High Frequency Trading Portfolio Optimisation:
Integration of financial and human factors

Šarunas Raudys and Aistis Raudys
Vilnius University, Faculty of Mathematics and Informatics
Naugarduko st. 24, LT-03225 Vilnius, Lithuania
sarunas.raudys@mif.vu.lt; aistis@raudys.com

Abstract—To use human factors together with financial ones in portfolio management task we analyze lengthy series of successes and losses of numerous automated high frequency trading systems that buy and sell assets. We found that in spite of sparse, bimodal non-Gaussian time series, modern Markowitz solutions can be applied to weigh up contributions of diverse trading strategies. Training history should be rather short in situations where technological, social, financial, economic and political situations are changing swiftly. The Markowitz portfolio coefficients finding algorithm can be improved by careful application of the regularization and matrix structurization methods.

Keywords—high frequency trading; Markowitz; portfolio optimisation; regularization; sample size

I. INTRODUCTION

Portfolio management (PM) task is one of very important research topics in financial markets (see e.g. classical Markowitz paper [1, 2] and reviews of thousands of papers aimed to improve the Markowitz’s solution [3, 4, 5]). During the last decade the financial markets turned towards the high frequency trading [9, 10]. It potentially promises good returns with relatively small risk [11], as positions are rarely held overnight. Positions typically last from a few seconds to a few hours or a day. This line of attack minimises the risk of big losses. A gain is possible due to very short term market inefficiencies.

Historically, majority of works aim at creating portfolio for a set of asset – an investment universe. We have rather different approach here. In real life we are facing hundreds and thousands of diverse trading strategies that acting together are influencing the financial markets. Diverse agents (market participants) are using individual portfolio managing strategies, decision making algorithms differing in trading logic, assumptions, concrete compositions of the multidimensional time series, exactness and timing of the empirical data, and methods used to estimate their statistical characteristics. Development of technologies and the society, information explosion, political and psychological aspects (a length of investment horizon, mispricing, investor credulity, preference to take risk, a desirable gain, a time of expected profits, etc., - we refer to them as human factors [6, 7, 8]), economic and financial indexes are changing swiftly.

For that reason, we analyse portfolio construction methods based on a large number of trading strategies. What distinguishes the new PM approach from previous ones is that here one does not create the portfolio based on assets. One creates the portfolio based on success and losses of automated trading systems that trade assets. On top of that, the underlying trading systems are short term / high frequency, so generated returns are vaguely correlated to variations of the underlying instruments.

A. Profit and loss (PNL) data matrix

Many aspects of investment analysis are said to be psychological in nature [6]. In order to manage portfolio “optimally” one needs to take into account both the financial and human factors. One may expect that diverse strategies are employing similar profit and loss data, however, strategies incorporate into the decision making algorithm their designers’ human factors. Therefore, one of the ways to take into account both types of factors is to consider long lasting history of the performance of a large number of the high frequency trading (HFT) strategies.

To integrate both kinds of information, one needs to simulate the large number of trading strategies (say, p ones) during a long time period (say n time moments) and calculate virtual profits obtained if each one of the p strategies would be applied. In such a way one obtains a n×p dimensional profit and loss data matrix

\[ X = [X_1^T, X_2^T, ..., X_n^T]^T, \]

where \( X_i = [x_{i1}, x_{i2}, ..., x_{ip}] \) are p-dimensional data vectors-
rows and superscript "^T" denotes a transpose.

Portfolio management approaches based on the assets and based on the successes of trading systems are two different lines of attack. Important characteristic feature of latter one is a fact that the data’s larger part (72% in this case) consists of zeros. The distribution of the data becomes bimodal. If the number of trading strategies, p, is large, small sample problems arise in PM. Sparse data matrix requires that the aggregation (fusion) rule should be as simple as possible. Thus, we will use linear rule, i.e. weights \( W = (w_1, w_2, ..., w_p)^T \) will be availed to assess p first level decisions

\[ PNL_i = w_1 x_{i1} + w_2 x_{i2} + ... + w_p x_{ip}, \quad (w_i > 0) \] (1)

