Non-Linear Factor Model for Asset Selection using Multi Objective Genetic Programming

Ghada Hassan
Department of Computer Science
University College London
Gower Street, London WC1E 6BT
g.hassan@cs.ucl.ac.uk

ABSTRACT

Investors vary with respect to their expected return and aversion to associated risk, and hence also vary in their performance expectations of the stock market portfolios they hold. In this work we present an empirical study of the use of the Multiobjective genetic programming (MOGP) technique for a real world problem of portfolio optimisation on UK FTSE-100 stocks. The MOGP evolves a nonlinear factor model of technical factors for asset ranking, and provides visual insight into the risk return trade-off involved in discovering an approximation to the risk-return efficient frontier of a portfolio optimization problem. We provide preliminary analysis of the set of factors chosen in the evolved solutions. We find evidence for the effect on risk-adjusted returns of firm size, return on equity and cash yield (and little or no evidence for the book to market ratio).

Categories and Subject Descriptors

I.2.M [**Artificial Intelligence**]: Miscellaneous

General Terms

Algorithms, Experimentation

Keywords

GP, Multiobjective Optimization, Factor Models, Portfolio Optimization, Finance

1. 1. INTRODUCTION

With the vast number of stocks available to choose from, the extensive information publicly available about traded firms, ease of access to values of economic indicators, and the increasing effect international markets have on each other, the stock market investors' job is becoming more difficult.

Investors are primarily interested in the expected return of their investment and the associated risk. Two theories provide the foundation for analyzing the trade-off between risk and return. The Capital Asset Pricing Model (CAPM) [11] is a linear model that predicts the stock return to be associated with the stock's systematic risk, which is the risk that cannot be diversified away by holding a portfolio of inverse correlated assets. The CAPM assumes asset returns are normally distributed, that variance is an adequate measurement of risk and no taxes or transaction costs are considered. The second theorem is the Arbitrage Pricing theorem (APT) [10]. The APT is a generalized form of the CAPM. It is a linear model of asset returns that depends on k multi factors, instead of a single factor of exposure to market risk as in the CAPM. It is essentially saying that the systematic risk of the CAPM should be modelled through sensitivity of the asset to several macroeconomic and/or fundamental factors, because there can hardly be one sole measure of risk. The APT, however, does not explicitly state what these factors are. In the work of Fama and French [3], [2], the authors showed that the returns on assets are significantly affected by three factors; company size, and book-to-market ratio and sensitivity to market movement. In [8], the interest rates, money growth, oil prices, and growth in industrial production were all candidates in the development of a multi-factor model that proved to have some predictive power on the markets tested. Recently, some researchers [1, 4, 9] questioned the linearity framework of the model. It was shown that the market exhibits evidence for nonlinear behaviour with effects asset pricing and expected returns. According to McMillan [1], the nonlinear behaviour is a consequence of the presence of market frictions and transaction costs, which are absent from the traditional theories, in addition to the interaction between heterogeneous traders, i.e. informed and noise traders.

Genetic programming (GP) [5], handles tree-sructured individuals. This allows for the evolution of non linear, variable size rules. Multi Objective Genetic Programming (MOGP) integrates Pareto dominance concepts into the framework of Evolutionary Algorithms (EAs) to allow for the comparison between individuals based on multiple conflicting objectives. Instead of producing one best solution, they produce a Pareto front of many solutions to the problem in one run.

This work addresses the evolution of factor models (used to select assets) using a MOGP. The MOGP is used to select factors that affect the risk-adjusted returns and uses them to rank assets and accordingly generate buy and sell decisions.

The paper is organized as follows. Section 2 describes the asset selection problem formulation, the investment model framework with real world constraints. Followed by, experiments and results in Section 3. Finally, Section 4 concludes.

GECCO'08, July 12–16, 2008, Atlanta, Georgia, USA.

2. PORTFOLIO OPTIMIZATION

2.1 General Problem

The general portfolio optimization problem is the choice of an optimum set of assets to include in the portfolio and the distribution of investor's wealth among them such that the objectives sought by holding the portfolio are maximized. Markowitz [7] assumed that the objectives of the investor are maximizing the return on investment and minimizing the associated risk. Hence, solving the problem requires the simultaneous satisfaction of maximizing the return:

$$E = \sum_{i=1}^{n} x_i \mu_i \tag{1}$$

,and minimizing the standard deviation:

$$V = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_{ij} \tag{2}$$

Where n is the number of securities in portfolio, x_i is the relative amount invested in security i ,and $\sum_{i=1}^{n} x_i = 1$. E is the expected portfolio return, V is the portfolio variance, which is the average squared deviation of the return from its expected mean value, and σ_{ij} is the covariance between assets i and j. These equations are solved by a set of points that constitute the efficient frontier of the problem.

