Portfolio Optimization Using Preference Relation Based on Statistical Arbitrage

Lovre Mrčela, Andro Merćep, Stjepan Begušić, and Zvonko Kostanjčar

Faculty of Electrical Engineering and Computing
University of Zagreb
Zagreb, Croatia

***Abstract*—We propose a new algorithm for portfolio optimization based on statistical arbitrage, that uses a multi-criteria decision making approach to obtain the most preferred assets. A preference flow graph of financial assets is constructed at each time step, with the aid of statistical arbitrage algorithm that describes preferences among the assets. Then, the individual preferences for each asset are obtained by using the potential method, and the most preferred assets are selected into the portfolio in accordance to them. A consistency measure of the preference flow graph is also obtained using the same method, and it measures the reliability of the decision making.**

The proposed method has been tested on a selection of S&P 500 constituent stocks from 1980 to 2004. The results indicate that the proposed method performs well in the considered market, which is indicated by high Sharpe ratios of the constructed portfolios. We also report that the algorithm performs better when provided with a larger number of assets, showing that the increased number of considered assets provides more insight into the market behavior.

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

With the rapid growth in amount of data and the increase in the frequency of decisions to be made, financial decision making is quickly becoming one of the most data intensive fields of interest [1], [2]. Both researchers and practitioners resort to quantitative and computational methods to analyze financial data, and thus enhance the decision making process, traditionally done by experts [3]. More specifically, the issues of asset allocation and portfolio optimization [4] remain some of the critical points where computational methods are needed in order to scale the traditional financial methods to the global universe of a large number of assets and a high frequency of decision making [5].

Here we consider the method of statistical arbitrage — a well-known algorithmic trading approach based on inefficiencies between asset prices [6], [7]. Classical statistical arbitrage methods take into account pairs of assets (e.g. stocks, bonds, commodities, etc.) where prices behave similarly to each other during a certain period of time [8]. Similarity is measured by cointegration, correlation, or some other relevant measure. Among those pairs, method finds a moment in time when those assets' prices overcome the interval that was statistically determined as highly confident. When such opportunities present themselves, one can take advantage of them by predicting that prices will likely return once again to the confident interval in the future, and trade in accordance with this prediction. This approach was initially defined on pairs of assets, and the problem of expanding it to larger sets of assets remained an open issue.

In this paper, we propose a new method based on predictions obtained by the statistical arbitrage method, using statistical measures as a proxy for describing the preference relations between pairs of assets. The idea of this method is to create a generalization of statistical arbitrage that is more robust and performs better when working with a larger number of assets by trying to take into account interaction of multiple assets [9]. This new method supplements the statistical arbitrage method by introducing a preference relation graph that has the potential to grasp total interaction among all the assets, whereas the former method only observes relations of individual pairs of assets. A graph is formed based on the estimated pairwise relations between a selection of assets. This graph imposes a preference relation among the assets that are included in it. Using the potential method [10], a multi-criteria decision making approach, we sort the considered assets by preference and include them into the portfolio accordingly.

We test the proposed method on 203 stocks which were constituent members of the S&P 500 market index from 1980 to 2004. Our results demonstrate the validity of the proposed approach and indicate that portfolios yielding high Sharpe ratio values can be obtained using the proposed method.

II. Concepts and Methods

A. Preference relation and utility function

Let Ω be any set of entities. Preference relation $\succ$ defined over $\Omega \times \Omega$ is a strict weak ordering that describes the way one entity is preferred over another. This relation is specific in that it is ($\forall x, y, z \in \Omega$):

- *irreflexive*: every entity x can not be preferred over itself,
- *asymmetrical*: if x is preferred over y, then y is not preferred over x,
- *transitive*: if x is preferred over y, and y is preferred over z, then x is also preferred over z,
- *transitive in incomparability* (noting that x and y may be *incomparable*, i.e. neither x is preferred over y, nor y is

