USING GENETIC ALGORITHM TO CONSTRUCT A MOMENTUM-BASED STOCK FUND

KODA SONG, SONG TANG*

Department of Finance & Risk Engineering, Tandon School of Engineering,

New York University, New York, USA

Abstract: Portfolio optimization is an important research field in modern finance. An important goal in portfolio optimization is to maximize risk-adjusted returns. In addition, momentum investing has gained a wide acceptance by asset managers. Besides, genetic algorithms (GA), which are based on the ideas of evolution and the concepts of Darwin's natural selection, have been widely used to generate high-quality solutions to optimization problems. In this paper, we propose an approach using genetic algorithms to construct a momentum-based 130-30 stock fund. We use the Sharpe Ratio as the fitness function for portfolio evaluation and a Mean-Variance Model with Monte Carlo simulation to optimize the portfolio's long and short positions. Using 2020 market data for the S&P500, our fund outperforms a variety of stock portfolios as well as the S&P500 ETF Fund SPY measured by Total Return, Sharpe Ratio, and Information Ratio.

Keywords: genetic algorithm; crossover; mutation; fitness function; portfolio optimization; momentum fund; Sharpe ratio; mean-variance model; Monte Carlo simulation.

2010 AMS Subject Classification: 91G60, 68T20.

1. INTRODUCTION

1.1 Genetic Algorithms

Genetic Algorithms [1] are a class of heuristic algorithms used to solve optimization problems without using close-formed analytical solutions. They are based on the ideas of evolution and the concepts of Darwin's natural selection, i.e., "survival of the fittest". In general, Genetic Algorithms (GA) is an iterative process which starts from a random population having a number of chromosomes, where each chromosome contains a list of genes. In each step (generation), the genetic algorithm selects chromosomes from a population based on their fitness value to produce new offspring chromosomes. Each chromosome is measured by its fitness, which is its capability to adapt to its "environment". Fitness in GA is measured by the value of an objective function to be optimized. Like natural selection, chromosomes with high fitness values are preserved and have a chance to pass their genes to the next generation while the chromosomes with low fitness values are replaced by offspring chromosomes. A new population is then formed with preserved chromosomes from last generation (with high fitness values) and offspring chromosomes, in the hope that it will survive better than the previous population. The number of chromosomes in the populations from generation to generation is constant. When the GA process is terminated, the chromosome with the highest fitness value is selected as the solution to the optimization problem.

Important operations in a GA process include chromosome selection, crossover, mutation, fitness calculation, and determination of termination criteria. They are detailed as follows and shown in Figure 1:

Chromosome Selection: The main purpose of the selection process is that the better an individual is, the higher its chance of being a parent [2]. Suppose there are N chromosomes in a population, we select M chromosomes with the highest fitness values ($1 < M < N$) to be parent chromosomes for crossover.

Crossover: is implemented by selecting a random point on the chromosomes where the gene exchange happens between two parent chromosomes. The offspring chromosomes from crossover will have genes chosen from both parents based on the exchange point. Such crossover is also called one-point crossover.

Suppose there are n genes s_1, s_2, ..., s_n on each chromosome. If $S_1 = \{s_{11}, s_{12}, ..., s_{1n}\}$, $S_2 = \{s_{21}, s_{22}, ..., s_{2n}\}$, are two parent chromosomes, select a random integer number $1 < r < n$, two offspring chromosomes, S_3 and S_4, acre created by crossing over genes as follows:

$S_3 = \{s_i \mid$ if $i \leq r$, $s_i \in S_1$, else $s_i \in S_2\}$,

$S_4 = \{s_i \mid$ if $i \leq r$, $s_i \in S_2$, else $s_i \in S_1\}$.

Mutation: is a genetic operation to maintain genetic diversity from one generation to the next. A mutation is carried out by selecting one gene at random from a chromosome to be replaced by another gene from a separate chromosome which is randomly selected in the population.

Suppose chromosome $S_1 = \{s_{11}, s_{12}, …, s_{1n}\}$, and S_k is another chromosome randomly selected from the population, $S_k = \{s_{k1}, s_{k2}, …, s_{kn}\}$. A mutation on S_1 occurs when a random selected gene s_{1r} ($1 \leq r \leq n$) from chromosome S_1 is replaced by the gene s_{kr} from the chromosome S_k, i.e., a mutation of $S_1 = \{s_{1i} \mid$ if $i=r$, $s_{1i} = s_{ki}$, else $s_{1i} = s_{1i}$, $1 \leq r \leq n\}$.
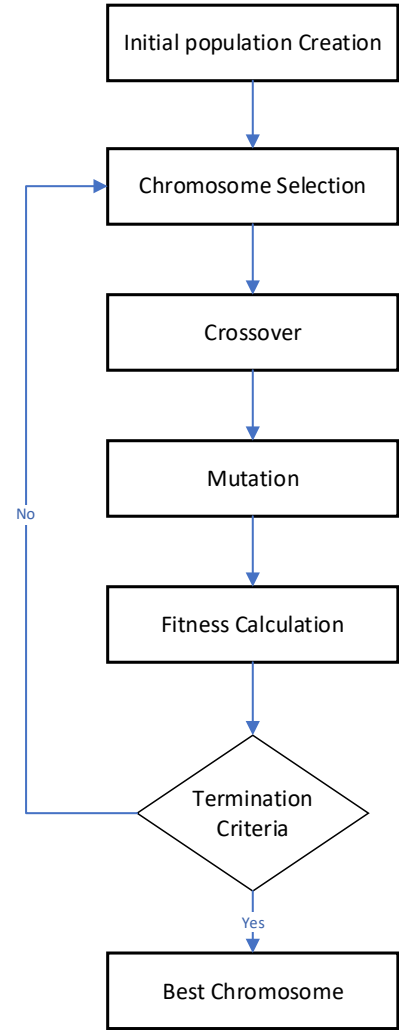