Sample based average \( \bar{X} \) and variance \( \text{var}_{PNL} \), of returns \( PNL_i \) are

\[ \bar{X} = \frac{1}{n} \sum_i X_i, \quad \text{var}_{PNL} = \frac{1}{n} \sum_i (X_i - \bar{X})^2, \]

where \( \bar{X} \) and \( \text{var}_{PNL} \) are training data based mean vector and covariance matrix, \( \bar{X} = \bar{X}_T \) - \( \bar{X} \) are centred p-dimensional data vectors. Finding the weight vector \( W \) can...
be performed as calculations of an efficient frontier for a number of parameter $q_i = W\mathbf{X}^T$ values, as is done in the standard portfolio management where proportions of various assets are determined [2]. For a set of a priori selected $q_i$ values, $q_1$, $q_2$, ..., $q_b$, the efficient frontier is found by minimizing $W\mathbf{S}_{\text{train}}W$. Minimization can be done by using a Lagrange multiplier.

Here we maximize a probability that return (1) is smaller as any a priori given value $P_{\text{given}}$ provided a distribution of the return (1) is Gaussian. The procedure actually maximizes the Sharpe ratio (sample average/standard deviation)

$$Sh = \frac{\text{PNL}}{\text{stdev}_{\text{PNL}}} = \mathbf{X}W(W^T\mathbf{S}_{\text{train}}W)^{-1/2},$$

where one is obliged to satisfy following two constraints:

$$w_j > 0 \ (s = 1, 2, \ldots, p), \quad \text{and} \quad \sum_w w_j = 1. \quad (3)$$

One more way to estimate $W = (w_1, w_2, \ldots, w_p)$ is straightforward: one generates very large number, say $m$ random vectors $W_j$ ($j = 1, 2, \ldots, m$), weighs up their quality and selects the best. If distribution of inputs is a mixture of multivariate distributions, instead of the Sharpe ratio (2) one would need to employ more complex non-linear objective functions [12,13]. Similar situation we face in our approach where majority of data points are “zeros”. Complex objective functions, however, require more training data [14]. So, we use simple linear profit and loss model and apply traditional Sharpe criterion

**B. Sample size issues in HFT portfolio creation**

The world is changing in a chaotic manner. Financial markets are affected by modern data storing, processing and transmission technologies. Objective and subjective active human interferences and intelligent mathematically based methods are changing even faster. Speaking in general, a lot of popular indicator-based trading rules turn out to be loss-making when applied individually in a systematic manner. Successful traders tend to adapt to market conditions by ‘dropping’ trading rules in situations where they become loss-making or when more profitable rules are found [16, 17]. This aspect forces the decision making algorithms adapt to changes as rapidly as possible.

In dynamic financial time series prediction, a decision rules based on short data sequences result more accurate predictions as using lengthy historical data. In [18] for a random drift model of the changes and multidimensional time series data model it was shown that the generalization error decreases at first with an increase in the length of training history, reaches minimum and starts increasing afterwards. This conclusion was confirmed in empirical analysis of prediction of commodity prices. The dilemma with optimal training set size cannot be avoided in the portfolio management too. In situations where time series are changing in time, genetic algorithms [16], evolvable multi-agent systems [19] can be used.

In Figure 1 we present dependence of the Sharpe ratio on the length of historic data used to find weight vector $W = (w_1, w_2, \ldots, w_p)^T$ in $p = 142$ dimensional space. We used artificially generated data of 142 trading strategies based on 8 years financial data time series history and the standard Markowitz algorithm “frontcon” realized in Matlab Financial Toolbox. Underneath it uses proprietarily algorithm QPLCPRG - positive-semi-definite quadratic programming based on linear complementary programming. In our approach, larger part of the data consists of zeros. In order to avoid the singularity of covariance matrix, we added constant $h = 0.1$ to diagonal elements of the matrix. We will call this algorithm as “standard Markowitz”.

This example illustrates a necessity to shorten the length of training history if the data are changing very often. If the size of training data is finite and dimensionality, $p$, is high, problems arise with estimation of the covariance matrix. Especially this question is important in our problem setting, where the data is sparse. In the eight years 142 dimensional data, 72% of positions were filled with zeros - prudent trading strategies refuse from active transactions very often. Possible ways to mitigate small sample size/dimensionality problems in machine learning are: 1) to use regularization, 2) apply diverse constraints (simplified theoretical models of the covariance matrix), 3) add a noise to training data [20].

