An Application of Cluster Random Subspace Method to Diversification & Portfolio Management

Research Project

Michael Guan
Email: wguan@uwo.ca

Supervisors: Dr. Stephen M. Watt and David Varadi

Supervisor Signatures

Dr. Stephen M. Watt: ______________________

UWO Computer Science Department

David Varadi: ______________________

Vice President of Economic and Strategic Development
Flexible Plan Investments

An Application of Cluster Random Subspace Method to Diversification & Portfolio Management

By
Michael Guan
Submitted to the UWO Department of Computer Science
on March 29th, 2014 in fulfillment for the requirements for
the degree of Bachelor of Science

Abstract

Equity portfolio managers are often required to estimate numerous statistical parameters for modelling large databases of stocks and/or assets. These input assumptions are often estimated with limited historical data which are then used to arrive at efficient portfolios through optimization procedures. Traditional optimization methods are based upon Modern Portfolio Theory, and require both return and variance/covariance inputs to solve for portfolio weights. Applying this framework to large-scale optimization is complicated by the "Curse of Dimensionality" due to the thousands or even hundreds of thousands of separate estimates that need to be derived from a limited dataset. The main purpose of this paper is to apply the principles of machine learning to specifically address the problem of Curse of Dimensionality in the optimization of large portfolios. Specifically, we apply the Random Subspace Method (RSO)- which is a well-accepted machine learning methodology for dealing with high dimensionality. RSO is a stochastic sampling procedure to reduce an initial space into a series of smaller sub-spaces that require fewer estimates. These subspaces are then aggregated to produce a more robust output since it relies upon fewer total parameter estimates. We improve upon the deficiencies of the RSO framework: 1) inefficient or poor diversification 2) vulnerability to universe bias and 3) introducing arbitrary sampling parameters that need to pre-specified that are dependent on both the size and homogeneity or heterogeneity of the universe of assets chosen. We instead introduce a novel RSO variation that uses clustering and also a heuristic to pre-define the number of samples required, that we term the "Cluster Random Subspace Method" CRSO. We compare the effectiveness of RSO and CRSO methodologies versus traditional approaches such as the "Maximum Sharpe Ratio" (MSR) or tangency

portfolio through large-scale optimization on a variety of asset class data. We show that both CRSO and RSO are superior to traditional methods, but CRSO is the best performer overall on a wide variety of metrics across a number of different universes. This makes CRSO a practical alternative to MSR for large-scale optimization problems.

ACKNOWLEDGEMENTS:

First and foremost, I would like to thank both of my supervisors for providing support and encouragements on the subject area I am pursuing. Their feedbacks and insights have been indispensable to my research and thought process.

Secondly, I would like to thank both Mike Katchabaw and John De Heuss for allowing me the last minute opportunity to take the course. Without them, this paper wouldn't be possible.

Lastly, I would like to thank Adam Butler from BPG & Associates with their generosity of sharing data and insight in to the asset allocation process.

Table of Contents

Abstract

Acknowledgement

Introduction

Description of Problem…………………………………………………………………1

Proposed Solution ……………………………………………………………………2

Data and Methodologies

Data Source ……………………………………………………………………….3

Modern Portfolio Theory and Diversification ………………………………………3

Curse of Dimensionality and the Clustered Random Subspace Method...……..8

Results

Homogenous Universe………………………………………………………………11

Heterogeneous Universe…………………………………………………………....11

Conclusion and Further Research…………………………………………...………16

References ……………………………………………………………………………..…24

1. Introduction

1.1. Description of Problem

There are many pitfalls associated with building an investment portfolio. Institutional investors like pension plans and sovereign wealth funds are constantly faced with how to best allocate capital to various assets. Given their sheer size, these funds face challenges and constraints that are not applicable to most retail investors. As a consequence, even reasonably low levels of turnover for such funds can move markets and increase the associated costs of trading. Rapid shifts in portfolio allocations are impractical net of trading costs and therefore it is paramount that these funds build their investment portfolios to be as diversified as possible. This usually entails the analysis and modeling of a large universe of securities in order to identify what will provide the best opportunity.

The analysis of large security universes requires a sound statistical framework and a strong theoretical framework. An investment portfolio that is built to achieve optimal exposure to equities, for example, will require the study and monitoring of thousands of investable publicly traded companies. The formulation of a quantitative equity portfolio requires the researcher to estimate numerous statistical parameters. These input assumptions, estimated with limited historical data, are then used to arrive at theoretically optimal portfolios. The conventional approach to the problem is derived from the work of Harry Markowitz based on his 1952 work "Portfolio Selection". This book lays the foundation for much of modern day asset allocation under the familiar moniker of "Modern Portfolio Theory". The derivation of the efficient frontier, and identification of the minimum variance and tangency portfolios have provided generations of investors with better insight into the risks and rewards associated with building an investment portfolio from various assets.

One shared characteristic of the minimum variance and tangency portfolios is that they are often highly concentrated or poorly diversified. As a consequence, asset weights are sensitive to the ever-changing input assumptions such as the variance-covariance structure. With even the slightest estimation error, the optimal weights would shift significantly. As the number of assets increases for modeling, the so-called "Curse of Dimensionality" greatly magnifies this error.

The curse of dimensionality is a common problem in Machine Learning (ML) whereby researchers are faced with a high dimensional feature space with limited amounts of training data. To illustrate the problem, imagine searching for an object along a one hundred meter one-dimensional line. This problem is relatively simple. But what happens if we extend the search space in to 2 dimensions? Now the search space expands to a much larger square area that is 100 meters by 100 meters or 10,000 meters total. As the dimensionality grows further (ie 3 dimensions) the search space in turn increases. A larger search space requires commensurately more time to search with the same degree of effectiveness. Drawing a parallel to portfolio management, as the number of assets in the universe increases, the number of total input estimates required to solve for portfolio weights climbs exponentially. This in turn requires that the amount of historical data used to arrive at meaningful estimates also must increase exponentially. Since time series data is often limited, this becomes a serious problem when optimizing for efficient portfolios. Even if there was enough data, there is still the challenge of latency—markets are non-stationary and more recent data is more meaningful than older data. As a consequence, we need an approach that can effectively reduce dimensionality without increasing latency.

