

# Portfolio Optimization Using Preference Relation Based on Statistical Arbitrage

Lovre Mrčela et al.

**Abstract**— We propose a new algorithm for portfolio optimization based on statistical arbitrage, using potential method to obtain the most preferred assets. A graph that represents preference flow among financial assets (i.e. if an edge exists going from asset  $A$  to asset  $B$ , then  $A$  is preferred over  $B$ ) is constructed at each time step, using a statistical arbitrage algorithm. Preference relation among the assets is imposed by the preference flow. Preference for each particular asset is then calculated using the potential method[1], from which the most preferred assets are selected into the portfolio for each time step.

Method has been tested on contiguous subset of shares from S&P 500 index. Trading cost of 0.1% is also included. Portfolios obtained by the algorithm yield annualized Sharpe ratios of over 1.0. Algorithm performs better when provided with a larger number of assets, as number pairs considered increases, providing more insight into the market behavior.

## I. INTRODUCTION

Classical statistical arbitrage methods take into account a pair of assets whose prices behave similarly during certain period of time. Similarity is measured by cointegration, correlation, or some other measure in order to find a moment in time when those assets' prices exceed what was statistically determined as highly confident range. When such opportunities present themselves, we can take advantage by predicting that prices will return once again to the confident range in the next time step, and do the trading in accordance with this prediction.

In this paper, we propose a new method based on those predictions that are obtained by the statistical arbitrage method, using statistical measures as a proxy for describing the preference relations between pairs of assets. Next, a graph is formed based on those relations, so that mutual interaction of assets might be analyzed. This graph imposes a preference relation among the assets that are included in it. Finally, assets are sorted by preference and included into the portfolio. The idea of this method is to create a generalization of statistical arbitrage methods that is more robust and performs better when working with a larger number of assets by trying to take into account mutual interaction of assets.

## II. CONCEPTS AND METHODS

Following are descriptions of the key components of the algorithm: outline of the preference relation and utility function, graph of preference flow, and potential method.

### A. Preference relation and utility function

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

- *irreflexive*: every entity  $x$  is not preferable over itself,
- *asymmetrical*: if  $x$  is preferable over some  $y$ , then  $y$  is not preferable over  $x$ ,
- *transitive*: if  $x$  is preferable over  $y$ , and  $y$  is preferable over  $z$ , then  $x$  is also preferable over  $z$ ,
- *transitive in incomparability* (noting that  $x$  and  $y$  may be *incomparable*, i.e. neither  $x$  is preferable over  $y$ , nor  $y$  is preferable 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 we describe relationships among the assets. Determining that some asset is preferred over another is easier than assigning a preference rank to each asset individually, especially for a larger number of assets. However, the latter is more useful for decision making, and therefore it is desirable to find a way of sorting assets in the order of preference.

Utility function  $U: \Omega \rightarrow \mathbb{R}$  is mapping from entities to real numbers, in such way that order of the mapping corresponds to the preference order of the entities, i.e.  $\forall x, y \in \Omega, U(x) > U(y) \Rightarrow x \succ y$ . One such mapping is obtained by using potential method that is described later in this paper. In addition to ordering of the entities, utility function also provides the intensity of preference for a particular entity, hence it is more informative when it comes to the decision making.

### B. Graph of preference flow

Graph of preference flow is a weighted directed graph without multiple edges and loops, whose nodes represent entities, edges represent preference for one entity over another, and edge weights correspond to the strength of the preferences. If an edge between nodes is missing, it is considered that neither entity is preferable over another (incomparability of entities). The graph as a whole describes preference flow among the entities. An example of graph is shown on Fig. 1.

Construction of the graph is based on a statistical arbitrage method. An edge going from node  $i$  to node  $j$  with weight  $w_{i,j}$  will be present in the graph if and only if assets represented by nodes  $i$  and  $j$  have demonstrated similar behavior during the past period, but have suddenly