Figure 1: The iterative process of genetic algorithms

The mutation operation mentioned above, which makes random change to one gene on a chromosome after crossover, is based the biological point mutation, and called single point mutation. This mutation operation involves a probability that a gene will be changed on a chromosome. The probability is called mutation rate or mutation probability, which is in the range of [0, 1], usually low, such as 3%. A common method of implementing the mutation operation involves generating a random variable for a gene on a chromosome. If the random variable is less than or equal to the mutation rate, the gene will be mutated [3].

Fitness Calculation: A fitness function measures the strength of a chromosome's ability to survive. The value calculated from the fitness function is used to select chromosomes for better surviving in the evolution process. In our GA, we use Sharpe Ratio as the fitness function.

Determination of Termination Criteria: A genetic algorithm performs the operations described above, i.e., chromosome selection, crossover, mutation, fitness calculation in sequence for a population. After these operations are performed, a new population is created for next generation (shown in Figure 1). Because of the nature of genetic algorithms, most of the time, it is not clear when the algorithm should stop so the algorithm must check the termination criteria at each generation. Termination criteria are usually based on statistical information such as the maximum number of generations, or convergence to a single fitness value for the best chromosome over multiple generations.

1.2 Portfolio Optimization

Portfolio optimization is a process to select the "best" portfolio among a set of potential portfolios. Modern portfolio theory [4] introduced the concept of the efficient frontier, which is a curve of the highest expected returns for each separately given risk level (Figure 2). Using the risk-free rate, we can plot a tangent from the expected return axis to the efficient frontier and the tangent point locates the portfolio with the highest Sharpe Ratio of return per risk [5]. Therefore, in our studies, the portfolio with the highest Sharpe Ratio is selected as the "best" portfolio for back-testing.
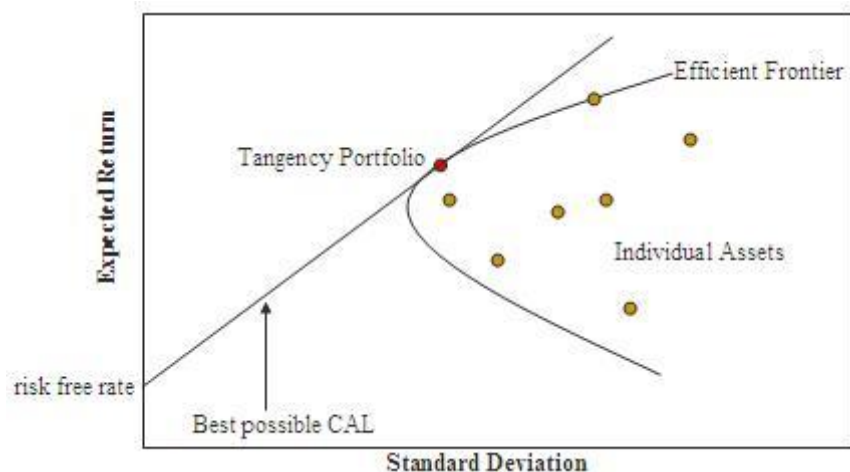
Figure 2: Efficient Frontier (en.wikipedia.org/wiki/Efficient_frontier)

Momentum investing assumes the past stock returns predict their future returns. A momentum portfolio is typically constructed by buying stocks with past high or positive returns and short-selling stocks with past low or negative returns. The academic literature has documented the efficacy of momentum investments over multiple time periods, in different markets, and among a variety of asset classes [6][7][8][9]. Additionally, it was found that 77% of US mutual funds that relied on momentum strategies realized significantly better performance as compared with other funds [10]. In this paper, we use GA to optimize the S&P500 component stocks and their weight in momentum portfolios. Back-testing with market data from 2020 indicates our optimized GA portfolio achieves the best Total Return and highest Sharpe Ratio by comparing with a set of momentum and mean-reversion portfolios as well as the S&P500 ETF Fund SPY.

Although portfolio optimization problem could be solved analytically, financial reality may complicate both the object functions and constraints facing financial agents, making it difficult to find an analytical solution for the optimization problem. Therefore, genetic algorithms could be used as an alternative approach [11]. For example, general algorithms were used to select technical rules for trading S&P500 index and found the selected rules had the ability to identify the periods to buy the index when the daily returns were positive and volatility was low, and sell when the reverse was true [12].