1.2. Solution

The plan of this paper is to explore the application of methods used in ML to reduce dimensionality. Traditional methods for dimensionality reduction used by ML researchers often include the use of Principal Component Analysis (PCA). PCA is able to find linearly d dimensions that are lower than the total number of variables in consideration. The dimensions are in form such that it represents the most variation of the original dataset. For example, given a 3 dimensional dataset, we can apply PCA to reduce the total number of dimensions via analyzing the principal components. When applying PCA, we get back eigenvectors and eigenvalues. Eigenvectors and eigenvalues are represented in pairs; for each column of eigenvector loadings, there exists an eigenvalue that summarizes that principal components explained variation. Therefore, if through our analysis finds that the first two principal components (ie, first 2 columns of the eigenvectors) explain 90% of the variability of the original dataset, we can decide to take out the last component thereby reducing our dimensionality.

The problem with PCA is the difficulty of applying it within the context of portfolio optimization. Instead, we propose using a modification of an algorithm called the Random Subspace Method (RSM), developed by Tin Kan Ho at Bell Labs, to directly address the problem. We show that RSM applied to the tangency portfolio optimizations facilitates improved returns over a large security universe. Furthermore, we will demonstrate that stability is improved when we overlay cluster analysis (Clustered RSM - CRSM). Through CRSM we are able to make the portfolio optimization universe agnostic.

2. Data and Methodology

2.1. Data

To validate our algorithm, we have chosen securities from a wide range of asset classes. This is to ensure our model is resilient to data pre-selection bias, which is widely prevalent in financial research when strategies and results are only valid for a small subset of asset universes.

All data are retrieved from Yahoo Finance and are adjusted for splits and dividends. In our universe of assets, we have selected both stocks from various indices and exchange traded funds (ETFs) to represent multiple asset types. Specifically, our data set covers asset classes ranging from single stock equity, equity indices, government bonds, corporate bonds, real estate, commodities, and precious metals. Our analysis will be conducted on data from 2006 to 2014 to facilitate testing on as wide a universe as possible, given that many ETFs were only listed in the past few years.

For added granularity, we have divided our universe into six different portfolios. Three of the six portfolios are built to contain highly correlated assets like equities. They include, the Standard and Poor's 100 stocks, U.S. equity sector ETFs, and international equity index ETFs. (Note that the components of the Standard and Poors 100 index are current day and therefore do not account for companies that transition into or out of the index during the period of our analysis.)

For the last three portfolios, they are built with diversification in mind; in other words, they contain assets that are dissimilar with respect to each other. Components of the portfolios includes equities, commodities, government bonds and corporate bonds. To test sensitivity we've also structure extreme portfolios that are highly skewed towards certain asset classes, mainly equities. For example, we place a bond ETF (IEF) into a highly homogenous asset universe composed of ether country equity index ETFs, or components of the S&P 100 index in order to illustrate the strengths and weakness of the different algorithms.

2.2. Modern Portfolio Theory and Diversification

Harry Markowitz introduced Modern Portfolio theory (MPT) in 1952, which provided investors with an elegant framework for quantitative portfolio construction. In this model, portfolio weights are derived from expected or historical asset returns and the variance covariance matrix in order to maximize return for a given level of risk. The core idea behind the framework stresses the importance of diversification - by mixing uncorrelated securities together, one is able to arrive at a final portfolio allocation that is less risky compared to holding any of the component securities individually. This is theoretically possible when you have asset returns that are not perfectly correlated. This imperfect correlation acts as an implicit hedge against risk.

We consider a portfolio that is composed of N assets. We denote $\boldsymbol{R}$ and $\boldsymbol{\sigma}$ to represent N dimensional return and risk vector respectively. Let the correlation matrix between each asset be denoted as $C = \{p_{ij}\}$, then the variance covariance matrix is simply $\sum = \left(p_{ij}\sigma_i\sigma_j\right)$ where $i \leq i, j \leq N$. For a given set of weight vector $\{w_1, \ldots, w_N\}$, the portfolio return and variance is given by matrix multiplication as:

$$R_p = w\boldsymbol{R}' \quad (1)$$

$$VAR_p = w'\sum w \quad (2)$$

From MPT, Markowitz also derived two timeless optimal portfolios known as the global minimum variance portfolio, and the maximum Sharpe portfolio or tangency portfolio. These portfolios lie on or toward the left-most border of what has come to be known as the efficient frontier. The efficient frontier lies on a Cartesian plane that comprises all possible combinations of efficient portfolios measured by their risk and return. Anything to the right of the efficient frontier is a feasible portfolio given a set of securities. The global minimum variance portfolio is then simply the portfolio that minimizes the portfolio level variance (furthest left) and is characterized by minimizing equation (2):

$$\text{argmin } VAR_p = w'\sum w \quad (3)$$

The Sharpe ratio of a portfolio is a risk adjusted performance measure that aims to quantify the risk associated with a rate of return. The higher the ratio, the better; the portfolio Sharpe ratio is:

$$Sharpe_p = \frac{w\boldsymbol{R'} - R_f}{\sqrt{w'\sum w}} \quad (4)$$

Maximizing this ratio given a set of weights will yield the tangency. In this paper, we will focus solely on the long only tangency portfolio for portfolio construction. As a further simplification, we will assume R_f to be 0.