978-1-5386-2101-1/17/$31.00 2017 IEEE

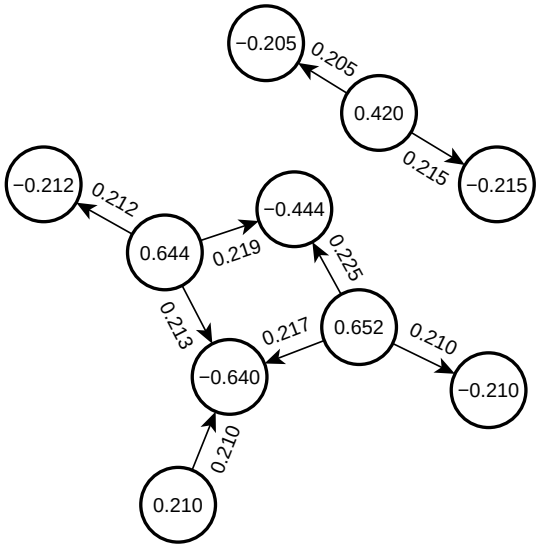

Fig. 1. An example of a preference flow graph. Preference is inscribed in each node, and preference flows are shown on edges.

preferred over x): if x is incomparable with y, and y is incomparable with z, then x is also incomparable with z.

We naturally assume this kind of relation when describing preferences among the assets. Deciding that one asset is more preferable than the other is an easier task than assigning a measure of preference to each asset directly, especially when dealing with a larger numbers of assets. However, the latter is more useful than the former; thus it is desirable to find a way of sorting assets in the order of preference.

A utility function $U\colon \Omega \to \mathbb{R}$ is a function that maps entities to real numbers, so that the ordering of the mapping reflects the ordering of the entities according to their preference, i.e. $\forall x, y \in \Omega,\ x \succ y \Leftrightarrow U(x) > U(y)$. One such mapping is obtained using the potential method that is described later in this paper. In addition to ordering the entities, the utility function also quantifies the relation by introducing the intensity of preference, i.e. when $x \succ y$, then x is more preferable than y by $U(x) - U(y)$. Hence it is more informative when it comes to decision making.

B. Preference flow graph

A preference flow graph is a weighted directed acyclical graph, whose nodes represent entities, edges represent preference for one entity over another, and edge weights correspond to the intensity of the preference. If an edge between nodes is missing, it is considered that neither entity is preferred over another (incomparability of entities). The graph as a whole describes preference flow among the entities. An example of a preference flow graph is shown in Fig. 1.

The construction of the graph is based on a statistical arbitrage method. A directed edge from node i to node j with weight $w_{i,j}$ exists in the graph if and only if assets represented by nodes i and j have demonstrated similar behavior during a considered lookback period, but have suddenly diverged at the moment, as determined by statistical measures. The weight $w_{i,j}$ corresponds to the magnitude of this divergence. A detailed description of used statistical measures the procedure is given later in III-A.

Connections in this graph impose a preference relation, and it already satisfies two following properties: neither node is in relation with itself (*irreflexivity*), and multiple connections are not allowed between any two nodes (implies *asymmetry*). However, problems arise with the aforementioned properties of *transitivity*, and *transitivity in incomparability*, which may not hold for an arbitrarily constructed instance of the graph [10]. This imposed preference relation should preferably be in compliance with all the aforementioned properties, but when it comes to larger number of entities, it may become infeasible to construct a graph of such qualities directly. Instead of aiming at a consistent preference relation, we use a consistency measure that describes similarity between the original graph and its nearest consistent reconstruction (i.e. a reconstruction which imposes a consistent preference relation), and use it as an additional parameter in decision making.

C. Potential method

From the obtained preference flow graph it is possible to tell which pair of assets has the highest preference flow. However, it is not yet possible to directly tell which are the most or least preferred assets, or obtain a measure of preference for individual assets. To calculate preferences for each node in the graph, we use the potential method [10]. The potential of a node corresponds to the difference of amount of flow directed towards and from the node.

For the observed graph $\mathcal{G}$, let there be a total of N nodes, and maximum of $E = \binom{N}{2}$ edges (in case of a complete graph). If $\mathcal{G}$ is not complete, we complete it by adding edges to it with weight 0 (the direction is irrelevant), thus forming a complete graph $\mathcal{G}$ with E edges. Let $\boldsymbol{B} \in \mathbb{R}^{E\times N}$ be the incidence matrix of $\mathcal{G}$. Let $\boldsymbol{f} \in \mathbb{R}^{E}$ be a vector containing edge weights (i.e. preference flows). Order of the edges must be the same as order of the edges in $\boldsymbol{B}$. As mentioned before, in place of missing edges we simply put 0. Let $\boldsymbol{\phi} \in \mathbb{R}^{N}$ be the vector containing potentials of each node, in order that is the same as order of the nodes in $\boldsymbol{B}$. Now, if $\mathcal{G}$ was consistent, then $\boldsymbol{B}$, $\boldsymbol{\phi}$, and $\boldsymbol{f}$ would satisfy the equation