**II. FUSION METHODS TO CONSTRUCT HFT PORTFOLIO**

**A. Sequential forward selection method – Comgen**

Modern portfolio theory in an inexplicit way assumes that PNL (or assets returns) are Gaussian random variables, variances and correlations between inputs (assets) do not change. In order to avoid the assumptions just mentioned in one of the fusion procedures investigated we applied rule with integer weights. Here sequential forward selection procedure was used to select the best subset of inputs (the HFT strategies). In training phase, the quality of the portfolio was evaluated on the particular training data set. Usage of this method was suggested by our time series provider – proprietary trading firm and was named “Comgen”, the portfolio Combination Generator.

In this rule initially we start with the best trading system, $T_{w}$, with the highest Sharpe ratio. In the second step we
consider \( p \) pairs, \( T_o^+T_r, \) \((j = 1, 2, \ldots, p)\), and select the pair with the highest Sharpe, say \( T_b = T_o^+T_r \) \( (r \) is the index of the best pair). And again, we select the best pair \( T_c = T_o^+T_r, \) \( (j = 1, 2, \ldots, p)\), where \( T_b \) contains previously selected systems. This process continues by adding the single system that appears to be the best when combined with previously selected trading systems. In each step we add value 1 to a selected position in the weight vector. As a result we get integer weights starting from 0. We repeat the procedure until we reach \( p \) - the maximum number of trading systems in our portfolio \( (p = 142 \) in this particular task).

The benefits of this system is in its simplicity and granularity, it creates integer portfolio weights that can be traded straight away. In principle, other heuristic data mining variable selection procedures such as backward selection, genetic algorithms or diverse variants of random search can be applied (see, e.g. [15] or [20], Section 6.3.2).

B. Standard weakly regularized Markowitz algorithm for aggregation of PM strategies

In our research we use two modifications of the Markowitz algorithm. The first one was the standard Markowitz algorithm ("fronticon" function in Matlab). As inputs served sample training data based mean vector \( \bar{X} \) and weakly regularized estimate of covariance matrix, \( S_{\text{train} \& \text{reg}} = S_{\text{train}} + hI \), where \( \bar{X} \) and \( S_{\text{train}} \) were defined above, \( h \) is very small regularization constant (we used \( h=0.1 \), when largest diagonal elements of matrix \( S \) exceeded \( 10^5 \)) to avoid the singularity of matrix \( S_{\text{train}} + hI \). To satisfy requirements (3) we brought negative weights to naught, if \( w_l \leq 0 \) and normalized \( W \) to meet with \( \sum w_l = 1 \).

The benefits of this system is in its simplicity and granularity, it creates integer portfolio weights that can be traded straight away. In principle, other heuristic data mining variable selection procedures such as backward selection, genetic algorithms or diverse variants of random search can be applied (see, e.g. [15] or [20], Section 6.3.2).

To satisfy requirements (3) we brought negative weights to naught, if \( w_l \leq 0 \) and normalized \( W \) to meet with \( \sum w_l = 1 \).

In the case of multimodal non-Gaussian data the complex methods can be less efficient as simple ones. Therefore, we used the training data based regularized matrix \( S_{\text{train} \& \text{reg}} = S_{\text{train}} + hI \) with large value of parameter \( h \) in our weight seeking algorithm. This regularization technique actually reduces the correlations. For that reason, we introduced a way to take into account “principal dependencies”: we started to use the first order dependence tree [25, 26, 20] to improve estimate of the covariance matrix.

The first order dependence tree model assumes each component of the input vector depends directly upon only one other component. The probability density function of random vector \( X = (x_1, x_2, \ldots, x_n) \) having the first order tree type can be written in the very simple form:

\[
p(x_1, x_2, ..., x_n) = \prod_{j=1}^{n} p(x_j | x_{m_j}) \quad (0 \leq m_j \leq n),
\]

where a sequence \( m_2, ..., m_n \) constitutes a graph of connections, \( m \) (an unknown permutation of the integers 1, 2, ..., \( n \)), and \( p(x_j | x_{m_j}) \), by definition, is equal to \( p(x_j) \); for details see [26] or [20], Appendix A2. In this model one estimates only \( p \) variances and \( p-1 \) correlations. An inverse of estimate \( S \) has only \( 2p-1 \) nonzero elements. In a general case, the matrix \( S \), however, is composed of \( p^2 \) nonzero elements. If assumptions concerning the first order dependence structure are correct, we can obtain significant improvements in accuracy in finite sample size/high dimensionality situations.

