Portfolio Optimization Using Preference Flow Based on Statistical Arbitrage

Lovre Mrčela et al.

***Abstract*— A new algorithm for portfolio optimization is proposed which is based on statistical arbitrage, with potential method used 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 the modified version of statistical arbitrage. Then, the preference of each asset is 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 dataset (TODO: which dataset), by simulating the portfolio obtained by the algorithm, including the trading cost of 0.1%. Sharpe ratios over 1.0 were observed.

I. INTRODUCTION

Classical statistical arbitrage methods take into account some pair of assets whose prices behave similarly during certain period of time measured by cointegration, correlation, or some other measure of similarity; then try to find a moment in time when those assets' prices go out of what was statistically considered as 'normal' range. When such opportunities show up, we can take advantage of them by predicting, with a certain statistical confidence, that they will return to the 'normal' range once again in the next time step, and do the trading accordingly to this prediction.

In this paper, we propose a new method that uses predictions obtained by statistical arbitrage method, with confidence as a proxy for describing the preference relations between pairs of assets. Next, a graph is formed based on those relations, so that in-depth analysis of assets' interaction with each other might be performed. Finally, assets are put into order by preference, and picked into the portfolio. The idea of this method is to create a generalization of statistical arbitrage methods that will be more robust and perform better when working on a larger number of assets.

II. CONCEPTS AND METHODS

Following are descriptions of the key components in the algorithm: graph of preference flow, and choosing assets for the portfolio...

A. Preference relations 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 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 preferrable over itself,
- *asymmetrical*: if x is preferrable over some y, then y is not preferrable over x,
- *transitive*: if x is preferrable over y, and y is preferrable over z, then x is also preferrable over z,
- *transitive in incomparability* (noting that x and y may be *incomparable*, i.e. neither x is preferrable over y, nor y is preferrable 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 relationships among the assets. Determining that some asset is preferred over another is an easier task than assigning a preference rank to each asset individually, especially when the number of assets becomes large. 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 $U\colon \Omega \to \mathbb{R}$ 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 a magnitude of preference for particular entity, and hence it is more informative when it comes to the decision making.

B. Graph of preference flow

Graph of preference flow is a simple weighted directed graph whose nodes represent entities, edges represent preference for one entity over another, and edge weights correspond to the strength of the preferences. In case of missing edge between two nodes, it is considered that neither entity is preferable over another (incomparability). The graph as a whole describes preference flow among the entities. An example of graph is shown on Fig. 1.

Connections in this graph impose a preference relation among entities that are in the graph, in a 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 (irrelexivity), 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 arbitrary instance of the graph. This imposed preference relation should preferably be consistent, but when it comes to larger number of entities, it may become infeasible to

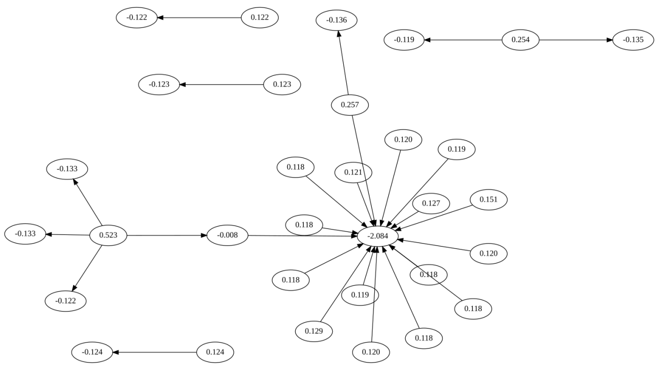

Fig. 1. (TODO: put image with weights on edges) An example of a graph of preference flow. Asset number and calculated preference are inscribed in each node, and preference flows are shown on edges.

construct a graph with such qualities. Instead of aiming at consistency of relations, we devise a consistency measure and use it as an additional parameter in decision making.

Construction of graph is based on a statistical arbitrage method. In short, an edge going from node i to node j with weight $w_{i,j}$ will be present in graph if and only if following happens: assets represented by nodes i and j have demonstrated similar behavior during the past period of time, but have suddenly diverged at the moment, as determined per statistical measures. Weight $w_{i,j}$ corresponds to the magnitude of this divergence. A detailed description of procedure is given later in III-A.

The main idea of preference flow is that preference of a particular node should depend on amount of preference flowing in and out of it. This allows for some kind of generalized statistical arbitrage over multiple assets at a time.

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 preferrable 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.