$$\boldsymbol{B}\boldsymbol{\phi} = \boldsymbol{f}. \tag{1}$$

The equation (1) states that the difference between the potential of any two nodes should result in the weight of the edge between them. This is only possible for consistent graphs, and most of the time the obtained preference flow graphs will be inconsistent. In that case we try to find an approximate solution $\boldsymbol{\phi}^*$ that minimizes the square error:

$$\boldsymbol{\phi}^* = \arg\min_{\boldsymbol{\phi}} \left\{ \|\boldsymbol{B}\boldsymbol{\phi} - \boldsymbol{f}\|^2 \right\}$$
$$\Downarrow$$
$$\frac{\partial \|\boldsymbol{B}\boldsymbol{\phi}^* - \boldsymbol{f}\|^2}{\partial \boldsymbol{\phi}^*} = \boldsymbol{0}. \tag{2}$$

Solving (2) via commonly used techniques of matrix calculus results in the following equation:

$$\boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{\phi}^* = \boldsymbol{B}^\top \boldsymbol{f}. \tag{3}$$

The equation (3) determines $\boldsymbol{\phi}^*$ up to a constant (i.e. solution has one degree of freedom), so the following constraint is also included:

$$\boldsymbol{j}^\top \boldsymbol{\phi}^* = 0, \tag{4}$$

where $\boldsymbol{j}$ is vector of ones of the same dimension as $\boldsymbol{\phi}^*$. This ensures a unique solution for which total amounts of positive and negative potential will be equal. Joining the previous two equations together by adding (4) to each row in (3) results in:

$$\begin{aligned} \boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{\phi}^* + \boldsymbol{J}\boldsymbol{\phi}^* &= \boldsymbol{B}^\top \boldsymbol{f} \\ \left[\boldsymbol{B}^\top \boldsymbol{B} + \boldsymbol{J}\right] \boldsymbol{\phi}^* &= \boldsymbol{B}^\top \boldsymbol{f}, \end{aligned} \tag{5}$$

where $\boldsymbol{J}$ is a matrix of ones of the same dimensions as $\boldsymbol{B}^\top \boldsymbol{B}$. Finally, solving (5) for $\boldsymbol{\phi}^*$ gives:

$$\boldsymbol{\phi}^* = \left[\boldsymbol{B}^\top \boldsymbol{B} + \boldsymbol{J}\right]^{-1} \boldsymbol{B}^\top \boldsymbol{f}. \tag{6}$$

Furthermore, the term $\left[\boldsymbol{B}^\top \boldsymbol{B} + \boldsymbol{J}\right]^{-1}$ in (6) may be simplified to $\frac{1}{N}\boldsymbol{I}$ due to $\boldsymbol{B}^\top \boldsymbol{B}$ being the Laplace matrix of a complete graph; thus (6) can be simplified futher into

$$\boldsymbol{\phi}^* = \frac{1}{N} \boldsymbol{B}^\top \boldsymbol{f}, \tag{7}$$

to get a more computationally optimal expression.

Afterwards, we can calculate the consistent reconstruction $\boldsymbol{f}^*$ of preference flow by simply plugging back $\boldsymbol{\phi}^*$ into (1):

$$\boldsymbol{f}^* = \boldsymbol{B}\boldsymbol{\phi}^*. \tag{8}$$

The reconstructed preference flow $\boldsymbol{f}^*$ compared to the original preference flow $\boldsymbol{f}$ may contain some new edges, lose some old edges, or both. In addition, $\boldsymbol{B}$, $\boldsymbol{\phi}^*$, and $\boldsymbol{f}^*$ now describe a consistent graph $\mathcal{G}^*$. It is now possible to define a consistency measure κ, as follows:

$$\kappa = \frac{\|\boldsymbol{f}^*\|}{\|\boldsymbol{f}\|}. \tag{9}$$

Equation (9) represents the cosine of the angle between $\boldsymbol{f}$ and $\boldsymbol{f}^*$ in the column space of matrix $\boldsymbol{B}$. The consistency measure κ describes how consistent graph $\mathcal{G}$ is compared to the $\mathcal{G}^*$. It ranges from 0 to 1, 0 meaning full inconsistency (virtually unreachable), and 1 meaning full consistency.

