A Combinatorial Method to Producing Portfolios

Samuel Koeut

November 2016

Abstract

There is a lack of research on portfolios and combinatorial methods in finance. In this paper, we outline a new method for producing long-term portfolios of stock using a combinatorial approach. A retrospective data analysis shows that this method produces profitable long-term portfolios.

1 Introduction

While methods of pricing individual stock are thoroughly researched and well developed, there is far less research on the topic of portfolios. Furthermore, there is a lack of combinatorial methods present in finance. This paper aims to resolve both by providing a combinatorial approach to produce portfolios of stock.

The method is based on the idea of 'not putting all your eggs in one basket.' Our method works by creating a portfolio of stock that are mutually unrelated in their growth from an arbitrary set of stock. This is opposed to the normal investigation of stock individually. We study a group of stock as a whole and cut out bits that correlate too strongly to each other. This significantly reduces the amount of time that it takes to construct a portfolio.

Our method relies primarily on graph theory and a defined relation to create portfolios.

2 Preliminaries

An arbitrary set of stock, each with a trading history of at least length t, in days, will be denoted by $S_t = \{s_{1,t}, s_{2,t}...s_{n,t}\}$. A trade vector T_i is vector such that the j^{th} of T_i is equal to the percent change of the stock $s_{i,t}$ from the j^{th} day to the $j+1^{th}$ day.

A graph G of order n is an ordered pair (V, E), where V is a set of vertices $\{v_1, v_2, ..., v_n\}$ and E is a set of edges $\{v_iv_j, v_kv_l...\}$. A completed graph K is a graph where every pair of vertices are connected by an edge. A clique C is a completed subgraph of G. The complement G' of a graph G is a a graph with vertex set V and edge set E', where $e = v_iv_j \in E'$ iff $e \notin E$.

3 The Method

Let S_t be a set of stocks with elements $s_{1,t},, s_{2,t}, ..., s_{n,t}$ and k be a real number such that $-1 < k < 1$.

Let G be a graph such that each vertex represents a stock in S_t and an edge exists between v_i, v_j if and only if the correlation coefficient between T_i, T_j is greater than k.

Our portfolio is then the maximal clique of the graph G'. In the case of a tie, one is picked arbitrarily.

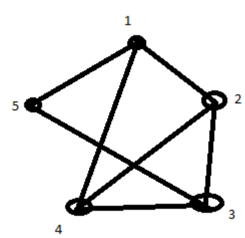
4 An Example

While this method generally requires a larger set of stocks to produce a viable portfolio, we use a small example to outline the method.

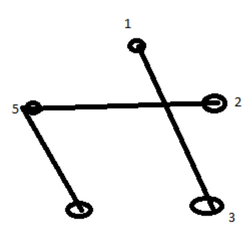
Let $S_10 = \{s_{1,10}, s_{2,10}, s_{3,10}, s_{4,10}, s_{5,10}\}$ be a set of five stock with a correlation coefficient of daily stock returns as a percent of the previous closing price as summarized in the following table.

	$s_{1,10}$	$s_{2,10}$	$s_{3,10}$	$s_{4,10}$	$s_{5,10}$
$s_{1,10}$	1	0.591	−0.939	0.770	0.424
$s_{2,10}$	0.591	1	0.190	0.940	−0.367
$s_{3,10}$	−0.939	0.190	1	0.851	0.326
$s_{4,10}$	0.770	0.940	0.851	1	−0.486
$s_{5,10}$	0.424	−0.367	0.326	−0.486	1

From this we construct the following graph.



We then find the complement.

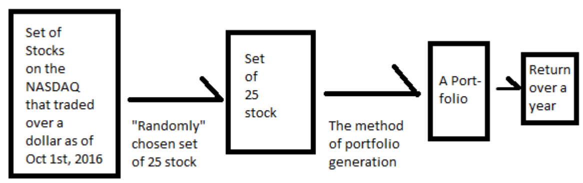


We find a maximal clique in the complement. This clique is our portfolio. Since there are ties, we pick one arbitrarily.

5 Testing Results

In order to test the method for it's viability, we tested seven sets of stock, each set containing 25 stocks randomly chosen from the NASDAQ. Stocks were limited to stocks that traded over a dollar as of October 1st, 2016. Our threshold value was $k = .1$.

These stocks each had ten years of data, from 2006 to 2016. In each set, we used the first nine years of data to check the mutual relatedness and create the portfolios. The last year of data was used to simulate investing in these portfolios. We assume an investment of 1 into each stock.



These sets returned portfolios with increases of -1, 1, 9, 16, 17, 21, and 40 percent over a year. We create a 90% confidence interval to check if the proportion of portfolios produced by this method with a positive return is greater than p=.5.

A resulting confidence interval is (.64,1). Therefore, we are 90% confident that the true proportion of portfolios created by this method with a positive return is in that interval. As such, we conclude that our method is better than picking stock arbitrarily.

6 Conclusion

With our usage of confidence interval, we conclude our method is successful. This research is based on the idea of relatedness being correlation coefficient and a threshold of .1. Interesting items to research in order to refine this method include:

- The size of the clique relative to the set of the input stock sets.
- The effect of changing the threshold on the returns of portfolios.
- Changing the definition of the relatedness.
- Changing the definition of relatedness to lack symmetry.
- Relating the return of the portfolio to the underlying stock set.