2. PRELIMINARIES

2.1 Long/Short Determination – Short Ratio

The method to determine long or short direction is based on a factor named "Short Ratio", which is the shares shorted by the market over the Average Daily Trading Volume (http://regsho.finra.org/regsho-Index.html). In a momentum strategy, a stock that is shorted with a low short ratio is expected to continue its upward momentum and so is purchased, while a stock with a high short ratio is sold short. Based on the "Short Ratio", we sort and divide all component stocks of S&P 500 into 10 groups, with groups given scores from 1 to 10 as the "Short Ratio" increased, each group will have 10% of S&P500 component stocks. In our GA method, stocks with ranks 9 and 10 are available for selling short, while stocks with ranks 1 and 2 are for purchase selection. Other stock groups are not taken into consideration. More details of our stock selection will be covered in next section (section 2.2).

2.2 The Description of Our GA Method

Our GA algorithm has 20 populations in each generation, with 50 portfolios (chromosomes) in each population, and each portfolio (chromosome) had 10 stocks (genes).

To generate a portfolio for the 1st generation, 5 stocks randomly selected from groups 1 & 2 are purchased, while 5 stocks from groups 9 & 10 are made short. The long and short stocks are arranged alternatively in each portfolio, i.e., the odd numbers are long stocks and even numbers are for short. The stock selection is a random draw with replacement, i.e., a stock could be present in different portfolios in a generation.

Optimal stock weights are calculated via the Markowitz Mean-Variance Model [4]. The details of our stock weight calculation procedure are presented in next section (section 2.3). The same procedure is applied 50 times to create the optimal stock weights for each portfolio in a population.

Then the Sharpe Ratio is used as the fitness for evaluating portfolios. Our portfolios are rebalanced every week, as detailed in section 3.1.

The Sharpe Ratio is calculated for each portfolio in a population and ranked among the 50 portfolios from high to low. Two portfolios from the top 30 portfolios are randomly selected to pair for one-point crossover (as defined in section 1.1). Each one-point crossover creates two offspring portfolios as shown in Figure 3. The procedure continues until 20 offspring portfolios are created. Therefore, a new population will be formed with the 30 top portfolios and the 20 offspring portfolios (the bottom 20 portfolios are replaced by the offspring portfolios).

Finally, mutation is performed to introduce a random modification to each offspring portfolio as shown in Figure 4.

- One-point crossover:

The first 5 stocks of Portfolio 1 join with the last 5 stocks of Portfolio 2 to generate the first offspring portfolio. Similarly, the last 5 stocks of Portfolio 1 are combined with the first 5 stocks of Portfolio 2 to create the second offspring portfolio.

Parent Chromosomes

Stock No	1	2	3	4	5	6	7	8	9	10
Symbol	A1	B1	C1	D1	E1	F1	G1	H1	I1	J1

Stock No	1	2	3	4	5	6	7	8	9	10
Symbol	A2	B2	C2	D2	E2	F2	G2	H2	I2	J2

Offspring Chromosomes

Stock No	1	2	3	4	5	6	7	8	9	10
Symbol	A1	B1	C1	D1	E1	F2	G2	H2	I2	J2

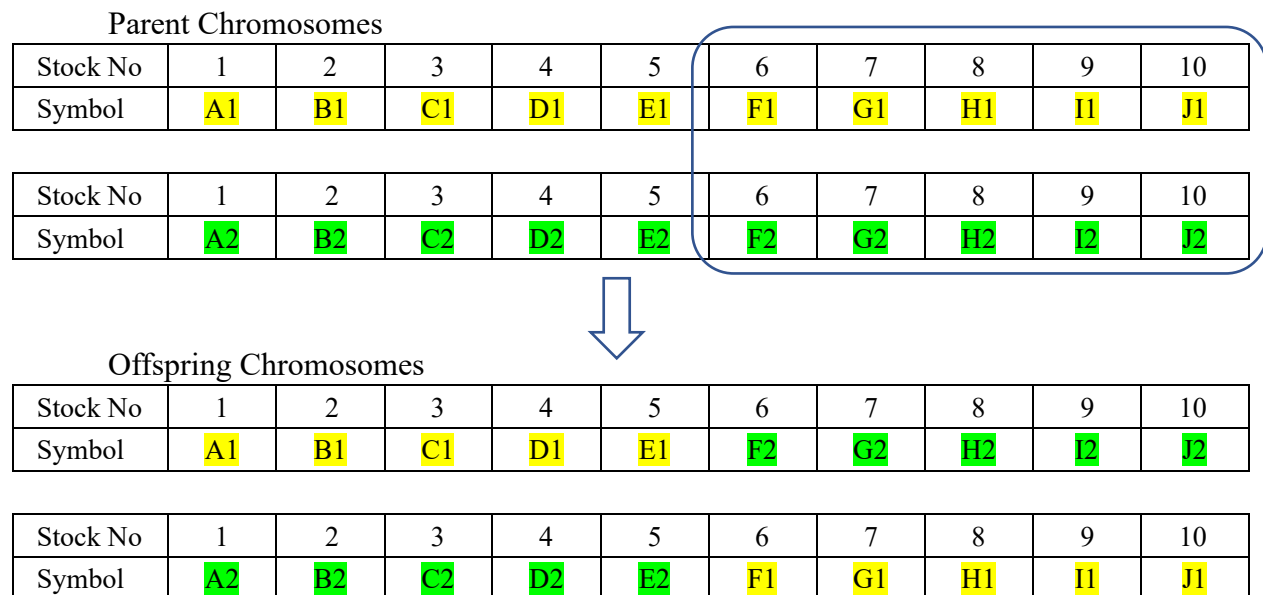| Stock No	1	2	3	4	5	6	7	8	9	10
Symbol	A2	B2	C2	D2	E2	F1	G1	H1	I1	J1