Going back to equation (1) and (2), the portfolio return is just a weighted average of asset returns. The novel contribution made by Markowitz relates to the portfolio risk equation. At the center of it all is the variance covariance matrix, which is composed of the variances and pairwise covariances between each asset. The level of risk associated with a portfolio is influenced by assets' individual risks , but also by the overall level of diversification among assets in the portfolio. But how does one analyze and measure diversification?

There are numerous methods one might utilize to measure diversification. The most straight forward way of measuring diversification is whether we are allocating proportionately across a given universe of assets; the aim here is not to be concentrated in only a few assets. In this perspective, given N assets with weight vector w, the most diversified portfolio allocates $\frac{1}{N}$ percentage of capital to each individual security.

The 1/n or equal weight method of constructing a diversified portfolio is naïve however, as it hides the most important characteristic: risk. Recent popularity in the concept of risk parity focuses entirely on proportionally allocating to the inverse of individual securities' volatility. To calculate the risk parity portfolio weights, we simply normalize the inverse of the asset volatility risk vector $\frac{1}{\sigma}$ to sum to 1. While risk parity accounts for risk, it doesn't take into account how correlation across assets can also help reduce risk. An alternative approach proposed by Maillard, Roncalli, and Teiletche aims to equalize the risk contribution of each asset to final portfolio risk. More formally, the risk contribution for each asset can be derived as $\sigma = w \times [\sum \boldsymbol{w}]$, where w is weight and $\sum$ is the covariance matrix. Note that within the bracket, we are doing matrix multiplication as oppose to simple multiplication. Taking in to account the covariance matrix allows us to account for the two main levers for controlling risk, correlations and volatilities, but it doesn't account for the various drivers of risk.

Another framework that takes risk allocation a step further is the concept of equal risk factor allocation. Risk factors are essentially building blocks that explains an asset's risk and return characteristics. For example, the majority of risk and return associated with investing in the corporate bonds market can be explain by the macroeconomic interest rate movements and the underlying company's performance. Focusing on risk factor is more superior to asset specific characteristics like correlation because it's a lot less nosier. If we decompose the returns of assets to their respective risk factor exposures, we can quantitatively understand if we are over allocating to unwanted risk factors. This allows us to penetrate through all the noise and focus on what constitutes a diversified portfolio.

In Managing Diversification (Meucci, 2009), Meucci formulated the framework on decomposing returns into uncorrelated latent risk factors through PCA. Given the variance covariance matrix $\sum$, we can perform the following eigen decomposition:

$$E'\sum E \equiv A \tag{5}$$

In the above expression, E represents the eigenvectors or loadings $\{e_1, \ldots, e_N\}$while A represents a vector of eigenvalues $\{\lambda_1, \ldots, \lambda_N\}$ sorted in decreasing order of variance. Each eigenvector is a column vector that corresponds to each principal components' (PCs) respective asset exposures. Interpreted from a finance perspective, each PC can be interpreted as an individual principal portfolio. Each principal portfolio's return can be derived from the following equation:

$$R_{PC} = E^{-1}\boldsymbol{R} \tag{6}$$

The derived return stream can be viewed as risk factor returns and each portfolio's respective variance is defined by its corresponding eigenvalue. Since the eigenvalues are sorted decreasingly by variance this also means each principal portfolio explains decreasing variance.

Based on the work in Partovi and Caputo (2004), given a weight vector in asset space, one can convert it to principal space. The weights in principal space can be viewed as exposures to the various risk factors. With this information, it will allow us to identify over- and under- allocation to various risk factors. To convert between asset space and principal space, we utilize the following equation:

$$\breve{w} = E^{-1}w \tag{7}$$

Weight vector $\breve{w}$ represents the factor exposure of asset space weights w. Given the factor exposures we can calculate the variance contribution from the ith factor given the following equation:

$$\breve{var}_i = \breve{w}_i^2 \lambda_i \tag{8}$$

As mentioned earlier, the eigenvalues corresponds to each principal portfolio's variance. Then the weighted sum of all the eigenvalues is equal to the aggregate portfolio variance:

$$VAR_p = w'\sum w \equiv \sum_{i=1}^{N} \breve{var}_i \quad (9)$$

The diversification distribution curve as outlined in Meucci (2009) can then be calculated as

$$p_i \equiv \frac{\breve{var}_i}{VAR_p} \quad (10)$$

This above equation allows us to measure the variance contribution of the nth principal portfolio. In other words, it can be viewed as a measure of how risk is distributed across the factors. Any over allocation to risk factors can be viewed as concentration of risk. To bring this full circle, we can calculate the true diversity of a given weight vector via the entropy equation:

$$n_{entropy} = exp(-\sum_{i=1}^{N} p_i ln(p_i)) \quad (11)$$

This above equation yields the effective number of bets (ENB), which helps us measure the effective diversification given an asset space weight vector. For a fully concentrated portfolio, where all equity is allocated to a single asset, the number of bets (n) is 1. On the other hand, the most diversified portfolio will yield a value equal to N. We will use the ENB to measure the exposure towards various risk factors across various weighing schemes and portfolio universes. Such analysis will help us determine whether a given weighing scheme depends on its universe composition.

2.3. Curse of Dimensionality and the Random Subspace Method

The curse of dimensionality problem in machine learning usually reveals itself in the feature selection process. When training a model for classification, many researchers face the problem of deciding what features should be included in their analysis such that out of sample model performance is maximized. To maintain the same density of samples with increases in the number of features, we will need a larger datasets.