III. Algorithm

We consider a total of N assets throughout a period of D days. Let the price of asset i at the time step t be $a_i^{(t)}$, for $i \in [1, 2, \ldots, N]$ and $t \in [0, 1, \ldots, D-1]$. The log–prices $b_i^{(t)}$, and log–price differences $c_{i,j}^{(t)}$ between assets i and j are obtained as follows:

$$b_i^{(t)} = \log\left(a_i^{(t)}\right) \tag{10}$$

$$c_{i,j}^{(t)} = b_i^{(t)} - b_j^{(t)}, \tag{11}$$

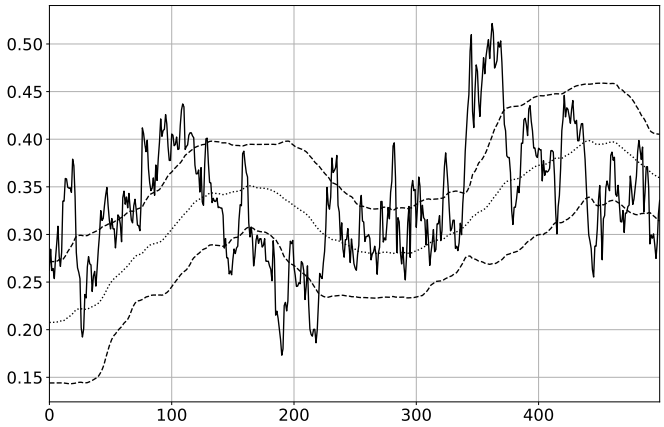

Fig. 2. Log price difference between a pair of assets (i, j) during a period of 500 time steps. The dotted line represents the mean value, and the region between two dashed lines represents the α – standard deviation range from mean value, here $\alpha = 1$; both are calculated on a rolling window of length T, here $T = 120$.

and rolling means $m_{i,j}^{(t)}$ and standard deviations $d_{i,j}^{(t)}$ of log price differences over the lookback period of size T are obtained as follows:

$$m_{i,j}^{(t)} = \frac{1}{T} \sum_{\tau=t-T}^{t-1} c_{i,j}^{(\tau)} \tag{12}$$

$$d_{i,j}^{(t)} = \sqrt{\frac{1}{T} \sum_{\tau=t-T}^{t-1} \left(c_{i,j}^{(\tau)} - m_{i,j}^{(t)}\right)^2}. \tag{13}$$

Note that in summation used in (12, 13) the time step t was intentionally excluded, therefore summation goes only to $t-1$. We use these calculations as basis for creating a portfolio based on preference relation and statistical arbitrage.

A. Preference flow graphs inference

Using the obtained $c_{i,j}^{(t)}$, $m_{i,j}^{(t)}$, and $d_{i,j}^{(t)}$ it is now possible to create a graph of preference flow among assets for each time step t. Considering one time step t, we find all such pairs of assets (i, j) that satisfy:

$$\left|c_{i,j}^{(t)} - m_{i,j}^{(t)}\right| > \alpha \cdot d_{i,j}^{(t)}, \tag{14}$$

i.e. current log–price difference is at least α deviations distant from mean value of the past time window. An illustration is shown in Fig. 2.

Afterwards, for each observed pair (i, j) that exceeds the threshold we add into graph vertices i and j, with a weighed edge of weight $w_{i,j}^{(t)}$ going from i to j. Weight $w_{i,j}^{(t)}$ is obtained as:

$$w_{i,j}^{(t)} = \left(c_{i,j}^{(t)} - m_{i,j}^{(t)}\right) \Big/ d_{i,j}^{(t)}\ . \tag{15}$$

Thus it is possible to create a graph of preference flow for each time step $t \in [T, T+1, \ldots, D-1]$. At some time steps it is possible that the graph could be empty, if it is the case that no pair (i, j) satisfies (14). Setting lower values for parameter α yields denser graphs, and setting $\alpha = 0$ always yields complete graphs.

B. Asset selection from preference flow graphs

We obtain the preference for each individual asset via the potential method, as described earlier in II-C. By obtaining the measure of preference for each asset it is possible to pick assets for the portfolio. The most preferred assets should be bought while the least preferred should be short-sold if possible.