Figure 3: One-point Crossover

- Mutation

To simulate mutation for the offspring portfolios, a stock is randomly selected from an offspring portfolio, and assigned with a normalized random number between 0 and 1. If the random number is less than the 3% mutation rate (as defined in Section 1.1), the stock will be replaced by any stock from the same set of either longed or shorted stocks, i.e., group 1 & 2 for longs and group 9 & 10 for shorts. For each offspring portfolio, the mutation operation would repeat 10 times, so the probability for mutation will be statistically applied to every stock in the portfolio. The mutation operation is equally applicable to all the 20 offspring portfolios in each new population.

Before mutation (An offspring portfolio from crossover)

Stock No	1	2	3	4	5	6	7	8	9	10
Symbol	A1	B1	C1	D1	E1	F2	G2	H2*	I2	J2

After mutation

Stock No	1	2	3	4	5	6	7	8	9	10
Symbol	A1	B1	C1	D1	E1	F2	G2	W2*	I2	J2

* Assume both stocks H2 and W2 are either long or short stocks.

Figure 4: Mutation

This process of portfolio (chromosome) selection, crossover, mutation, and fitness calculation (Sharpe Ratio) will iterate until the termination criteria are meet. The termination criteria are either the best portfolio converged upon a single value or the max number of generations, which is 100 generations in our studies, is reached. The best portfolio from each population in the last generation is selected for back-testing.

2.3 Weight Determination of Portfolio Constituents and Portfolio Fitness

Our portfolios are constructed as 130-30 funds, in which a ratio of 130% of starting capital is allocated to long positions by taking in 30% of the starting capital from shorting stocks. Therefore, each of our portfolios had a 130% leverage built in. A 130-30 fund is a "long-short" approach often lumped into long-short equity mutual funds [13][14]. As we mentioned previously, for our portfolios, stocks with ranks 9 and 10 are sold short and stocks with ranks 1 and 2 are purchased. Stocks in other ranks are not taken into consideration.

The Markowitz Mean-Variance Model [4] is applied to determine the optimized dollar weight for each stock in a portfolio. Specifically, Monte-Carlo Simulation is used to generate 10,000 random weight combinations for the 10 stocks in a portfolio. With 10,000 weights, each portfolio has 10,000 different fitness values (Sharpe Ratios). The highest Sharpe Ratio is selected as the portfolio's fitness, and the corresponding weights are the optimal weights for the stocks in the portfolio. The same process is applied to all the 50 portfolios in each population.

The process of calculation is shown below:

Since each portfolio contained 10 stocks, a matrix of 10×10 is built to store the covariance between any two stocks in this portfolio.

$$\Sigma = \begin{pmatrix} \sigma_{1,1} & \cdots & \sigma_{1,10} \\ \vdots & \ddots & \vdots \\ \sigma_{10,1} & \cdots & \sigma_{10,10} \end{pmatrix}$$

W is a matrix of 10×10000. Each column has 10 rows, which represent the weight for each stock.

$$W = \begin{pmatrix} w_{1,1} & \cdots & w_{1,10000} \\ \vdots & \ddots & \vdots \\ w_{10,1} & \cdots & w_{10,10000} \end{pmatrix}, \ w_i = \begin{pmatrix} w_{1,i} & \cdots & w_{10,i} \end{pmatrix}^T$$

The expected Return of 10 stocks is calculated based on historical data from eodhistoricaldata.com and Yahoo Finance in 2019 and 2020:

$$\mu = (\mu_1 \ldots \mu_{10})^T$$

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} r_i, where\ r_i\ is\ the\ historical\ daily\ return\ of\ this\ stock$$