Relating this back to finance, one of the central statistical input parameters used for optimization is the covariance matrix. Since it takes into account the pairwise correlation structure of input variables, the number of unique estimates for a covariance matrix equals:

$$\frac{N(N+1)}{2} \tag{12}$$

Where N equals the number of assets. It is important to recognize that as the number of asset increases, the number of unique values grows at approximately a quadratic pace or $O(n^2)$. To maintain the same data density, data points also have to grow at a rapid pace. For example, if 250 days of data for each asset is enough for estimating a reliable covariance matrix for a 3 assets universe, then when we increase our universe size to contain 10 assets, we will need 4167 days of data for each asset. If we ignored the data requirements and performed portfolio optimization, we are essentially conducting an over-fit exercise to fit our model to past data. In problems whereby additional data is not available, we may be better off just assuming that assets are uncorrelated, in other words assuming the entire covariance matrix is zero except the diagonal variance estimates. Another important factor that should be considered is whether the underlying distribution of the data is known. If we are dealing with an unknown distribution, it is always prudent to incorporate more data for parameter estimation.

One method used in machine learning that has the capabilities of reducing the impact of curse of dimensionality is the Random Subspace Method (Ho, 1999). Developed by Ho at Bell Laboratories, it is an ensemble meta-algorithm originally developed to construct robust decision trees. In this paper, we have adapted it towards portfolio optimization and show that through the application of this method, we are able to improve results when dealing with large universes of securities.

The main idea behind RSM is that it acknowledges the fact that through combining weak learners we are able to come up with a strong learner. From a ML perspective, RSM aims to randomly split the entire feature space into smaller subspaces. From there, a classifier is trained on each subspace. In order to produce a final prediction, each individual subspace classifier must generate a prediction, which will then be aggregated by ether simple averaging or voting.

From a finance perspective, optimization that is conducted directly on the sample covariance matrix derived from a large universe of assets is prone to estimation error as the estimated statistical input parameters are noisy and sparse. To reduce the impact of sparsity, we randomly choose k assets from the portfolio s times with replacement and perform mini optimizations on each of these mini portfolios, or subspaces. The final weight vector is the average of all the individual weights from each of the mini optimizations. This method is similar to the concept of 'divide and conquer', and its' superiority over direct optimization results from the fact that focusing on a smaller subspace of assets reduces the dimensionality of the problem relative to the fixed amount of data we have.

2.3.1 Universe Specification

While RSM overlaid on top of large-scale portfolio optimization was designed to reduce the impact of the curse of dimensionality, the model is heavily dependent upon the type of securities within the portfolio. For example, if the universe contains 6 equity-like asset classes, 2 bonds and 2 commodities, the probability of sampling an equity asset in to a subspace is 60%, while the probability of sampling a bond or commodity asset is just 20%. ***This means that if our asset universe contains a large number of securities that belong to the same group, applying RSM on top of portfolio optimization will result in under diversification and over concentration***.

To combat this problem, we propose to introduce a clustering filter before performing RSM, which we call Cluster Random Subspace Method (CRSM). Cluster analysis is a group of unsupervised learning algorithms that aims to sort random variables into groups. Variables within groups are highly similar while variables between different groups are highly dissimilar. In our analysis, we employed the common K-means clustering algorithm. To cluster assets we employ the correlation structure across assets as the distance matrix. More specifically, our distance matrix is simply $dist = 1 - \rho$, where ρ is the correlation of asset returns. To avoid parameter bias we have chosen to vary the number of centers with the elbow algorithm. The elbow algorithm chooses the number of centers that maximizes the marginal percentage of variance explained.

The main reason for applying clustering before performing the RSM is that it equalizes the probability that assets from each source of risk will be drawn in each subspace sample, limiting the probability of sub-space over-concentration.

There are two approaches for applying the CRSM, one with replacement and one without. As mentioned above, the sampling probabilities for RSM doesn't take in to account the portfolio compositions so there is a chance that a large proportion of sampled subspaces contain assets that behave similarly. CRSM with replacement is structured in such a way that it accounts for this probability distortion. The algorithm is as follows: for c clusters, we will randomly choose a number between $\{1 \dots c\}$, k times. These k randomly chosen numbers represent clusters from which assets will be selected for each subspace sample. The replacement feature ensures that the probability of choosing assets from any cluster stays the same across k random samples. On the other hand, CRSM without replacement means that once we have chosen an asset from that cluster, we will not be able to sample from this cluster again. In other words, each sample will contain an asset from separate clusters.

These two modified algorithms' parameters, k and s, will be dynamic to ensure model simplicity and robustness. For parameter k, we will set it equal to the number of dynamically derived clusters at each rebalance period, while for parameter s, we will apply the following simple heuristic:

$$s = N * \sqrt{c} \qquad (13)$$

Where c equals the number of clusters. This equation is formulated in such a way that it takes enough samples to properly account for all assets, which improves diversity.

3. Results and Discussion

3.1. Homogenous Universe

Applying the Markowitz algorithm - Max Sharpe Optimization (MSO) - to a large homogenous asset universe produces sub-optimal performance. This is primarily caused by the fact that the covariance structure of such universes tend to be very noisy. What is shown to be useful information in one period will likely be noise that will dissipate in another. Tables 1-6 shows the summary statistics for the various optimization procedures conducted on our six uniquely selected portfolios. As our benchmark, we have chosen to focus on an Equal Weight Portfolio and the Tangency portfolio. The equal weight portfolio assumes we know nothing about the risk, return, and correlation characteristics of our portfolio of assets.

At first glance, the daily mean returns and volatilities in the first three columns of table 1 and 2 (SP100, Sector, and Country), confirms how poorly the tangency portfolio performed when confronted with a highly homogenous security universe. More specifically, the daily mean return and volatility for the tangency portfolio is approximately 40% less and 13% higher than the simple equal weight portfolio respectively. In each of these instances, the investor would have performed better simply by just allocating equal proportions to each asset and periodically rebalancing as oppose to having to deal with the intricacies of estimating risks, returns and correlations. But through ignoring these input parameters, the investor makes the assumption that all securities are the same. The macroeconomic relationship across business cycles are not constant, for example, it wasn't always the case that bonds and stocks had negative correlation.