Let $\boldsymbol{\phi}^{(t)} = \begin{bmatrix} \phi_1^{(t)} & \phi_2^{(t)} & \dots & \phi_N^{(t)} \end{bmatrix}$ denote the vector of preferences of assets at time step t and let $\phi_i^{(t)}$ denote the preference for asset i at time step t. When selecting assets for the portfolio we take into consideration the consistency measure κ as well. Lower values of κ suggest that we should diversify our portfolio by including some more assets in the order of preference, while higher values suggest that it is safe to invest in smaller number of assets. Portfolio diversification might be seen as a strategy for protection from fundamental risks, e.g. risk of asset default.

The bound on the assets which will be taken into portfolio is proportional to the consistency measure κ. Depending on the nature of assets we may tune the consistency measure κ to be more or less inclined to diversification by transforming it to κ':

$$\kappa' = a + (1 - a)\kappa^b, \tag{16}$$

where $a \in [0, 1], b \in \mathbb{R}^+$. For default values of $a = 0$, $b = 1$, κ' equals κ.

For determining the assets that should be held in the portfolio at time step t, we find such assets i for which holds:

$$\phi_i^{(t)} \geq \kappa' \cdot \Phi, \tag{17}$$

where Φ is $\max_j \{|\phi_j|\}$. Likewise, for short-selling we choose those assets i for which holds:

$$\phi_i^{(t)} \leq -\kappa' \cdot \Phi. \tag{18}$$

For $a = 0$ diversification completely depends on the consistency κ, while for $a = 1$ only the most preferred asset is held in the portfolio (no diversification). On the other hand, when $0 < b < 1$, the algorithm is less inclined to diversify even when consistency is low, and when $b > 1$, algorithm is more inclined to diversify even when consistency is high.

IV. Results

We test the proposed method on a set of 203 stocks that were contiguously included in S&P 500 index from Jan 1st, 1980 until Dec 31st, 2003, which includes 6261 trading days. From 203 stocks a total of 20503 pairs were probed for statistical arbitrage at each time step. Transaction costs of 0.10% were also included in the analysis. The summary of results for various parameters is shown in Table I. The annual Sharpe ratio [11] is defined as: $S = \frac{\mu_r}{\sigma_r}$ — the ratio between annual mean returns and volatilities of the considered portfolio. For each individual trade we analyze the profit by evaluating the amount of gain, loss and net profit, as well as gain/loss ratio. We measure the algorithm accuracy as the ratio of trades resulting in gain to the total number of trades, and we calculate the average turnover ratio as the average percetnage of the portfolio which needs to be rebalanced at each point. Finally, we include the transaction costs and recalculate the portfolio gains. The highest profits and Sharpe ratios have been achieved when using $\alpha = 0$. The naive "Buy & Hold" algorithm was used as a baseline algorithm for comparison. "Buy & Hold" simply holds equal fractions of all available stocks from the beggining to the end. Comparison of methods is shown in Fig. 3.

These results indicate that the proposed method does indeed yield portfolios which are able to perform multi-criteria statistical arbitrage on a large set of assets. In addition, we report that the method adapts to the inconsistence of preferences by picking variable number of assets into the portfolio. This speaks to the resilience of the algorithm to various market conditions, also demonstrated by the obtained portfolio performance. Another interesting finding is the fact that the average gain is much higher than the average loss, meaning that the algorithm errors cost less than the gains obtained when the algorithm is correct. The method is shown to produce rational turnover ratios, and the fact that the obtained Sharpe ratios remain high despite transaction costs additionally affirms the validity of

TABLE I
Results for $T = 60, \alpha = 0$.

Parameter:							
a	0.0			0.5			1.0
b	0.5	1.0	2.0	0.5	1.0	2.0	/
Annual return	0.95339	0.88967	0.84463	0.98336	0.95663	0.89704	1.00223
Annual volatility	0.77042	0.76595	0.74077	0.77905	0.77054	0.76660	0.78363
Annual Sharpe ratio	1.23750	1.16152	1.14020	1.26225	1.24150	1.17015	1.27896
Profit:							
gain	89.27624	88.89440	88.32548	89.58775	89.29840	89.04414	89.55020
loss	-58.37779	-59.03715	-58.41396	-58.24852	-58.32385	-59.02220	-58.05316
net profit	30.89846	29.85725	29.91152	31.33923	30.97456	30.02195	31.49704
positive to negative ratio	1.52928	1.50574	1.51206	1.53803	1.53108	1.50866	1.54256
Average accuracy	0.36485	0.39276	0.43413	0.34902	0.36458	0.39145	0.33241
Average turnover ratio	0.59976	0.64224	0.73597	0.57585	0.59947	0.64089	0.55112
Net profit w/ 0.1% transaction cost	23.46019	21.89215	20.78402	24.19757	23.53996	22.07361	24.66204