2.2 Investment Model

The investment model employed is inspired by real world fund management practices. The portfolio held consists of one cash line and has a fixed cardinality of $n = 25$ stocks. The initial portfolio value is $C_o = £1,000,000$ in cash with no stock holdings. After that, the portfolio will constitute of n securities, and the current cash holding will be denoted by C, where we try to keep C less than or equal a maximum bound $C_{max} = 3\%$ of the total fund value. S is the universe of equities, S_n is the set of securities held in the portfolio. For all buying and selling decisions in any day, it is assumed that we can trade at the opening price of that day. During the holding period, interest received on cash holdings is ignored.

For the duration of holding period, we do the following. At the start of each month, we calculate attractiveness of each stock in S according to the nonlinear factor model examined, and sort them accordingly. If any of the stocks we currently hold falls in the bottom quartile of the rank, it is sold. If the number of stocks currently in the portfolio is less than n or $C > C_{max}$, then we need to buy stocks from the top quartile, starting with the most attractive. The proportion to be invested in each stock is C_i, and is decided by:

$$C_i = min(\frac{C}{n - |S_n|}, 4\%\ of\ total\ fund\ value) \tag{3}$$

If we still have cash more than C_{max}, and there are some stock holdings with less than 4% of the total fund value, then we use all remaining cash to bring each of these stock holdings up to 4% or at least up to the maximum that the extra cash allows us to.

Several realistic constraints were included in the system: a portfolio cardinality of 25, lower and upper bounds on investment per stock, maximum cash holding, and 2% transaction costs. With the addition of constraints, no analytic method exists for solving it, otherwise it can be solved in an exact manner by quadratic programming.

Table 1: Definition of Financial and Economic Factors Used

Close Price	Previous day last reported trade price
Price Momentum	Price per USD price change
Volume	Total sum of shares that have traded in the security for the current or most recent days on its primary trading market place
Price-Cash Ratio	Compares stock price with cash flow from operations per outstanding shares
Price to Book Ratio	Price of stock is divided by reported book value the of the issuing firm
Price-Earnings Ratio	Financial Ratio that compares stock price with earnings per share
30-Day moving average	Mean of the previous 30 days' closing prices
Moving average changes	
Volatility	The degree of price fluctuations of the stock - expressed by variance or standard deviation
Dividend yield	The Company's annual dividend payments divided by its market capitalization, or the dividend per share divided by the price per share
BVPS	A measure to determine the level of safety associated with each individual share after all debts are paid. It represents the amount of money that the holder of a share would receive if the stock was liquidated
Market capitalization	Price per share multiplied by the total number of shares outstanding
Change Return on Equity	return on equity (current year) - return on equity (previous year)
Revenue Growth	The rate at which revenue has increased annually. Can be negative. $= \frac{current year's revenues}{previous year's revenues} - 1 * 100$
1Y Earn Growth Momentum	$\frac{last year EPS - previous year EPS}{absolute previous year EPS} * 100$
Adjusted Dividend Yield	A stocks return calculated using the capital gains and dividends
Earnings per Share (EPS)	Net income for a period is divided by the total number of shares outstanding
Equity-Asset	Total assets divided by shareholder equity
Earning on Equity	Net income divided by share holders equity . Measure of the net income a firm earns as a percent of stockholders' investment
Adjusted EPS	Calculates earning per share using only normal trading profits and excluding returns made from exceptional items and on offs. These are excluded as they don't help investors estimate future cash flows
Altman Z-Factor	The technique uses a statistical technique to predict the probability of a company's failure
CPS-DPS	Ratio of cash to debt per share
Cash Share Yield	The ratio of the annual return from an investment, through dividend and capital gains, to the amount invested

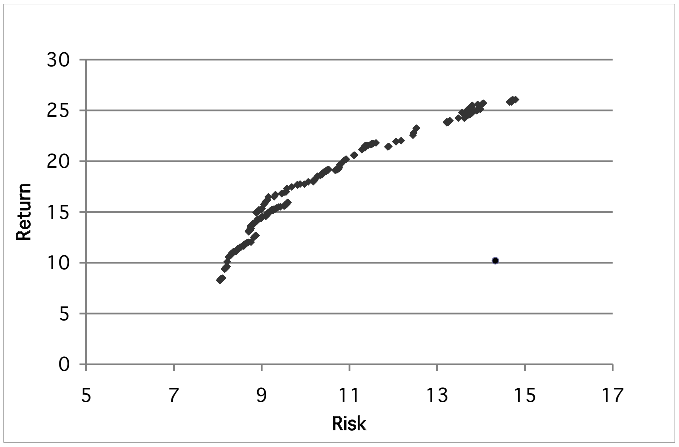

Figure 1: Efficient Frontier of Portfolios. The single dot shows the risk/return of the Index fund.