Our overlay of the Random Subspace Method on top of portfolio optimization shows an improvement over to the tangency portfolio. While returns are not as high when compared to equal weight portfolio, we can see that RSM, CRSM-R, and CRSM show higher Sharpe ratios (Table 6) relative to our benchmarks. In other words, while our algorithm wasn't able to generate the highest returns, it was able to generate the most *stable* returns. A more intuitive visualization is displayed in tables 7-12, where we rank the relative measures (higher the better).

The effectiveness of overlaying the RSM on portfolio optimization is attributed to the fact that by sampling numerous different times, we get subspaces that are representative of all securities inside a universe. Once we aggregate all the subspaces via equal weight averaging, we end up with a more proportional allocation. By spreading out our weights across all assets, we inherently build a buffer for error out of sample.

For example, for a given universe of equity assets, we can expect several stocks to be in the same industry, maybe even serving the same geographical area. These fundamental factors may give them similar future prospects which are not reflected in the input assumptions that go into MSO. With MSO's built in feature of holding concentrated assets, the margin for error out of sample is much larger to a more diversified portfolio.

3.2 Heterogeneous Universe

The poor performance of equal weight portfolios is most evident when we test on heterogeneous universes such as a diversified basket of asset classes. By ignoring relative returns, risks and correlations, an equal weight portfolio is exposed to comparatively severe drawdowns during volatile market conditions. The naïve RSM method whereby no clustering is applied before optimization is useful for homogeneous asset universes. However, as we discussed, the primary weakness of RSM is that it is highly dependent on the security composition of the universe. For a universe that incorporates many equity-like assets along with a few non-equity asset classes like commodities and fixed income, the sampling probabilities will be largely skewed towards equities because that is the asset class that is most representative in the universe. We use the heterogeneous universes to highlight this deficiency. RSM performs poorly compared to MSR and CRSM variants in these universes, and does not adequately take advantage of the risk reducing asset classes.

CRSM solves this problem by adjusting the sampling probabilities via clustering. CRSM with replacement is a direct extension to RSO. It adjusts the probabilities by grouping assets into their respective clusters before randomly choosing a cluster to sample from with replacement. This increases the chances that it will identify assets or combine portfolios that will reduce risk. In tables 1-12 we can see that the CRSM and CRSM-R can easily self-adjust to changes in universe composition. The performance of the CRSM variations are more consistent versus RSM and MSR in both homogenous and heterogeneous and control asset universes.

The results show that both CRSM methods - with and without replacement - improve diversification significantly as measured by each portfolio's average exposure to risk factors (ENB). In the IEF plus S&P 100 universe for example, the effective number of bets for CRSM without replacement is 5 times higher than it is with standard RSM. For all three universes, CRSM methods reduce risk and limit maximum drawdowns significantly.

In table 13 we aggregated the performance statistics across universes and show a final score for each sizing algorithm. We can see that all three RSM algorithms rank in on the top 3, with our CRSM variations showing the best performance. This is expected given the adaptive nature and their ability to curb the curse of dimensionality present in portfolio optimization on large data sets.

4. Conclusion

In this paper, we have demonstrated the application of different RSM methods to large-scale portfolio optimization. In addition, we modified the RSM method to create a novel approach- CRSM- that can dynamically adjust to different asset universes by using clusters to draw samples. Such novel procedure allows the algorithm to become more adaptive to the asset composition within an arbitrarily selected universe. Our results show that the pure RSM approach is much less stable across a variety of universes relative to CRSM. For example, in heterogeneous universes- such as a broad mix of asset classes- RSM applies naive diversification and holds a lot of assets but does not create allocations that capture opportunities for superior risk reduction. The larger maximum drawdowns and higher volatility of RSM versus MSR and CRSM portfolios on the heterogeneous universes highlight this inefficiency. Overall, CRSM demonstrates the highest aggregate scores/metrics across universe in comparison to alternative methods such as MSR (traditional mean-variance) and equal weight, and dominates traditional RSM. Given this nature, CRSM is an excellent procedure to utilize when optimizing for a large universe of securities.

The charts below (Figure 1-6) show the aggregate metric computed for each allocation method across all of the different universes used for testing. This is a measure of the average success of each allocation method. CRSM is Cluster Random Subspace Method, CRSM-R is Cluster Random Subspace Method with Replacement, MSR is the Max Sharpe/Tangency portfolio in Markowitz, RSM is the original Random Subspace Method, Equal Weight is an equal weight portfolio of all securities in the chosen universe.