The Expected Return, Volatility, Sharpe Ratios and Fitness for Portfolio i are:

$$R_i = w_i^T \cdot \mu = (w_{1,i} \ldots w_{10,i}) \cdot (\mu_1 \ldots \mu_{10})^T$$

$$Vol_i = w_i^T \cdot \Sigma \cdot w_i = (w_{1,i} \quad \ldots \quad w_{10,i}) \cdot \begin{pmatrix} \sigma_{1,1} & \cdots & \sigma_{1,10} \\ \vdots & \ddots & \vdots \\ \sigma_{10,1} & \cdots & \sigma_{10,10} \end{pmatrix} \cdot (w_{1,i} \quad \ldots \quad w_{10,i})^T$$

$$SR_i = \frac{R_i - R_f}{Vol_i} \qquad R_f\ is\ the\ risk\text{-}free\ rate.$$

$$Fitness\ for\ Portfolio\ i = max\ (SR_1\ \ldots\ SR_{10000})$$

3. MAIN RESULTS

3.1 Our Way of Portfolio Rebalancing

The back-test period is the entire year of 2020. This is a weekly rebalanced strategy, where we open positions on the first trading day of the week and exit on the last trading day of the week. Each week, our GA method generates 20 populations, based on the daily price data from the first trading date of 2019 to the last trading day before the week for rebalancing, as described in Section 2.2. Each population has 50 portfolios, and each portfolio has 10 stocks, with 5 stocks from groups 1 and 2 for long positions and 5 stocks from groups 9 and 10 as short positions. The best portfolio from each population is selected, and 5% of our current capital is invested into each portfolio with 130% leverage. The optimized weight of each stock in each portfolio is determined by Mean-

Variance Model with Monte Carlo simulation as described in section 2.3. All the long and short positions will be closed out on the last trading date of the week.

3.2 Funds for Comparison

There are two common and simple strategies for S&P 500 stocks: 1) current best performers will continue to be the future best performers; 2) current worst performers will become the future best performers. Therefore, at the 1st trading day of a week we select best and worst performers based on their cumulative return in the week just past for rebalancing and application during the week.

In addition, we create a 130-30 fund with the 5 best performed stocks and 5 worst performed stocks and rebalance the fund weekly. We call the fund "WML" (Winner Minus Loser). Therefore, we have a total of 6 investment strategies to compare:

- The S&P500 ETF fund, SPY, which serves as the benchmark for market performance.
- The Long-only fund with 10 weekly best performers, named as WIN in the following diagrams and tables.
- The Long-only fund with 10 weekly worst performers, named as LOS.
- The 130-30 momentum fund, with 5 weekly best performers and 5 weekly worst performers, WML.
- The 130-30 momentum fund created by our genetic algorithm strategy with equally dollar-weighted stocks, GA-EW-MOM.
- The 130-30 momentum fund created by our genetic algorithm strategy, while using Mean-Variance Model/Monte Carlo optimization of stock's weights, GA-MOM.

Besides SPY, the component stocks of each fund are updated weekly based on our rebalance strategies.

3.3 Fund Performance Evaluation

The following (Figure 5 and Table 1) are the performance measurements of the 6 funds, i.e., GA-MOM, GA-EW-MOM, WML, WIN, LOS, and SPY.

Figure 5: Cumulative Returns of the 6 Funds in 2020

Metrics	GA-MOM	GA-EW-MOM	WIN	LOS	WML	SPY
Total Return	**59.0%**	49.2%	-24.8%	49.2%	-43.7%	14.2%
Volatility	32.9%	**30.3%**	41.9%	53.5%	54.2%	28.0%
Max Drawdown	17.6%	**15.5%**	43.2%	41.3%	52.7%	21.1%
Alpha	**49.0%**	39.0%	-40.5%	29.1%	-57.3%	0
Beta	**0.72**	0.73	1.09	1.42	0.95	1.0
Sharpe Ratio	**1.8**	1.63	-0.59	0.92	-0.81	0.51
Information Ratio	**1.34**	1.18	-1.18	0.93	-1.8	-

Note:
1. The best value in each row is shown as highlighted.
2. Total Return is the cumulative profit for all the trading days in 2020.
3. Volatility is the annualized standard deviation of daily returns for all the trading days in 2020.
4. Max drawdown is the maximum loss from the local maximum, expressed as a percentage, annualized.

5. Beta is a measure of how an individual asset moves (on average) when the overall stock market increases or decreases.

$$Beta = \frac{Covariance(R_{portfolio}, R_{market})}{Variance(R_{market})}$$

6. Alpha is referred to as the annualized excess return to describe how much a strategy beats the market.

$$Alpha = R_{portfolio} - R_f - Beta * (R_{market} - R_f)$$

7. Sharpe Ratio (SR) is a measure of return used to compare the performance of portfolios by adjusting for risk.

$$SR = \frac{avg(R_{portfolio} - R_f)}{std(R_{portfolio} - R_f)}$$

    $where$ "avg" and "std" $are\ the\ mean\ and\ standard\ deviation\ of\ portfolio's\ daily\ return$

8. Information Ratio (IR) is a measurement of portfolio returns beyond the returns of a benchmark, usually an index, compared to the volatility of those returns. It is calculated by the difference between the return of a portfolio and the benchmark divided by the standard deviation of the difference:

$$IR = \frac{avg(R_{portfolio} - R_{market})}{std(R_{portfolio} - R_{market})}$$

Table 1: Performance Measurements of the 6 Funds in 2020

Both the table and graph (Figure 5 and Table 1) indicate that our GA-MOM fund performed the best, having the highest Total Return, Alpha, Sharpe Ratio, and Information Ratio while the Max Drawdown and Volatility were controlled well despite the several circuit breakers in 2020.