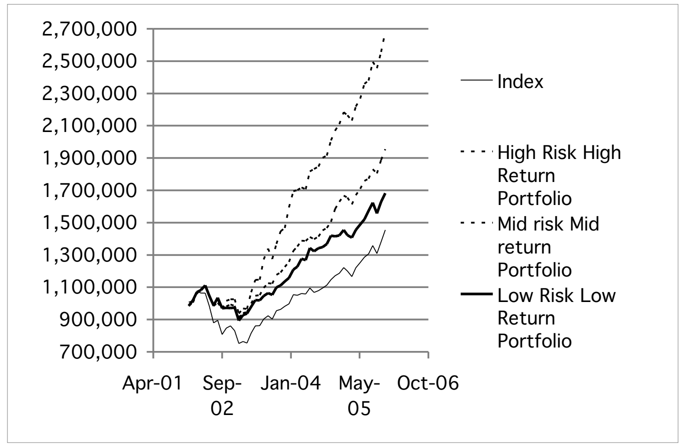

Figure 2: Performance of 3 Investment Strategies

2.3 System Architecture

Our system consists of a multiobjective GP, as well as the embedded simulation of an investment strategy, which is used for fitness assessment of solutions. The MOGP fitness function passes an individual (an equation representing a factor model) to the simulator; the equation is used to rank stocks on a monthly basis during simulation. The rankings drive buy/sell decisions and at the end of simulation various metrics (e.g. return, risk) are returned to the fitness function.

The multiobjective algorithm used is SPEA2 [12]. The implementation (in Java) is based on the ECJ package [6]. Experiments have a population size of 2000, archive size 600, and run for 60 generations. The method of tree generation is ramped half and half [5]. The terminal set for the tree consists of technical and fundamental financial factors describing a company's performance and constants. The set of factors chosen are described in Table 1. The function set includes addition, subtraction, multiplication, division, power 2, and power 3. The MOGP has two conflicting objectives to satisfy; return maximization and risk minimization. Return is defined as the annualized average return, and risk is the standard deviation of the annualized average return. The MOGP solutions are trees, each of which represents a non-linear model of financial factors.

2.4 Financial Data

Research was conducted on historical data from the London Stock Exchange market, the FTSE-100, for 48 months from January 2002 to December 2005. Our stock universe consisted of 82 stocks. For each stock, the data consisted of the monthly values of factors describing the company performance.

In all experiments, our bench mark is the performance of an "Index Fund" over he same period. The index fund was constructed by using the initial sum of one million pounds to invest equally into all eighty two stocks that constitute our traded index.

3. SIMULATION RESULTS

During the training, the model is continuously evolving trying to achieve the best possible objectives values in response to monthly data. At the end of the training, we obtain a set of solutions that represent the best tradeoffs between objectives found so far.

3.1 Efficient Frontier and Performance of Factor Models

The result of merging the efficient frontier of 5 different runs of the MOGP is shown in Figure 1. We extracted three factors models, such that one is achieving a high return, high risk, another medium risk medium return, and the third low risk low return as representatives of three different investment strategies. Using these models to rank assets and form the portfolio, we plotted the fund value for the 48 months training period against the index fund value (invested with equal proportion in the 82 stocks). Results are shown in Figure 2.

3.2 Models Evolved and Financial Factors

French [3], [2] reported that small stocks outperform large stocks, and value stocks (high book to market ratio) outperform growth stocks in the majority of markets and time periods studied. Their research is considered a landmark in multifactor models that explain asset returns. We were interested to investigate which factors were chosen by the MOGP to form the factor models for each of the three risk-return trade-off classes. We plotted a histogram for the frequency that each factor was used in 100% of individuals in each of the risk-return trade-off classes. Results are presented in Figure 3.

Results show evidence that the following factors affect risk-adjusted returns in all solutions:

- price momentum;
- change of return on equity, and
- cash yield.

The *firm size* factor has a greater affect on the *medium* and *low* risk/return strategies. By contrast, the *moving-average-changes* factor has a greater affect on the *high* risk/return strategies. No effect of the *book-to-market* ratio is evident in the period studied.

It is noted that some factors are used as mutual alternatives across runs. We speculate that this is due to factor interaction and different possible ways of combining different factors while affecting stock performance in the same way.