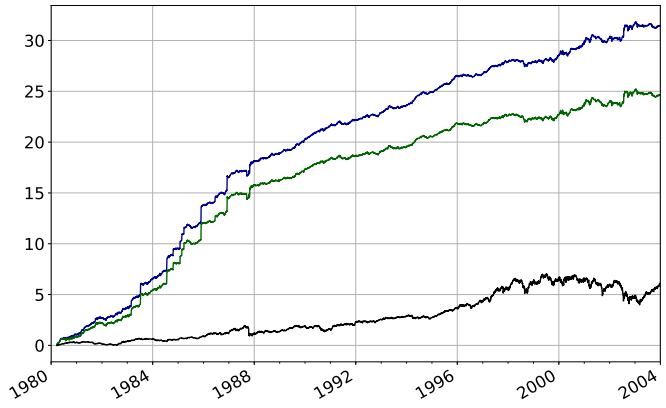

Fig. 3. Performance of the portfolios constructed on 203 stocks from the S&P 500 index during 6261 working days, using the proposed algorithm without transaction costs (blue), proposed algorithm with 0.1% transaction costs (green), and the buy & hold strategy (black).

the approach. These findings suggest that the proposed method produces consistent returns and may be feasible in a live market setting.

V. Conclusion

In this paper we present a method for portfolio optimization based on pairwise statistical arbitrage principles. The algorithm works on pairs of assets, looking for those deviations which are uncommon, and constructs a preference flow graph in order to select the most preferred assets to be included in the portfolio. The method has been tested on a contiguous subset of 203 shares from the S&P 500 market index, from 1980 to 2004. The results suggest that the proposed method yields portfolios with superb market performance, as indicated by the high Sharpe and low turnover ratios. This demonstrates the applicability of the propsed method for portfolio optimization on large sets of financial assets.

Acknowledgment

This work has been supported in part by the Croatian Science Foundation under the project 5349.

References

[1] C. Yingsaeree, G. Nuti, and P. Treleaven, "Computational Finance," *Computer*, vol. 43, no. 12, pp. 36–43, dec 2010.

[2] D. Johnston and P. Djurić, "The Science Behind Risk Management," *IEEE Signal Processing Magazine*, vol. 28, no. 5, pp. 26–36, sep 2011.

[3] X.-P. S. Zhang and F. Wang, "Signal Processing for Finance, Economics, and Marketing: Concepts, framework, and big data applications," *IEEE Signal Processing Magazine*, vol. 34, no. 3, pp. 14–35, may 2017.

[4] W. Sharpe, "Asset Allocation: Management Style and Performance Measurement," *Journal of Portfolio Management*, vol. 18, pp. 7–19, 1992.

[5] R. Cont, "Statistical Modeling of High-Frequency Financial Data," *IEEE Signal Processing Magazine*, vol. 28, no. 5, pp. 16–25, sep 2011.

[6] O. Bondarenko, S. A., and V. R., "Statistical Arbitrage and Securities Prices," *Review of Financial Studies*, vol. 16, no. 3, pp. 875–919, jul 2003.

[7] M. Avellaneda and J.-H. Lee, "Statistical arbitrage in the US equities market," *Quantitative Finance*, vol. 10, no. 7, pp. 761–782, aug 2010.

[8] A. Pole, *Statistical arbitrage : algorithmic trading insights and techniques*. J. Wiley & Sons, 2007.

[9] C. Zopounidis and M. Doumpos, "Multi-criteria decision aid in financial decision making: methodologies and literature review," *Journal of Multi-Criteria Decision Analysis*, vol. 11, no. 4-5, pp. 167–186, jul 2002.

[10] L. Čaklović, "Measure of Inconsistency for the Potential Method," in *9th International Conference on Modeling Decisions for Artificial Intelligence*. Springer, Berlin, Heidelberg, 2012, pp. 102–114.

[11] W. F. Sharpe, "The Sharpe Ratio," *The Journal of Portfolio Management*, vol. 21, no. 1, pp. 49–58, jan 1994.