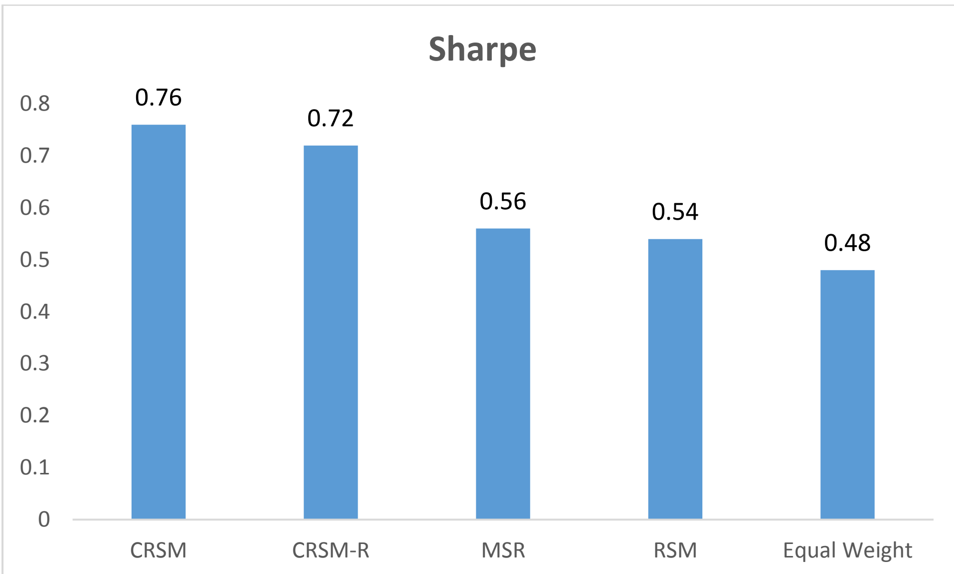

Figure 1

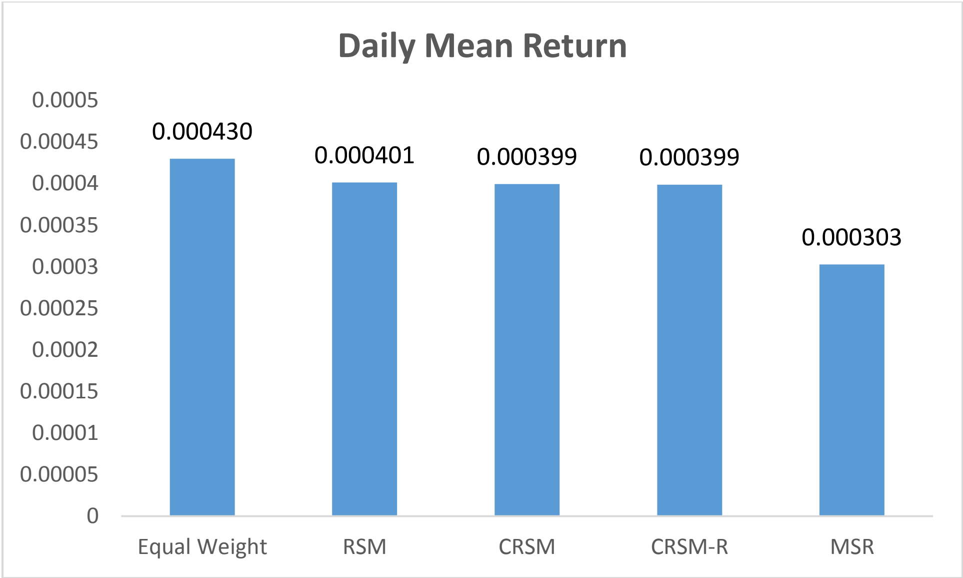

Figure 2

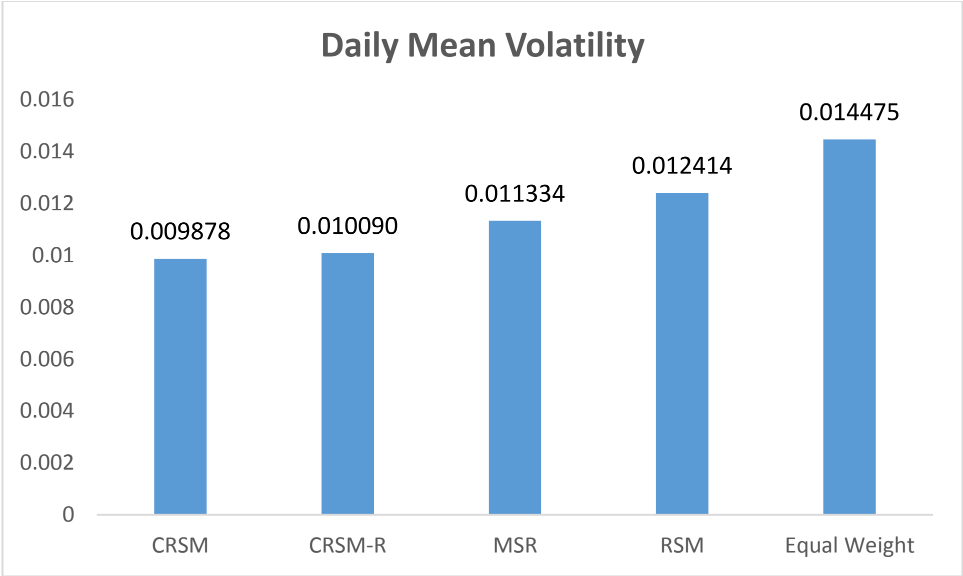

Figure 3

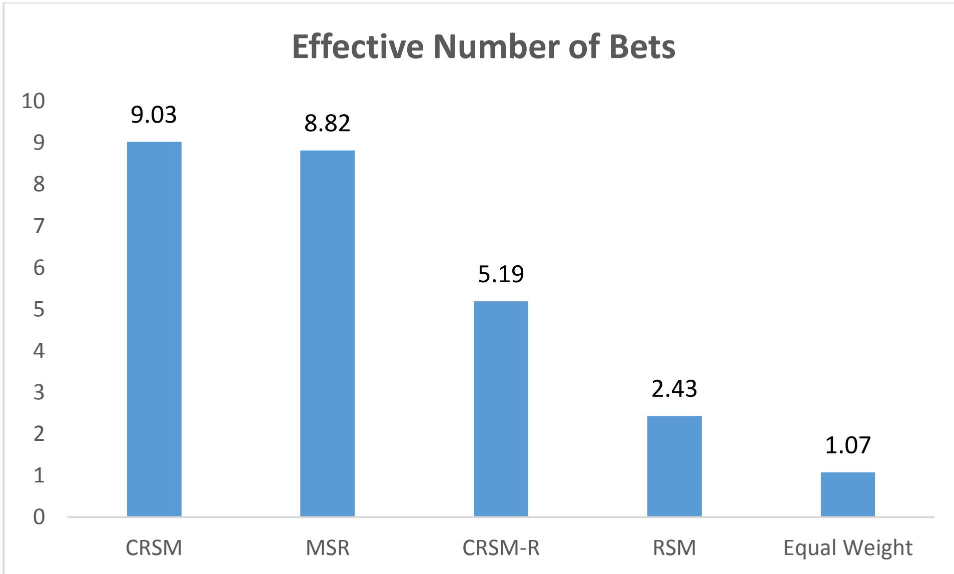

Figure 4

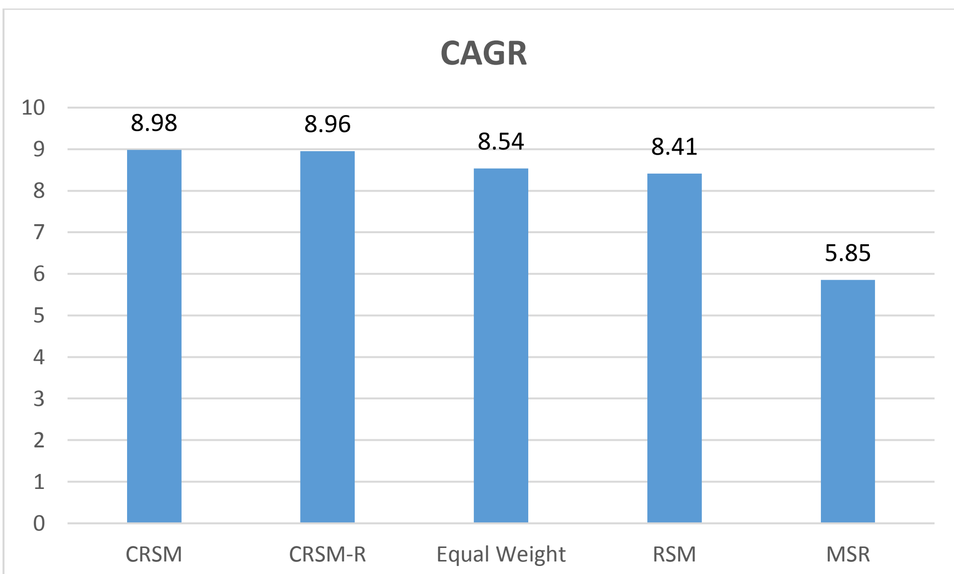

Figure 5

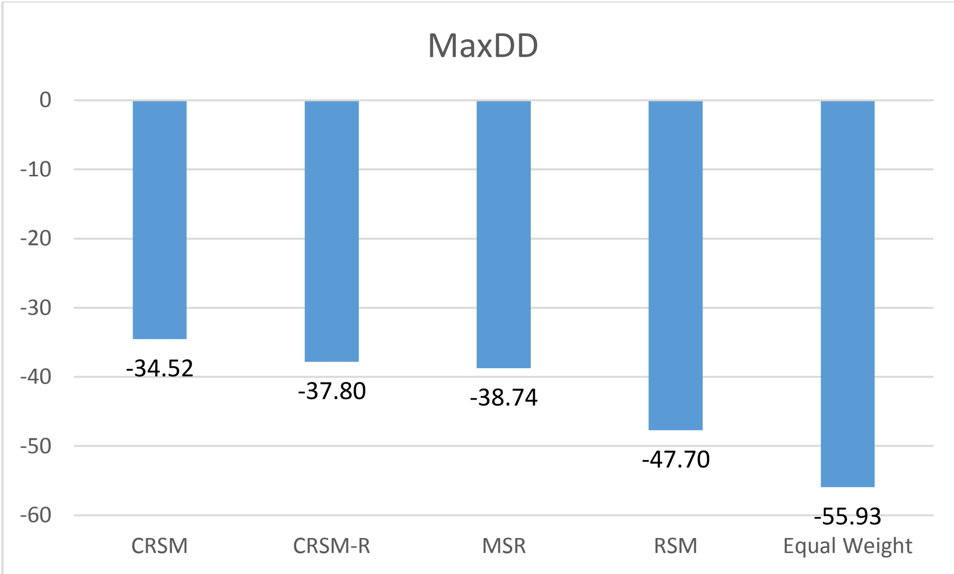

Figure 6

The tables below (tables 1-3) show the summary breakdown across metrics by universe and allocation method.

Daily Mean Return

	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	0.000545627	0.000391347	0.000376714	0.000371627	0.000541921	0.000352608	0.000429974
MSR	0.000326534	0.00018831	0.00025871	0.000305526	0.000390666	0.000345643	0.000302565
RSM	0.000452585	0.000359609	0.00039042	0.000378042	0.000464465	0.000363075	0.000401366
CRSM	0.000506759	0.00031664	0.000458884	0.000388708	0.000403149	0.000322486	0.000399438
CRSM-R	0.000493297	0.000346773	0.000421441	0.000369694	0.000470638	0.000290946	0.000398798

Daily Mean Volatility

	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	0.013953124	0.012947882	0.016417757	0.015932916	0.013778493	0.013817289	0.014474577
MSR	0.018029372	0.013222754	0.0176378	0.005969947	0.007222773	0.005922934	0.011334263
RSM	0.013806491	0.012721254	0.016190019	0.012121749	0.012025092	0.007620161	0.012414128
CRSM	0.013081129	0.012696417	0.016346063	0.006099334	0.005503027	0.00554429	0.009878376
CRSM-R	0.012899883	0.012588736	0.016057391	0.007122088	0.006269274	0.005601091	0.010089744

