Generalized Neutral Portfolio

Sandeep Bhutani
bhutanister@gmail.com

December 10, 2009

Abstract

An asset allocation framework decomposes the universe of asset returns into factors either by Fundamental Analysis, Factor Analysis or Principal Component Analysis. Asset allocation methods then attempt to neutralize the sensitivity of the portfolio to a select factor. However, Factor Analysis or PCA fail to take into account moments higher than the second moment. This paper outlines a method based on Independent Component analysis to effectively neutralize a portfolio to a component. This method is illustrated for building a market neutral portfolio in the Mean variance framework, and can be extended to encompass any factor(s).

1 Introduction

Capital Asset Pricing Model (CAPM) postulates that any stock return can be decomposed into a linear combination of a return component correlated to the Market Portfolio and an idiosyncratic component. This simplified model can be used to design a portfolio which is insensitive to market gyrations, but still has a positive return. A simplified example is the pairs trading method: A pair of highly correlated stocks are chosen and the trader goes long the cheaper stock and short the expensive stock to create a market neutral pair. This pair has zero correlation with the market. The method of pairs trading can be generalized to create a linear combination of arbitrary stocks which gives a net market exposure of zero.

The pairs trading method can be generalized by decomposing a stock specific return into various factors. The factors could be fundamentals based, which include for instance, market capitalization, industry/sector, rankings, analyst ratings etc. Or the factors can be determined by statistical techniques like Factor Analysis, Principal Component Analysis (PCA) to determine the dominant return factors. Then a portfolio manager can neutralize his/her portfolio to a chosen factor based on his/her beliefs or alpha model. This is achieved by the Portfolio optimization technique which minimizes the variance of the returns of the portfolio for a desired return, under the budget and factor neutrality constraints.

The method of using PCA with variance minimization considers only the first two moments, mean and variance. Factor analysis makes a Gaussian assumption which is routinely violated by the stock returns. The market returns are negatively skewed and fat-tailed. We saw an evidence of the higher moments in the past year (2008) when the market-neutral hedge funds suffered with the drop in the market.

To overcome the Gaussian distribution assumption, this paper proposes to modify the method of extracting factors using Independent Component Analysis to decompose the stock returns into factors. The resulting factors would be from independent distributions and not just uncorrelated. Then using Mutual Information as a metric of similarity the Independent factor corresponding to Market would be identified and used as a constraint in portfolio optimization. The resulting return profile is shown to be independent of the market by using the Mutual Information metric.

The next section describes existing work in explaining stock returns by means of Independent Component Analysis. The followng section describes the data used for the study, which is followed by a simplified mean variance optimization solution. We then describe the PCA and ICA formulations which are followed by the evidence based on our analysis of the stock returns data. The analysis section demonstrates the ineffectiveness of neutralizing the portfolio to the market using only the second moments. Additionally by using ICA it demonstrates the market neutrality of the resulting portfolio weights.

2 Existing Work

Back and Weigend in [1] made first use of Independent Component Analysis in analysing stock returns. They applied it to the daily returns of the 28 of the largest Japanese stocks and used ICA factors as explanatory variables for the stock market. A very similar approach was used by Siu-Ming Cha and Lai-Wan Chan in [2] to analyze the returns of the Hang Sheng market. Their primary purpose too was to explain the underlying data generating process of the stock returns. An innovative use of ICA to analyze hedge fund returns can be found in [4] where Jan uses ICA to explain the factors driving the hedge fund returns of a select group of hedge fund managers and demonstrates superiority over using PCA for attribution analysis. Analysis of S&P500 returns in [5] about the non-normality of stock returns by Weigend et al serves as the motivation for this paper.

3 Data

The data used for this problem was obtained from the Fama French website [3]. The website contains daily returns for a set of ten portfolios which seggregate the US stock universe based on the size of the companies. Each one of these portfolios is treated as one asset for the purpose of this paper. The market returns are determined by the daily returns on the index **SPY** and downloaded from Yahoo. The period of study is from February 2, 1993 to September 30, 2009.