In our weight \( W \) determining algorithm we incorporated both regularization means, the regularization with \( h=30,000 \) and the first order dependence structurization. Use of the penalty terms in Eq. (4) tolerates the inaccuracies arising in the weight vector estimation. In such a way, the penalty terms serve as the third additional “regularizer” in the weight vector seeking algorithm. Therefore, we will refer to the novel modification as the “Three times regularized Markowitz algorithm”.

Both Markowitz PM based algorithms create fractional weights. One has to scale and round them to get tradable weights (one cannot trade fractional part of the trading system). The downside is that the rounding reduces portfolio accuracy and the Sharpe ratio (−8-10%).

\[
\begin{align*}
\frac{\partial S_{\text{reg}}}{\partial W} = 2WS_{\text{train} \& \text{reg}} + 2A(W \bar{X}^T - q_j) \bar{X} + 2A(W^T - 1) = 0, \quad \text{and} \\
W = (q_j \bar{X} + 1) (S_{\text{train} \& \text{reg}} / A + \bar{X}^T \bar{X} + 1^T 1)^{-1}.
\end{align*}
\]

698 2011 11th International Conference on Intelligent Systems Design and Applications
III. SIMULATION STUDY

A. An objective of the study

The objective is to compare efficacies of three novel schemes of PM rules with benchmark one:

a) equally weighted - when all HFT schemes were evaluated with the same weights ($w_1 = w_2 = \ldots = w_p = 1$). This is the original weighting as proprietary trading house was using (the benchmark method);

b) Comgen, our heuristic algorithm with integer weights obtained in sequential forward selection;

c) standard Markowitz algorithm - linear weighting of $p$ PM trading strategies realized in Matlab Financial Toolbox;

d) three times regularized Markowitz algorithm - linear weighting of $p$ HFT trading strategies with the novel modification of the Markowitz algorithm (Eq. (6)), simple regularization, $S_{\text{train+reg}} = S_{\text{train}} + hI$, and the first order dependency tree structure (Eq. (7)) of the covariance matrix.

B. Data used

From the high frequency trading firm we received a set of 142 profit and loss series representing simulated track records of real life trading systems. The firm was trading them equally weighted and the aim was to create a portfolio maximizing risk-reward ratio – the Sharpe ratio. These systems were selected qualitatively by the analyst looking at maximization of the Sharpe ratio. These systems allows increasing the Sharpe ratio.

All experiments were performed in out-of-sample (oos) regime using walk forward methodology. One trains the model with $N$ $p$-dimensional vectors of time series and tests on subsequent $M$ vectors. Next, one walks $M$ days forward: we shift the training and validation sets’ windows by $M$ days. We used 10 walk forward iterations with $M = 100$ day period. So, for each size of training set, we trained the fusion algorithms ten times. As a result we got total of $10 \times 100 = 1000$ days out of sample trading history of values $XW_i$ ($i = 1, 2, \ldots, 10$), to estimate the Sharpe ratio.

To have fair comparison, in both Markowitz PM algorithms the training set information and $R = 50$ values were used to calculate the efficient frontiers. The portfolios were designed on training samples of length $N = 200$, $300$, $\ldots$, $1000$ and tested on following $M$ days.

In the first series of the experiments we considered $p=142$ HFT strategies (trading robots). Results obtained from single data set study with 2100 days history are not satisfactorily conclusive. For that reason, to increase the validity of experimental results we formed $K=25$ randomly chosen sets of HFT strategies composed of $p=100$ trading robots each:

- From 142 set select 100 HFT strategies.
- Perform walk forward portfolio optimisation of all four methods using:
  - $N=200$, $300$, $\ldots$, $1000$ days training set and $M=100$ days forward periods,
  - Compute the Sharpe ratios on all four systems in out of sample data.
- Repeat the experiments $K=25$ times.
- Results are averaged from 25 experiments

In such experiment design instead of single estimates we were able to evaluate averages and standard deviations of the Sharpe ratio.

D. Results

The experiments performed with (c) and (d) aggregation schemes showed that both Markowitz’s solution based fusion schemes require regularization $S_{\text{train+reg}} = S_{\text{train}} + hI$. After short test, for all future experiments we selected $h=1000$ for standard Markowitz algorithm and $h=30,000$ for the new modification. In Fig. 2a we present graphs of dependence of the Sharpe ratio on training sample size in the single series of experiments in experiments with $p=142$ trading schemes.