Effective Number of Bets

	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	1.124946119	1.077017905	1.030540608	1.031653241	1.125419094	1.053042824	1.073769965
MSR	6.925211545	2.25232099	3.039992021	9.421003059	21.35909671	9.929639401	8.821210621
RSM	2.938038047	1.590169964	1.744143852	1.911937665	2.437340805	3.976697859	2.433054698
CRSM	4.481467811	2.072392648	2.539826743	10.24272308	24.64923418	10.16944142	9.025847646
CRSM-R	3.605086881	1.865448432	2.038612409	4.325088318	10.17128274	9.128156391	5.188945862

Table 1,2,3

CAGR

	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	11.94	8.05	6.28	6.35	11.91	6.69	8.536666667
MSR	4.2	2.57	2.62	7.51	9.61	8.61	5.853333333
RSM	9.41	7.28	6.74	7.92	10.37	8.74	8.41
CRSM	11.19	6.12	8.53	9.76	10.26	8.04	8.983333333
CRSM-R	10.88	6.97	7.64	9.05	12.02	7.17	8.955

MaxDD

	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	-52.43	-53.51	-61.87	-60.67	-51.94	-55.18	-55.9333333
MSR	-56.73	-56.55	-68.89	-14.35	-20.25	-15.65	-38.7366667
RSM	-49.76	-52.58	-63.20	-48.36	-44.65	-27.63	-47.6977778
CRSM	-51.23	-52.81	-63.3	-10.36	-12.54	-16.87	-34.5183333
CRSM-R	-50.91	-53.58	-65.45	-22.51	-16.93	-17.44	-37.8033333

Sharpes (0%)

	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	0.62	0.48	0.36	0.37	0.62	0.41	0.476666667
MSR	0.29	0.23	0.23	0.81	0.86	0.93	0.558333333
RSM	0.52	0.45	0.38	0.50	0.61	0.78	0.541111111
CRSM	0.61	0.4	0.45	1.01	1.16	0.92	0.758333333
CRSM-R	0.61	0.44	0.42	0.82	1.19	0.82	0.716666667

The tables below (tables 4-12) show the ranking by metric by allocation method and by universe. (Higher ranks are better)

Daily Mean Return

	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	5	5	2	3	5	4	24
MSR	1	1	1	1	1	3	8
RSM	2	4	3	4	3	5	21
CRSM	4	2	5	5	2	2	20
CRSM-R	3	3	4	2	4	1	17

Daily Mean Volatility

	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	2	2	2	1	1	1	9
MSR	1	1	1	5	3	3	14
RSM	3	3	4	2	2	2	16
CRSM	4	4	3	4	5	5	25
CRSM-R	5	5	5	3	4	4	26

Effective Number of Bets

	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	1	1	1	1	1	1	6
MSR	5	5	5	4	4	4	27
RSM	2	2	2	2	2	2	12
CRSM	4	4	4	5	5	5	27
CRSM-R	3	3	3	3	3	3	18

Table 7 8 9

CAGR	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	5	5	2	1	4	1	18
MSR	1	1	1	2	1	4	10
RSM	2	4	3	3	3	5	20
CRSM	4	2	5	5	2	3	21
CRSM-R	3	3	4	4	5	2	21

MaxDD	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	2	3	5	1	1	1	13
MSR	1	1	1	4	3	5	15
RSM	5	5	4	2	2	2	20
CRSM	3	4	3	5	5	4	24
CRSM-R	4	2	2	3	4	3	18

Sharpes (0%)	SP100	Sector	Country	Country Bond	SP100 Bond	Diversified	Average
Equal Weight	5	5	2	1	2	1	16
MSR	1	1	1	3	3	5	14
RSM	2	4	3	2	1	2	14
CRSM	3	2	5	5	4	4	23
CRSM-R	3	3	4	4	5	3	22

Table 10,11,12

Aggregate Rankings	Daily Mean Return	Daily Mean Volatility	Effective Number of Bets	MaxDD	Sharpes (0%)	Aggregate
Equal Weight	24	9	6	13	16	68
MSR	8	14	27	15	14	78
RSM	21	16	12	20	14	83
CRSM	20	25	27	24	23	119
CRSM-R	17	26	18	18	22	101

Table 13

References

Dalio, Ray. 'Engineering Targeted Returns & Risk'. *Bridgewater Associates* (2014): n. pag. Web. 3 Nov. 2014.

Ho, Tin Kam. "The random subspace method for constructing decision forests." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20. 8 (1998): 832--844. Print.

Kirchner, Ulrich and Caroline Zunckel. "Measuring Portfolio Diversification." *arXiv preprint arXiv:1102.4722*, (2011): Print.

Kritzman, Mark. "Are Optimizers Error Maximizers?." *The Journal of Portfolio Management*, 32. 4 (2006): 66--69. Print.

Maillard, Sébastien, Thierry Roncalli, and Jerome Teiletche. 'On The Properties Of Equally-Weighted Risk Contributions Portfolios'. *SSRN Journal* n. pag. Web.

Markowitz, Harry. "Portfolio selection*." *The journal of finance*, 7. 1 (1952): 77--91. Print.

Meucci, Attilio. "Managing diversification." (2009): Print.

'Risk Parity Is About Balance'. *Bridgewater Associates* (2014): n. pag. Web. 3 Nov. 2014.

Varadi, David and Michael Kapler. "Cluster Risk Parity." *CSSA*, 2013. Web. 01 Jan 2014. <http://cssanalytics.wordpress.com/2013/01/03/cluster-risk-parity/>.

Varadi, David. "Random Subspace Optimization." *CSSA*, 2013. Web. 01 Jan 2014. < http://cssanalytics.wordpress.com/2013/10/06/random-subspace-optimization-rso//>.