The performance of GA-EW-MOM was close to GA-MOM. But the simple 130-30 momentum without using GA performed worst among 6 funds in almost all the matrices.

For both long funds, although the LOS fund (Long-only fund with 10 weekly worst performers) lost considerably during March, it rebounded quickly afterwards, and exceeded GA-MOM several times. However, its Volatility was extremely high, with several big Drawdowns. The WIN fund (Long-only fund with 10 weekly best performers) was shown to be a poor portfolio with respect to return.

3.4 Quarterly Performance Evaluation

Although our momentum fund GA-MOM had the best cumulative return in 2020, other funds performed better in some quarters (Figure 6 and Table 2).
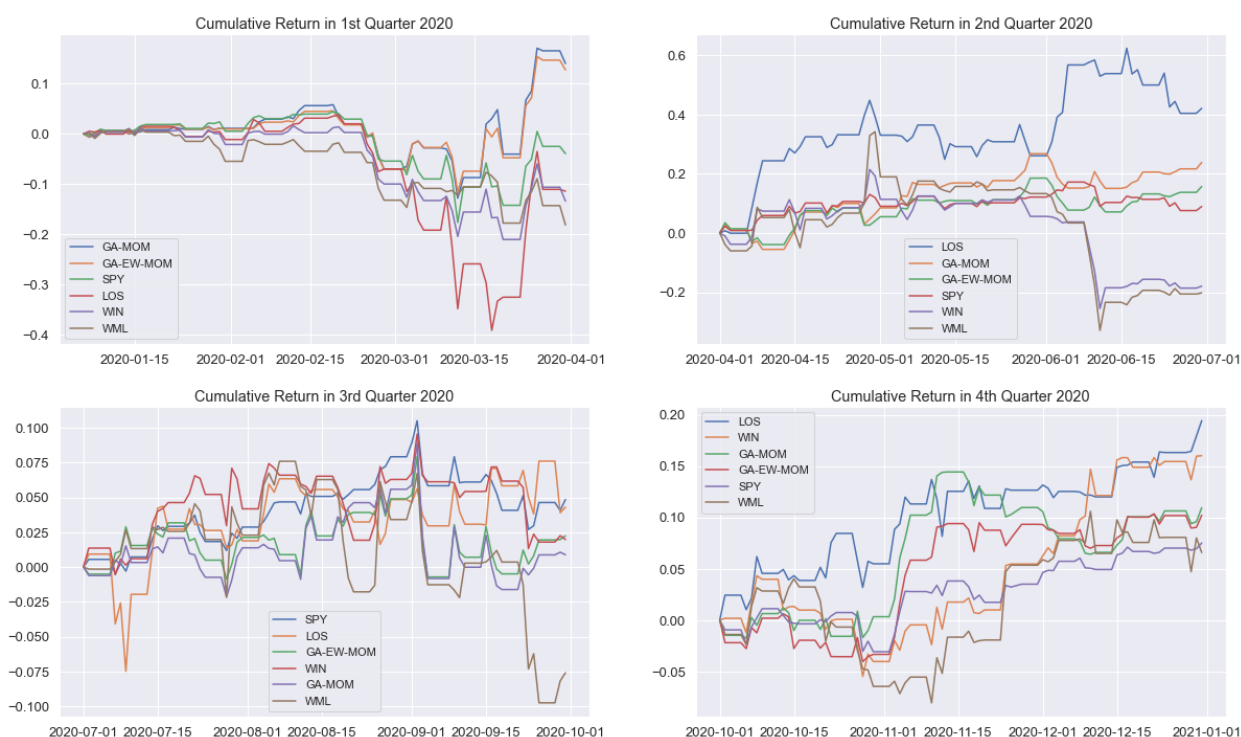

Figure 6: Quarterly Cumulative Returns for the 6 Funds in 2020

1st Quarter						
Metrics	GA-MOM	GA-EW-MOM	WIN	LOS	WML	SPY
Total Return	**14.0%**	12.7%	-13.4%	-11.4%	-18.1%	-3.9%
Annualized Return	**58.8%**	53.4%	-56.2%	-47.9%	-76.2%	-16.6%
Volatility	49.1%	44.7%	42.9%	85.9%	**36.8%**	46.2%
Max Drawdown	17.6%	**15.5%**	22.3%	41.3%	18.9%	21.1%
Alpha	**72.9%**	67.5%	-41.9%	-22.6%	-67.7%	0
Beta	0.85	0.85	0.87	1.53	**0.51**	1.0
Sharpe Ratio	**1.2**	1.19	-1.31	-0.56	-2.07	-0.36
Information Ratio	2.47	**3.03**	-2.62	-0.41	-1.95	-