In Fig. 2b we present average values obtained in $K=25$ series of experiments with $p=100$ randomly selected collections of trading systems. In Fig 3a we provide more details of the experiments with the 25 runs. We show mean values, $m$, for both variants of the Markowitz algorithm calculated for 9 diverse sample sizes. We also demonstrate deviations from $m$ by two standard deviations of the means. In Fig. 3b we present a scatter diagram of Sharpe ratio. The graphs and the diagram indicate that the improvement of the Three time regularized Markowitz rule is significant.

E. Evaluation of the results

Experiments show that the portfolio based on the weighted evaluation of success of automated trading systems allows increasing the Sharpe ratio of the portfolio. In comparison, trading the assets is notably less successful as one has to absorb not only up trends but down trends as well. It means that the information acquired from analysis of multiplicity of the HFT strategies was useful.

The benchmark ("equally weighted") algorithm was notably outperformed by the Comgen. The standard Markowitz resulted in similar accuracy (Fig.2). In Fig.2a we fused $p=142$ single HFT schemes and Fig.2b we see averages of 25 additional experiments with 100 randomly selected trading schemes. We can see that additional regularization ($h=1000$) is very useful to increase the profit.
The regularization and covariance matrix structurization are necessary if the fusion weights are found from Eq. (6). Value of the optimal regularization is not very crucial parameter. Differences between mean values of the Sharpe ratio of the novel and old methods are larger than two standard deviations/√K for all values in training sample size.

IV. CONCLUDING REMARKS

To take into account human factors acting together with financial and economy ones we started analyzing long lasting histories of successes of employment of diverse HFT strategies in the portfolio management tasks. Instead of creating the portfolio based on assets we create the portfolio based on successes and losses of multiple automated trading systems that trade the assets.

The first difficulty we faced was sparse, bimodal time series data. We found that in spite of the non-Gaussian distribution of the training data, the Markowitz solution can be successfully applied to weight contributions of diverse HFT strategies in automatic portfolio design. We become aware of that the training data history should be rather short in situations where the technological, financial, economic and political situations are changing swiftly. The tolerances used in optimisation moderately fix sample base estimates of the covariance matrix in the Matlab Financial toolbox. The Markowitz PM algorithm can be additionally improved by application of the regularization and covariance matrix structurization methods. The simulation experiments showed that regularization techniques reduce the decline in the accuracy and the Sharpe ratio when the sample size is decreasing or increasing too much. This fact gives a hope that the novel methodology could be applied in the future if the environmental situations will change more abruptly and unexpectedly.

We also state that sample size is important in portfolio management task. From our experiments we recommend at least 400 days of history to include in PM task. In the case of regularized covariance matrix this does not seem to significantly change further increasing training size. That of
course is subjective and may depend on the number of trading systems, types of the systems and a frequency and power of the changes. The algorithmic trading is affected by the outbreak of use of modern computers and telecommunications means and counts its history a decade of years. The HFT activities are affecting the markets. For that reason, the data histories useful for comparison of different trading methods are not sufficiently lengthy. Like in pattern recognition, the sample size/complexity considerations [14] are very important in portfolio management research in rapidly changing environments. We considered only a small number of means that are useful in this task, however, the number of covariance matrix simplification tools is much larger and constitutes a subject for future wide scale investigations.

The Efficient-Market Hypothesis (EMH) asserts that the market is efficient and whenever new information comes up, the market absorbs it by correcting itself [27, 28, 29]. Simultaneous analysis of numerous long lasting success histories of diverse trading robots brings in additional information and generates the model where it is assumed that the future asset prices depend on the predictions of all participants of the market. It provides some theoretical basis for explanation of EMH and can be useful for analyzing simulation results obtained by exploration of prediction models of the financial time series when asset prices depend on predictions. Theoretical considerations and preliminary experiments showed that proposed regularization methodology can also be used for liquidity analysis and transaction cost management.

ACKNOWLEDGMENT

This research was funded in part by a grant (No. MIP-20/2010) from the Research Council of Lithuania. The authors also want to express their appreciation to Vilnius University.

REFERENCES