A concise summary of the method is as follows:

1) 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.
2) If $\mathcal{G}$ is not complete, we complete it by adding edges to it with weight 0 (direction doesn't matter). Thus, from now on $\mathcal{G}$ is a complete graph with $\binom{N}{2}$ edges.
3) Let $\boldsymbol{B}$ be the $E \times N$ incidence matrix of $\mathcal{G}$.
4) Let $\boldsymbol{f}$ 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 $\boldsymbol{B}$. As mentioned before, in place of missing edges we simply put 0.
5) Let $\boldsymbol{\phi}$ be $N \times 1$ vector that contains potentials of each node, in order that is the same as order of the nodes in $\boldsymbol{B}$.
6) Now, if $\mathcal{G}$ was consistent, then $\boldsymbol{B}$, $\boldsymbol{\phi}$, and $\boldsymbol{f}$ would satisfy the equation
$$\boldsymbol{B}\boldsymbol{\phi} = \boldsymbol{f}. \tag{1}$$
Equation (1) simply means that the difference between potential of any two nodes should result 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 $\boldsymbol{\phi}^*$ that minimizes the square error:
$$\boldsymbol{\phi}^* = \arg\min_{\boldsymbol{\phi}} \left\{ \left\| \boldsymbol{B}\boldsymbol{\phi} - \boldsymbol{f} \right\|^2 \right\}$$
$$\Downarrow$$
$$\frac{\partial \left\| \boldsymbol{B}\boldsymbol{\phi}^* - \boldsymbol{f} \right\|^2}{\partial \boldsymbol{\phi}^*} = \boldsymbol{0}. \tag{2}$$
Solving (2) via commonly used techniques of matrix calculus brings us to the following equation:
$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\phi}^* = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}. \tag{3}$$
Equation (3) determines $\boldsymbol{\phi}^*$ up to a constant (i.e. solution has one degree of freedom), so the following constraint is also included:
$$\boldsymbol{j}^{\mathrm{T}}\boldsymbol{\phi}^* = 0 \tag{4}$$
where $\boldsymbol{j}$ is vector of ones with same dimension as $\boldsymbol{\phi}^*$. This ensures an unique solution for which total amounts of positive and negative potential will be equal.
7) Joining the previous two equations together by adding (4) to each row in (3) results in:
$$\begin{aligned} \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\phi}^* + \boldsymbol{J}\boldsymbol{\phi}^* &= \boldsymbol{B}^{\mathrm{T}}\boldsymbol{f} \\ \left[\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{J}\right]\boldsymbol{\phi}^* &= \boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}, \end{aligned} \tag{5}$$
where $\boldsymbol{J}$ is a matrix of ones with same dimension as $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$. Finally, solving (5) for $\boldsymbol{\phi}^*$ gives us:
$$\boldsymbol{\phi}^* = \left[\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{J}\right]^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}. \tag{6}$$
8) Furthermore, the term $\left[\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{J}\right]^{-1}$ in (6) can be simplified to $\frac{1}{N}\boldsymbol{I}$ due to $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ being the Laplace matrix of a complete graph; thus, we can simplify (6) some more:
$$\boldsymbol{\phi}^* = \frac{1}{N}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}, \tag{7}$$
to get as computationally optimal expression as possible.
9) Afterwards, we can calculate the 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 even contain

some new and/or lose some old edges. In addition, $\boldsymbol{B}$, $\boldsymbol{\phi}^*$, and $\boldsymbol{f}^*$ now describe a consistent graph $\mathcal{G}^*$. This allows now 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 $\boldsymbol{B}$. Consistency measure κ tells us how consistent graph $\mathcal{G}$ was, compared to the $\mathcal{G}^*$. κ ranges from 0 to 1, with 0 meaning full inconsistency (virtually unreachable), and 1 meaning full consistency.

III. ALGORITHM

Parameters of the algorithm are:

- T - length of the past time window,
- α - the deviation threshold,
- β - asset selection threshold.

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 $i \in [1\,..\,N]$ and $t \in [0\,..\,D-1]$. The log prices $b_i^{(t)}$, and log price differences $c_{i,j}^{(t)}$ between assets i and j are obtained as follows:

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

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

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:

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

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

Note that in summation used in (12, 13) 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) for which holds:

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

i.e. current log price difference is at least α deviations distant from mean value of the past time window. An example 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:

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

Thus it is possible to create a graph of preference flow for each time step $t \in [T\,..\,D-1]$. 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.

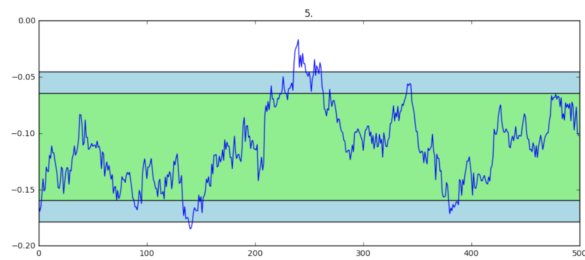

Fig. 2. (TODO: put a better image, this is just for a placeholder) Log price difference of a pair of assets during period of $T+1$ time steps. During the last observed timestep, it goes (TODO: broj) times standard deviations away from mean value of past time window of size T. This would be measure of preference flow from asset with higher price to the asset with lower price.