2nd Quarter					
Metrics	GA-MOM	GA-EW-MOM	WIN	LOS	WML	SPY
Total Return	23.8%	15.7%	-17.9%	**42.1%**	-20.2%	9.0%
Annualized Return	95.4%	62.8%	-71.8%	**168.5%**	-80.8%	35.8%
Volatility	33.3%	**29.9%**	61.3%	55.0%	88.1%	25.2%
Max Drawdown	**9.2%**	9.6%	38.7%	13.8%	50.0%	8.2%
Alpha	86.8%	52.7%	-134.0%	**115.3%**	-153.9%	0
Beta	**0.24**	0.28	1.73	1.48	2.04	1.0
Sharpe Ratio	2.86	2.1	-1.17	**3.06**	-0.92	1.42
Information Ratio	1.3	0.59	-2.04	**2.98**	-1.34	-

3rd Quarter					
Metrics	GA-MOM	GA-EW-MOM	WIN	LOS	WML	SPY
Total Return	0.9%	2.2%	2.0%	4.3%	-7.6%	**4.9%**
Annualized Return	3.5%	8.8%	7.8%	16.9%	-29.9%	**19.1%**
Volatility	23.6%	22.4%	21.1%	28.7%	34.1%	**14.3%**
Max Drawdown	9.7%	8.0%	7.5%	8.3%	16.1%	**7.1%**
Alpha	-18.0%	-8.8%	-8.8%	-2.8%	-49.9%	0
Beta	1.13	1.07	**0.87**	1.03	1.05	1.0
Sharpe Ratio	0.15	0.4	0.37	0.59	-0.88	**1.34**
Information Ratio	-0.85	-0.59	-0.64	**-0.08**	-1.52	-

4th Quarter					
Metrics	GA-MOM	GA-EW-MOM	WIN	LOS	WML	SPY
Total Return	10.9%	10.2%	16.0%	**19.4%**	6.6%	7.5%
Annualized Return	43.1%	40.2%	63.1%	**76.5%**	25.9%	29.6%
Volatility	19.2%	**19.0%**	24.8%	23.2%	29.6%	13.9%
Max Drawdown	7.0%	**4.6%**	9.4%	4.9%	11.6%	4.1%
Alpha	23.9%	17.4%	29.9%	**47.7%**	-4.2%	0
Beta	**0.65**	0.77	1.12	0.97	1.02	1.0
Sharpe Ratio	2.25	2.12	2.55	**3.29**	0.88	2.13
Information Ratio	0.68	0.59	1.63	**2.31**	-0.13	-

Table 2: Quarterly Performance Measurements of the 6 Funds in 2020

In the first quarter, our momentum fund GA-MOM performed the best, with well controlled risk and Max Drawdown. This proves that in a steeply declining and extremely volatile market, our momentum fund can be profitable and antifragile [15].

In the second quarter, the market recovered, and SPY earned about 9%. During this time, the past losers performed the best in respect of Total Returns, Alpha, Sharpe Ratio, and Information Ratio. However, its Volatility and Beta were high, which means the high return of past losers was mainly from the market. In contrast, our momentum fund GA-MOM was the second best with low Max Drawdown and Beta. More importantly, as shown in Table 3, we calculated the Pearson Correlation Coefficient between market (SPY) and all the 5 funds for the 2nd Quarter in 2020:

Fund Name	Coefficient	P-Value
WIN	0.713721	5.19E-11
LOS	0.679851	8.94E-10
WML	0.584282	4.97E-07
GA-EW-MOM	0.238972	0.059264
GA-MOM	0.180884	0.155984

Table 3: Coefficient and P-Value between Market (SPY) and the Five Funds

It is indicated that our GA-MOM fund has lowest correlation coefficient value, 0.180884, while its p-value is the highest, which means that the correlation between GA-MOM fund and market is the least significant. This an important characteristic for a momentum fund in a volatile market.

In the third quarter, no portfolio beat the market. SPY stood out among all the funds. The market stabilized with reduced volatility and generally trended higher in the 3rd quarter. We believe in such a market situation, index funds could perform better than our momentum funds, which presents an interesting area for further research.

In the fourth quarter, all portfolios performed relatively well. Although our momentum fund GA-MOM did not achieve the best performance, it still had lowest Beta and its risk-adjusted return beat market.

3.5 Summary

To summarize, although our GA momentum fund might not have been the best performer in every quarter, it performed best while the market was in its worst decline in the year and most volatile condition in 2020. Furthermore, by comparison to other funds, it had a relatively high Sharpe Ratio in most quarters. In other words, it accumulated profits through risk-adjusted returns. Therefore, in the end, it had the highest Total Return, Alpha, Sharpe Ratio and Information Ratio, Lowest Beta, second-lowest Max Drawdown, and second-lowest Volatility.

We also constructed an "Equally Dollar-Weighted Momentum" fund (GA-EW-MOM) to compare with our optimized weighted "Momentum" fund (GA-MON). As presented in the above diagrams and tables, the trend of these two funds' returns is very similar and most metrics are also close. However, the optimized portfolio (GA-MOM) improves the Total Return by about 20% in 2020, which successfully proves the effect of optimization.