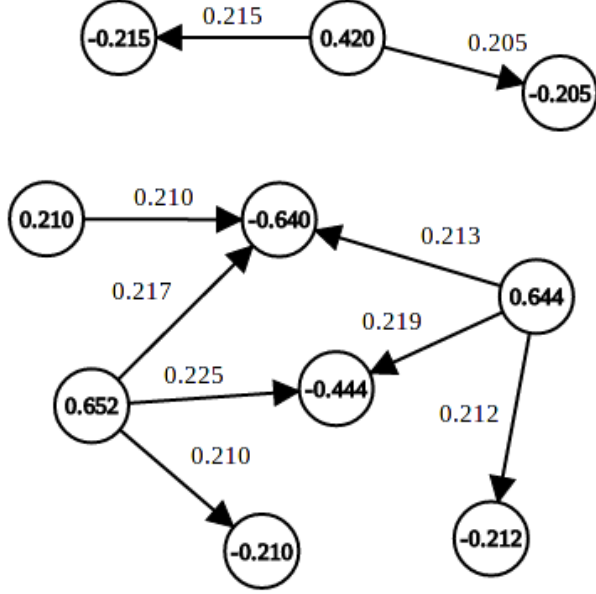


Fig. 1. (TODO: put image with weights on edges) An example of a graph of preference flow. Asset number and calculated preference are inscribed in each node, and preference flows are shown on edges.

diverged at the moment, as determined per statistical measures. 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 among entities that are presented by the graph, in the way that an edge that goes from node  $A$  to node  $B$  means that  $A$  is preferred over  $B$ . It is the case that neither node is in relation with itself (irreflexivity), and that no multiple connections are allowed between 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. 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 devise a consistency measure that describes similarity between the original graph and its nearest consistent reconstruction, and use it as an additional parameter in decision making.

### C. Potential method

From previously obtained 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 preferable assets, or obtain the measure of preference for individual assets. To calculate preferences for each node in the graph, we use the potential method[1]. The potential of a node corresponds to difference in amount of flow going in and out of 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 (direction doesn't matter), thus forming a complete graph  $\mathcal{G}$  with  $\binom{N}{2}$  edges.

Let  $\mathbf{B}$  be the  $E \times N$  incidence matrix of  $\mathcal{G}$ . Let  $\mathbf{f}$  be  $E \times 1$  vector that contains edge weights (i.e. preference flows). Order of the edges must be the same as order of the edges in  $\mathbf{B}$ . As mentioned before, in place of missing edges we simply put 0. Let  $\phi$  be  $N \times 1$  vector that contains potentials of each node, in order that is the same as order of the nodes in  $\mathbf{B}$ .

Now, if  $\mathcal{G}$  was consistent, then  $\mathbf{B}$ ,  $\phi$ , and  $\mathbf{f}$  would satisfy the equation

$$\mathbf{B}\phi = \mathbf{f}. \quad (1)$$

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

$$\begin{aligned} \phi^* &= \arg \min_{\phi} \left\{ \|\mathbf{B}\phi - \mathbf{f}\|^2 \right\} \\ &\Downarrow \\ \frac{\partial \|\mathbf{B}\phi^* - \mathbf{f}\|^2}{\partial \phi^*} &= \mathbf{0}. \end{aligned} \quad (2)$$

Solving (2) via commonly used techniques of matrix calculus brings us to the following equation:

$$\mathbf{B}^\top \mathbf{B}\phi^* = \mathbf{B}^\top \mathbf{f}. \quad (3)$$

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

$$\mathbf{j}^\top \phi^* = 0 \quad (4)$$

where  $\mathbf{j}$  is vector of ones with same dimension as  $\phi^*$ . This ensures an 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:

$$\begin{aligned} \mathbf{B}^\top \mathbf{B}\phi^* + \mathbf{J}\phi^* &= \mathbf{B}^\top \mathbf{f} \\ [\mathbf{B}^\top \mathbf{B} + \mathbf{J}]\phi^* &= \mathbf{B}^\top \mathbf{f}, \end{aligned} \quad (5)$$

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

$$\phi^* = [\mathbf{B}^\top \mathbf{B} + \mathbf{J}]^{-1} \mathbf{B}^\top \mathbf{f}. \quad (6)$$

Furthermore, the term  $[\mathbf{B}^\top \mathbf{B} + \mathbf{J}]^{-1}$  in (6) can be simplified to  $\frac{1}{N}\mathbf{I}$  due to  $\mathbf{B}^\top \mathbf{B}$  being the Laplace matrix of a complete graph; thus, we can simplify (6) some more:

$$\phi^* = \frac{1}{N} \mathbf{B}^\top \mathbf{f}, \quad (7)$$

to get as computationally optimal expression as possible.

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

$$\mathbf{f}^* = \mathbf{B}\phi^*. \quad (8)$$

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

$$\kappa = \frac{\|\mathbf{f}^*\|}{\|\mathbf{f}\|}. \quad (9)$$

Equation (9) represents the cosine of the angle between  $\mathbf{f}$  and  $\mathbf{f}^*$  in the column space of matrix  $\mathbf{B}$ . Consistency measure  $\kappa$  tells us how consistent graph  $\mathcal{G}$  was, compared to the  $\mathcal{G}^*$ .  $\kappa$  ranges from 0 to 1, with 0 meaning full inconsistency (virtually unreachable), and 1 meaning full consistency.