4 A note on Mutual Information metric

As the goal of this project is to demonstrate that the resulting distribution of returns is **independent** of the market returns, we need a metric that will work on distributions. Kolmogorov-Smirnov (KS) test can be used to compare distributions, however it can not determine whether the given distribution is part of the mixture. Kullback Leibler (KL) Divergence would tell us if the two distributions are independent, but the KL divergence can not be used as a metric since it does not obey triangle inequality, which prevents it from being used to compare non-zero KL divergences. There is a method based on KL Divergence, the Mutual Information metric [7] which obeys triangle inequality, and is a true metric. This is the method we use in this paper as this allows us to determine when two distributions are independent and if they are not independent, the one with the larger metric shows greater dependence.

5 Modern Portfolio Theory

Asset allocation in the modern portfolio theory [6] is achieved by determining the weights on each of the assets to maximize the total return while minimizing the risk under a given budget constraint. For a vector of portfolio weights of assets denoted by w, return vector R, asset covariance matrix Σ, the portfolio optimization is defined as a solution to:

$$\min w^T \Sigma w - \lambda R^T w \tag{5.1}$$

$$s\ t : \sum_{i=1}^{m} w_i = 1$$

The covariance matrix is created by $cov(R)$, where R is a matrix of return vectors corresponding to each asset. Additional constraints can be included in (5.1) based on the portfolio design requirements.

6 Principal Component Analysis

For PCA we determine the zero based auto-correlation matrix and then use the *princomp* in MATLAB to determine the components. We form portfolios corresponding to each one of the Principal Components and compare the returns of these portfolios to the Market Portfolio using the Mutual Information metric. The component with the largest Mutual Information (the largest eigenvalue) corresponds to the Market. We modify (5.1) to include this constraint:

$$R^T w_{pca^{(i)}} = 0$$

which indicates that the returns corresponding to this component have to be zero. We then determine the returns of the optimized portfolio and compute the efficient frontier. The set of returns computed are once again compared to the market using the Mutual Information metric. We find the mutual information to be non-zero in this case indicating that the returns are not indepdent of the market.

7 Independent Component Analysis

While there are several algorithms for Independent Component Analysis, we use the FastICA algorithm. We compute the Independent Components and determine the component corresponding to the Market Portfolio based on the Mututal Information metric. This is used to add the constraint to the (5.1):

$$R^T w_{ica^{(i)}} = 0$$

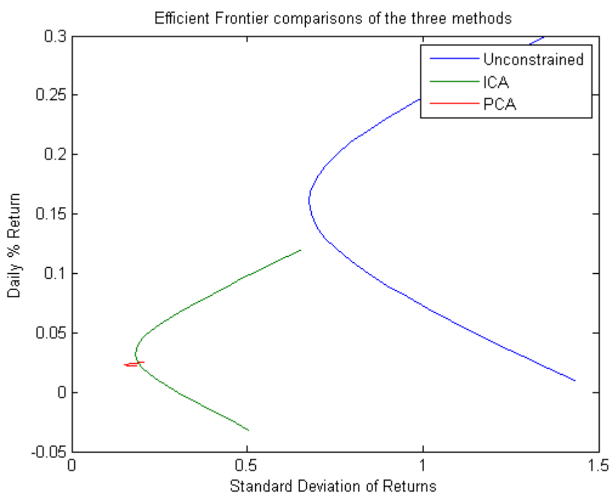
We once again compute the efficient frontier of the new portfolio using ICA. We determine the mutual information of the returns of the ICA portfolio and compare it to the mutual information of the returns of the PCA.