B. Choosing assets from graph

We obtain preference of each asset via the potential method, as described earlier in II-C. By obtaining the measure of preference of each asset it is now 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 $\boldsymbol{\phi}^{(t)}$ denote vector of preferences of assets at time step t and $\phi_i^{(t)} \in \boldsymbol{\phi}^{(t)}$ denote the preference of asset i at time step t. When picking the assets for the portfolio we take into consideration the consistency measure κ as well. Lower values of κ suggest that we should hedge our portfolio by including some more assets in the order of preference, while higher values suggest that it is safer to do trading with smaller number of assets. So, the bound on the assets which will be taken into portfolio is proportional to the consinstency measure κ. Depending on the nature of assets we may tune the consistency measure κ to be more or less inclined to hedging by transforming it to

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

where $a \in [0,1], b \in \mathbb{R}^+$. For default values of $a=0, b=1$, κ' equals κ. In our dataset, values $a=0.5$ and $b=0.5$ achieved better performances than the default ones.

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

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

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

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

IV. RESULTS

Results were obtained by testing on dataset (TODO: dataset). Short selling was disabled.

T=90

α	β	Sharpe ratio	turnover rate	profit
3.0	1.0	1.12033	**0.64613**	18.63180
	0.9	1.14333	0.68407	19.03378
	0.8	**1.15759**	0.71548	**19.35692**
	0.7	1.13153	0.75715	19.00914
	0.6	1.06598	0.79802	17.70338
	0.5	1.03619	0.85536	17.16405
3.25	1.0	1.04127	0.68351	17.27340
	0.9	1.07932	0.71015	17.90707
	0.8	1.12707	0.75099	18.72864
	0.7	1.10705	0.79748	18.42081
	0.6	1.05853	0.84321	17.51641
	0.5	1.01289	0.89148	16.76236
3.5	1.0	0.99809	0.72792	16.40651
	0.9	1.02397	0.76663	16.88082
	0.8	1.03420	0.81126	17.09970
	0.7	1.05821	0.86305	17.48784
	0.6	1.01590	0.91392	16.89037
	0.5	0.98479	0.96461	16.27100
3.75	1.0	0.98829	0.80549	15.84890
	0.9	1.02361	0.84603	16.40503
	0.8	1.04266	0.89192	16.74853
	0.7	1.03920	0.95609	16.73169
	0.6	1.05219	1.00587	17.00883
	0.5	1.03340	1.06956	16.67687
4.0	1.0	1.00574	0.91970	15.02475
	0.9	1.02546	0.95710	15.37962
	0.8	1.04724	0.99747	15.74329
	0.7	1.04383	1.03972	15.76506
	0.6	1.01710	1.08639	15.39334
	0.5	1.01925	1.15608	15.43147

Fig. 3. Results. TODO: comment.

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 inconsistence of preferences by picking variable number of assets into the portfolio.

References

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

T=120

α	β	Sharpe ratio	turnover rate	profit
3.0	1.0	1.02488	**0.57236**	17.63623
	0.9	0.99552	0.60648	17.16067
	0.8	1.00498	0.64939	17.24051
	0.7	1.01485	0.68910	17.43257
	0.6	1.04164	0.72308	17.91063
	0.5	**1.08388**	0.78830	**18.33920**
3.25	1.0	0.95261	0.60842	15.76037
	0.9	0.97329	0.63420	16.11968
	0.8	0.98852	0.67298	16.35240
	0.7	0.97404	0.72470	16.12448
	0.6	0.94128	0.77388	15.67659
	0.5	0.97199	0.81790	15.97750
3.5	1.0	0.89277	0.66365	14.63103
	0.9	0.93685	0.69743	15.37556
	0.8	0.94367	0.73989	15.52817
	0.7	0.95012	0.79145	15.73655
	0.6	0.96529	0.83751	15.97783
	0.5	0.99559	0.88986	16.08797
3.75	1.0	0.90895	0.74222	14.27841
	0.9	0.96107	0.77696	15.10384
	0.8	0.98794	0.82238	15.53080
	0.7	1.00128	0.87246	15.92569
	0.6	0.98468	0.92541	15.67132
	0.5	1.01009	0.98699	16.09083
4.0	1.0	0.99691	0.83501	14.67240
	0.9	1.01115	0.87406	14.89908
	0.8	1.06557	0.91977	15.71279
	0.7	1.06097	0.95934	15.76776
	0.6	1.03053	0.99846	15.34007
	0.5	1.04213	1.05134	15.47247

Fig. 4. Results. TODO: comment.