### III. ALGORITHM

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

$$b_i^{(t)} = \log(a_i^{(t)}) \quad (10)$$

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

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

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

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

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

#### A. Creating the graph of preference flow

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:

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

i.e. current log price difference is at least  $\alpha$  deviations distant from mean value of the past time window. An illustration is shown on 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:

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

Thus it is possible to create a graph of preference flow for each time step  $t \in [T, T+1, \dots, D-1]$ . 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  $\alpha$  yields denser graphs, and setting  $\alpha = 0$  always yields complete graphs.

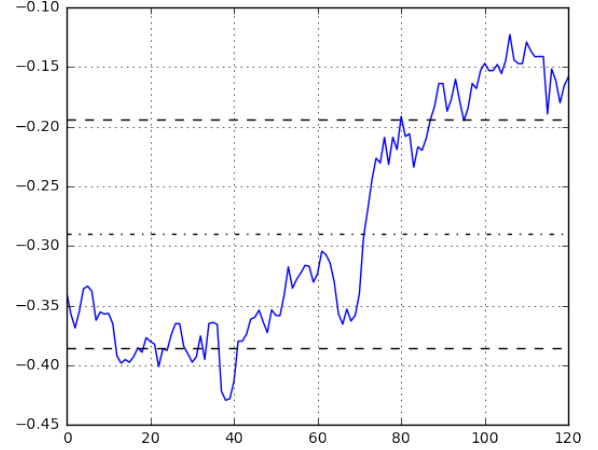


Fig. 2. Log price difference between a pair of assets  $(i, j)$  during a period of  $T+1$  time steps, where  $T = 120$ . Dashed and dotted line represents mean value, and region between two dashed lines represents  $\alpha$  standard deviation range from mean value, here  $\alpha = 1$ ; both are calculated in the first  $T$  time steps. During the time step  $T+1$ , log price difference goes over  $\alpha$  standard deviations above mean value of past period of size  $T$ . This would mean that two of the assets  $i$  and  $j$  would be added to the graph of preference flow at the time step  $T+1$ . Weight  $w_{i,j}$  describes current deviation from mean value.

#### B. Choosing assets from graph

We obtain preference for each 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  $\phi^{(t)} = [\phi_1^{(t)} \ \phi_2^{(t)} \ \dots \ \phi_N^{(t)}]$  denote vector of preferences of assets at time step  $t$  and  $\phi_i^{(t)}$  denote the preference for asset  $i$  at time step  $t$ . When picking the assets for the portfolio we take into consideration the consistency measure  $\kappa$  as well. Lower values of  $\kappa$  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 do trading with 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  $\kappa$ .

Depending on the nature of assets we may tune the consistency measure  $\kappa$  to be more or less inclined to diversification by transforming it to  $\kappa'$ :

$$\kappa' = a + (1 - a)\kappa^b, \quad (16)$$

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

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

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

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

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

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

#### IV. RESULTS

Results were obtained by testing on a set of 203 stocks that were contiguously included in S&P 500 index from Jan 1st, 1980 thru Dec 31st, 2003, which includes 6261 trading days. A total of 20503 pairs of assets were probed for statistical arbitrage at each time step. Summary of results for various parameters is shown in the table IV. Best profit and Sharpe ratio has been achieved when using  $\alpha = 0$ .

[bt] tableResults 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	/
Average return (year)	0.95339	0.88967	0.84463	0.98336	0.95663	0.89704	1.00223
Volatility (year)	0.77042	0.76595	0.74077	0.77905	0.77054	0.76660	0.78363
Sharpe ratio (year)	1.23750	1.16152	1.14020	1.26225	1.24150	1.17015	1.27896
Profit:							
positive only	89.27624	88.89440	88.32548	89.58775	89.29840	89.04414	89.55020
negative only	-58.37779	-59.03715	-58.41396	-58.24852	-58.32385	-59.02220	-58.05316
total	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
Actual profit, if transaction cost were 0.1%	23.46019	21.89215	20.78402	24.19757	23.53996	22.07361	24.66204

#### V. CONCLUSIONS

Algorithm works on pairs of assets, looking for those deviations which are uncommon, so generally it is expected to perform better where there is larger number of assets as more deviations will be discovered. It adapts to the inconsistency of preferences by picking variable number of assets into the portfolio.

#### REFERENCES

- [1] L. Čaklović, Decision Making by Potential Method