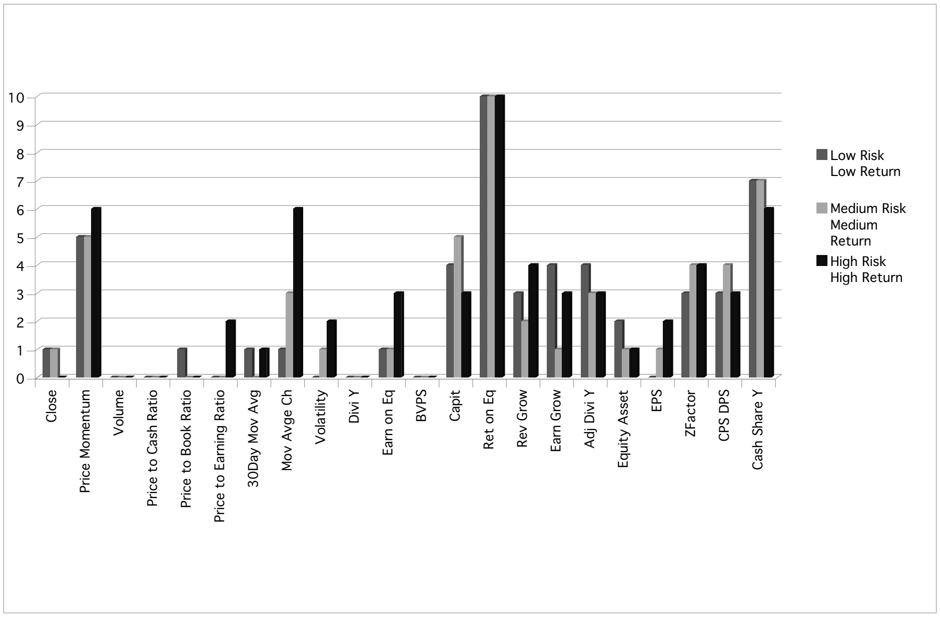

Figure 3: Histogram of Factors used in Investment Strategies Evolved - The y-axis indicates number of runs out of 10

4. CONCLUSION

A fund manager facing the problem of portfolio optimization is often handling various portfolios for different investors. Each investor will have his own set of objectives. Traditionally, the two objectives of risk and return have been considered, with investors varying in their expected return and tolerated risk. If we consider these two objectives, then an efficient portfolio is the one that have the lowest standard deviation given certain expected return, or a portfolio with the highest expected return, given a certain standard deviation. Identifying the efficient frontier is important not only because rational investors are expected to choose a portfolio from the efficient frontier, but also because it can give a sense of the trade-off between objectives. We have shown that the MOGP is a suitable tool to draw the efficient frontier of risk-return trade-off. We have also shown that the MOGP was able to derive a multi-factor model that was used to rank how attractive a stock is for investment. The results further establish the effectiveness of MOEA in financial applications. The use of evolutionary algorithms has the added advantage of being able to view the model and analyse decision based on it. We have provided a preliminary analysis of the set of factors chosen for each class of investment with a general tendency to the tradeoffs of objectives on the efficient frontier. We have found that certain factors are consistently selected as important factors for ranking assets across the three classes of investment such as the return on equity. Some were more important for selecting assets in specific investment class with the capitalization more important in case of medium risk/medium return strategies; moving average changes in high risk/high return strategies as two examples.

5. REFERENCES

[1] McMillan D.G. Non-linear predictability of UK stock market returns. Oxford Bulletin of Economics and Statistics, December 2003.

[2] Eugene Fama. Multifactor portfolio efficiency and multifactor asset pricing. *The Journal of Finance*, 31(4):441–465, December 1996.

[3] Eugene Fama and Kenneth R French. Multifactor explaination of asset pricing anomalies. *The Journal of Finance*, 51(1):55–84, March 1996.

[4] Angelos Kanas. Non-linear forecasts of stock returns. *Journal of Forecasting*, 22:299–315, 2003.

[5] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press Cambridge MA USA, 1992.

[6] Sean Luke et al. A java-based evolutionary computation research system, version 15. http://www.cs.gmu.edu/ eclab/projects/ecj/.

[7] H.M Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[8] M. Hashem Pesaran. A recursive modelling approach to predicting uk stock returns. *The Economic Journal*, 110:159–191, January 2000.

[9] Min Qi. Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business and Economic Statistics*, 17:419–429, 1999.

[10] S.A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(1):341–60, 1976.

[11] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

[12] Eckart Zitzler, Marco Laumanns, , and Lothar Thiele. Spea2: Improving the strength pareto evolutionary algorithm for multiobjective optimization. *Evolutionary Methods for Design, Optimization, and Control, CIMNE*, pages 95–100, 2002.