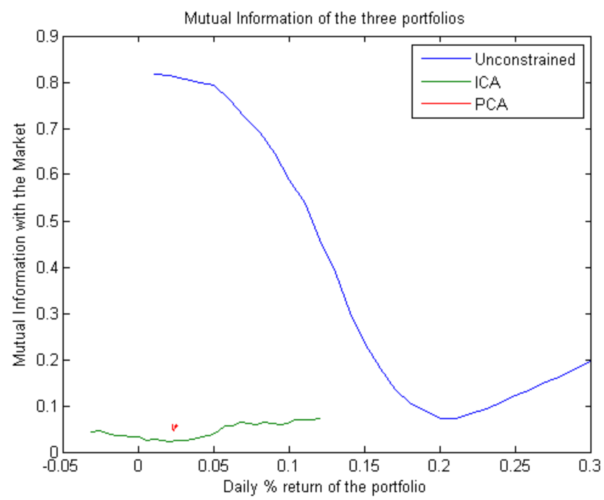
8 Method and Results

This section illustrates the comparison of the efficient frontier obtained by the three methods: Unconstrained Optimization, Optimization with the domiant PCA eigenvector set to zero and Optimization with the dominant ICA component set to zero. The choice of the component to neutralize for both ICA and PCA was made by picking the component which had the highest Mutual Information with the returns to the market. The constraint was implemented by regenerating the returns matrix by multiplying the original returns matrix with a mixing matrix which had the dominant column set to zero for PCA and the dominant row set to zero for ICA.

As expected, the PCA based returns have non-zero mutual information when compared to the market. However, the Mutual Information of ICA constrained portfolio based returns is still non-zero albeit much lower than the PCA based portfolio for the same level of returns. Another observation from the plots is that the possible risk and returns from the ICA constrained portfolio are higher than the possible risk and returns from the PCA constrained portfolio indicating that there is improvement to be gained by using an ICA based method.

The efficient frontier shows that the returns from the unconstrained portfolio are higher than the constrained returns owing the presence of the market which is expected since the ICA and PCA portfolios can't make use of the market. A striking result of the efficient frontier is that the unconstrained results have a minimum with respect to the Mutual information to the market which indicates another possibility of achieving a market neutral portfolio. To test this, we tried optimizing the orignal equation (5.1) with the constraint that the mutual information with respect to the market be zero and that fails to converge in MATLAB.





9 Conclusion

This paper illustrates a method for creating a generalized neutral portfolio to a component based on using Indepdent Component Analysis as a constraint in the Markowitz Portfolio Theory framework. The resulting portfolio returns are evaluated for their mutual information with the market and the results from an ICA based constraint have lower mutual information with the market when compared to a PCA based constraint. Furthermore, ICA based results give access to a higher range of return/risk values than the PCA based methods enabling higher returns from existing portfolios. This kind of portfolio formulation can be extended to include any other factor or a set of factors. An added benefit of this formulation is that when we remove the exposure to market distribution, we greatly reduce the higher moment exposure to our portfolio without directly optimizing for it. This is something that we expect to demonstrate in future work.

References

[1] Andrew D. Back and Andreas S. Weigend. A first application of independent component analysis to extracting structure from stock returns. *Working Paper IS-97-22*, 1997.

[2] Siu-Ming CHA and Lai-Wan CHAN. Applying independent component analysis to factor model in finance. *Intelligent Data Engineering and Automated Learning - IDEAL 2000, Data Mining, Financial Engineering, and Intelligent Agents*, pages 538–544, 2000.

[3] Kenneth R. French and Eugene Fama. Fama french portfolios.

[4] Jan Olszewski. Unraveling hedge fund returns: An introduction to independent component analysis as an analytical tool. *Available at SSRN: http://ssrn.com/abstract=884266*, October 2006.

[5] Andreas S.Weigend and Shanming Shi. Predicting daily probability distributions of s&p500 returns. *Available at http://hdl.handle.net/2451/14307*, Aug 1998.

[6] Wikipedia. Modern portfolio theory. *Available at http://en.wikipedia.org/wiki/Modern_portfolio_theory*, 2009.

[7] Wikipedia. Mutual information. *Available at http://en.wikipedia.org/wiki/Mutual_information*, 2009.