using the most sophisticated investment tools is no better than that which could be achieved by gambling in a casino [29].

These observations seem to be at odds with the commonly accepted wisdom that the stock market outperforms risk-free assets over the longer term, as supported by the fact that U.S. equities delivered an average of 4.3% real annual return during the 20th century [30]. However, Jorion and Goetzmann have argued that reliance on historical U.S. data for long-term estimates of expected returns is a serious problem [30]. They argue that such estimates are subject to survivorship bias, in that the U.S. has been specifically identified by investors because of a historical level of economic success which may not endure. They reveal that the high equity premium obtained for U.S. equities during the 20th century was the exception rather than the rule, with only a 0.8% return registered on average worldwide. Indeed, in the first decade of the 21st century, the total real return on an investment in the S&P 500 was -3.4%, even when taking dividends into account. It may be the case that investors who made significant profits from U.S. equities in the 20th century were relying more on luck than on prescience. Jorion and Goetzmann's findings suggest that, as well as being risk-adjusted, long-term historical returns from the S&P 500 should also be adjusted for selection bias.

VIII. CONCLUSION

In conclusion, we have developed a novel evolutionary algorithm for optimising portfolio selection based on a multi-objective analysis of historical data. Although the algorithm successfully identified strong solutions, performance on test data suggested that the significance of these solutions was not above chance. Further study is required to ascertain whether compensating for overfitting can facilitate the identification of portfolios with a genuine risk-adjusted return.

REFERENCES

- [1] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.
- [2] N. Bradshaw, C. Walshaw, C. Ierotheou, and A. Parrott, "A multi-objective evolutionary algorithm for portfolio optimisation," *Proceedings of AISB09*, 2009.

- [3] R. Merton, "On estimating the expected return on the market: An explanatory investigation," *Journal of Financial Economics*, vol. 8, pp. 323–361, 1980.
- [4] Z. Bodie, A. Kane, and A. Marcus, *Essentials of Investments*. New York: McGraw-Hill, 2004.
- [5] E. Elton, "Presidential address: Expected return, realized return, and asset pricing tests," *The Journal of Finance*, vol. 54, no. 4, pp. 1199–1220, 1999.
- [6] P. Skolpadungket, K. Dahal, and N. Harnpornchai, "Portfolio optimization using multi-objective genetic algorithms," *Proceedings of the IEEE Congress on Evolutionary Computation*, 2007.
- [7] J. Jia and J. Dyer, "A standard measure of risk and risk-value models," *Management Science*, vol. 42, no. 12, pp. 1691–1705, 1996.
- [8] Y. Carma and M. Schyns, "Simulated annealing for complex portfolio selection problems," *Financial Modelling*, vol. 150, no. 4, pp. 546–571, 2003.
- [9] F. Modigliani and L. Modigliani, "Risk-adjusted performance: How to measure it and why," *The Journal of Portfolio Management*, vol. 23, no. 2, pp. 45–54, 1997.
- [10] W. Sharpe, "Mutual fund performance," *Journal of Business*, vol. January, pp. 119–138, 1966.
- [11] K. Dowd, "Financial risk management," *Financial Analysts Journal*, vol. 55, no. 4, pp. 65–71, 1999.
- [12] X. Gao and L. Chan, "An algorithm for trading and portfolio management using q-learning and sharpe ratio maximization," *Proceedings of the International Conference on Neural Information Processing*, pp. 832–837, 2000.
- [13] G. Hassan and C. Clack, "Robustness of multiple objective gp stock-picking in unstable financial markets," *Proceedings of GECCO 09*, 2009.
- [14] J. Holland, *Adaptation in Natural and Artificial System*. University of Michigan Press: Ann Arbor, 1975.
- [15] F. Allen and R. Karjalainen, "Using genetic algorithms to find technical trading rules," *Journal of Financial Economics*, vol. 51, pp. 245–271, 1999.
- [16] C. Neely, P. Weller, and R. Dittmar, "Is technical analysis in the foreign exchange profitable? a genetic programming approach," *Journal of Financial and Quantitative Analysis*, vol. 32, no. 4, pp. 405–426, 1996.
- [17] A. Brabazon and M. O'Neill, *Biologically Inspired Algorithms for Financial Modelling*. Springer: Berlin, 2006.
- [18] M. Tapia and C. Coello, "Applications of multi-objective evolutionary algorithms in economics and finance: A survey," *IEEE Congress on Evolutionary Computation*, pp. 532–539, 2007.
- [19] Y. Becker, H. Fox, and P. Fei, "An empirical study of multi-objective algorithms for stock ranking," *R.L. Riolo, T. Soule and B. Worzel (eds.), Genetic Programming Theory and Practice V*, pp. 239–259, 2007.
- [20] A. Messac, C. Pueumi-Sukam, and E. Melachrinoudis, "Aggregate objective functions and pareto frontiers: Required relationships and practical implications," *Optimization and Engineering*, vol. 1, pp. 171–181, 2000.
- [21] D. Whitley, S. Rana, and R. Heckendorn, "The island model genetic algorithm: On separability, population size and convergence," *Journal of Computing and Information Technology*, vol. 7, no. 1, pp. 33–47, 1998.
- [22] B. Miller and D. Goldberg, "Genetic algorithms, selection scheme, and the varying effect of noise," *Evolutionary Computation*, vol. 4, no. 2, pp. 113–131, 1996.
- [23] T. Back, "Self-adaptation in genetic algorithms," *Proceedings of the First European Conference on Artificial Life*, pp. 263–271, 1992.
- [24] W. Bernstein and D. Wilkinson, "Diversification, rebalancing, and the geometric mean frontier," *Working paper for William J. Bernstein Company, Inc.*, 1997.
- [25] G. Grimmett and D. Stirzaker, *Probability and random processes*. Oxford: Oxford University Press, 2001.
- [26] E. Fama, "Efficient capital markets: A review of theory and empirical work," *Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.
- [27] P. Mishra, K. Das, and B. Pradhan, "Empirical evidence on indian stock market efficiency in context of the global financial crisis," *Global Journal of Finance and Management*, vol. 2, pp. 149–157, 2009.
- [28] K. Izumi, F. Toriumi, and H. Matsui, "Evaluation of automated-trading strategies using an artificial market," *Neurocomputing*, vol. 72, no. 16–18, pp. 3496–3476, 2009.
- [29] S. Lay, "The myth of persistent sharpe ratio," *The Hedge Fund Journal*, vol. January, 2010.
- [30] P. Jorion and W. Goetzmann, "Global stock markets in the twentieth century," *The Journal of Finance*, vol. LIV, no. 3, pp. 953–980, 1999.