4. CONCLUSION AND DISCUSSION

This paper has applied Genetic Algorithms in selecting and optimizing S&P500 investment portfolios for a 130-30 momentum fund. The portfolio's stock weights are optimized through Mean-Variance Model and Monte Carlo simulation. The fitness of each portfolio is measured by its Sharpe Ratio. Through portfolio selection, crossover, and mutation, the "better" portfolios are preserved through multiple generations while abandoning the "worse" portfolios. In the end, the portfolio with the highest fitness in the last generation is added to our fund. The same strategy is repeated until the 20 best portfolios are selected with 5% investment funding for each portfolio. From the results of the back-testing, we conclude a simple 130-30 fund fails to beat the market, while our GA constructed 130-30 fund excels in total return as well as risk-adjusted return. This indicates that the application of Genetic Algorithms could be a useful and effective approach in portfolio optimization.

In our current studies, we have not included transaction costs for the weekly portfolio rebalancing for back testing, as we assume the same transaction costs are applicable to all the funds in comparisons described in section 3.2. We are further investigating the robustness of our results by including transaction costs of 0.1% or 0.5% of the total trade value. In addition, we are expanding the time duration of our studies to 22 years, from the year 2000 to current date, and trying longer rebalance periods, such as monthly, every 3, 6 and 12 month frequencies.

In the future, more fitness functions, and a strategy to avoid short squeezing will be tested to further increase portfolio robustness and reduce maximum drawdown in a volatile market. In addition, it would be interesting to find out at what market conditions passive investment strategies such as ETFs and index funds would perform better than our momentum fund.

5. TECHNICAL ASPECTS OF OUR RESEARCH

Market Data used in our research are from EOD Historical Data Services (https://eodhistoricaldata.com) and Yahoo Finance. More specifically, S&P500 component stock list, S&P500 historical daily prices, and fundamental data, S&P500 Index ETF (SPY) historical prices are downloaded from eodhistoricaldata.com. Treasury Yield 10 Years (TNX) is used as a risk-free rate and downloaded from Yahoo Finance. Daily Short Sale Data are from FINRA (http://regsho.finra.org/regsho-Index.html).

As shown in the Table 4, the parameter settings for our GA strategy are:

No	Parameters	Values
1	Population Size	50
2	Crossover	Single Point
3	Mutation Rate	3%
4	Chromosome Selection	Ranking
5	Fitness Calculation	Sharpe Ratio
6	Termination Criteria	Fitness Value Convergence or 100 generations

Table 4: All the Parameter Settings for Our GA

All the coding is done in C++ and Python. A SQLITE database is built for data persistency. All the source codes and the database are available to share upon request.

ACKNOWLEDGE

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] J. H. Holland, Adaption in natural and artificial systems, University of Michigan Press, 1975.

[2] A. Shukla, H.M. Pandey, D. Mehrotra, Comparative review of selection techniques in genetic algorithm, in: 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), IEEE, Greater Noida, India, 2015: pp. 515–519.

[3] A. Hassanat, K. Almohammadi, E. Alkafaween, E. Abunawas, A. Hammouri, V.B.S. Prasath, Choosing mutation and crossover ratios for genetic algorithms-a review with a new dynamic approach, Information. 10 (2019), 390.

[4] H. Markowitz, Portfolio selection. J. Finance, 7(1) (1952), 77-91.

[5] W.J. Bernstein, D. Wilkinson, Diversification, rebalancing and the geometric mean frontier, 1997, Available at SSRN: http://dx.doi.org/10.2139/ssrn.53503.

[6] N. Jegadeesh, S. Titman, Profitability of momentum strategies: An evaluation of alternative explanations, J. Finance, 56 (2001), 699-720.

[7] J. Okunev, D. White, Do momentum-based strategies still work in foreign current markets? J. Financ. Quant. Anal. 38 (2003), 425-447.

[8] C.B. Erb, Claude, C.R. Harvey, The strategic and tactical value of commodity futures, Financ. Anal. J. 62 (2006), 69-97.

[9]   C.S. Asness, T.J. Moskowitz, L.H. Pedersen, Value and momentum everywhere, J. Finance 58 (2013), 829-895.

[10] M. Grinblatt, S. Titman, R. Wermers, Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior, Amer. Econ. Rev. 85 (5) (1995), 1088-1105.

[11] S.H. Chen, Genetic Algorithms and Genetic Programming in Computational Finance, 2002, Springer.

[12] F. Allen, R. Karjalainen, Using genetic algorithms to find technical trading rules, J. Financ. Econ. 51 (1999), 245-271.

[13] C. Armfelt, D. Somos, Performance, Benefits and Risks of Active-extension Strategies. Master's Thesis in Finance, Stockholm School of Economics, 2008, available at http://arc.hhs.se/download.aspx?MediumId=474

[14] M. Horejs, Long-short equity handbook, 2012, http://advisor.morningstar.com/uploaded/pdf/Alt_Long-ShortEquity.pdf

[15] N.N. Taleb, Antifragile, things that gain from disorder, Random House, New York, 2012.