A Machine Learning approach for Asset Allocation

Wanda Tsewu (wanda@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)

Supervised by: Professor Phillip Mashele
North-West University, South Africa

12 October 2018

Submitted in partial fulfillment of a structured masters degree at AIMS South Africa

African Institute for Mathematical Sciences SOUTH AFRICA

Abstract

Quadratic optimizers, in particular Markowitz's brilliant Critical Line Algorithm (CLA), although mathematically correct, generally return unreliable solutions due to their instability, concentration and underperformance. CLA's instability issue is given rise by the need to invert a positive-definite covariance matrix, especially when it is numerically ill-conditioned. Graph theory and machine learning techniques allow for the exploration of a hierarchical clustering based asset allocation method called Hierarchical Risk Parity (HRP) developed by M. Lopez de Prado. HRP does not require this invertibility, instead, it exploits the hierarchical structure inherent in the covariance matrix to model a less complex, stable and intuitive solution to the asset allocation problem. In this paper we have closely studied the inner workings of the HRP algorithm and used it to allocate assets for a portfolio problem which consists of 13 Johannesburg Stock Exchange (JSE) equities.

Keywords: Portfolio Theory, Hierarchical Clustering, Asset Allocation, Machine Learning, Risk Parity

Declaration

I, the undersigned, hereby declare that the work contained in this research project is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

Wanda Tsewu, 12 October 2018

Contents

1. Introduction

Harry Markowitz, in 1952, set out on a quest to find a mathematical solution to optimize the most recurrent financial problem of asset allocation. What he found was the notion of the 'efficient frontier' which he described as: "the investor does (or should) consider expected return a desirable thing and variance of return an undesirable thing" (Markowitz, 1952). This was the answer to the problem the investor faces when trying to optimally allocate and manage assets in an uncertain environment (Meucci, 2005). Markowitz's model, named the mean-variance portfolio, revolutionized the financial world and became the most widely used model in the asset allocation problem. James Tobin and William Sharpe further extended Markowitz's great work. They both, as their predecessor did, received the economics Nobel prize for their work. (Tobin, 1958) was able to show that the efficient frontier becomes a straight line when one asset is risk-free in a two asset portfolio model. (Sharpe, 1964), along-side John Lintner and Jack Treynor developed the capital asset pricing model (CAPM theory) where they found that the expected risk premium of the asset (difference between the expected return) and its beta (systematic risk with respect to the tangency portfolio) is directly proportional in a competitive market (Roncalli, 2013). This gave rise to passive investment management (like risk management and insurance) while active investment management continued to use Markowitz's model to optimize portfolios.

Like many models, Markowitz's optimization has numerous strengths and weaknesses. It is a scientific approach to a finance problem that is easy to use and also simple to explain. Its shortcomings have been well tested by the major disruptions in the economy over the past decades, namely the dot-com crisis in the 90's and the 2008 financial crisis (Roncalli, 2013). It failed dismally in the latter with financial institutions and investors losing large amounts of money. The primary problem with the optimization model is that in its solution it requires the inversion of a positive definite covariance matrix (as we shall later explore) and thus it is very sensitive to input parameters such as expected returns. These expected returns are forecasted using averaged long-term historical data which is then added to the current interest rate (Merton, 1980), and any small change in the expected return of an individual asset further exacerbates the matrix inversion problem.

Before we examine the matrix inversion problem, let's take a step back and take a look at an overview of Modern Portfolio Theory and how Harry Markowitz revolutionized the world of asset management with his model. We also introduce some of the jargon and notions used such as how risk and return interact to form the efficient frontier. This introductory chapter is then concluded with the pitfalls of Markowitz's quadratic optimization model (matrix inversion problem).

In Chapter 2 we look to explore how brilliant researchers such as (Black and Litterman, 1992) and (Ledoit and Wolf, 2003) tended back to the proverbial drawing board in Markowitz's behalf, in pursuit of a solution to his model's shortcomings, before we introduce the more modern machine learning technique used in this paper that addresses those very issues.

Chapter 3 is a dissection of the machine learning algorithm called Hierarchical Risk Parity. We examine its in and outs, showing how it works using local data from the Johannesburg Stock Exchange (JSE) and then present the research results obtained.

1.1 Modern Portfolio Theory

Harry Markowitz's revolutionary optimization model allows an investor to manage risk by investing in a number of shares (or assets/securities) at the same time. The combination of these shares is called a portfolio (Jordaan, 2018).

1.1.1 Efficient Frontier. Markowitz portfolio analysis is a mathematical tool that is used to find the optimum portfolios (Francis and Kim, 2013). In his 1952 paper, Markowitz found that these optimum portfolios (not just an individual one) lied on the efficient frontier (Roncalli, 2013). This frontier is illustrated in Figure 1.1 as the curvature formed as a result of varying combinations.of shares in a portfolio. These efficient portfolios have the following features (Francis and Kim, 2013):

- They have the greatest expected return for a given level of risk, or
- offer the lowest risk for a given level of expected return.

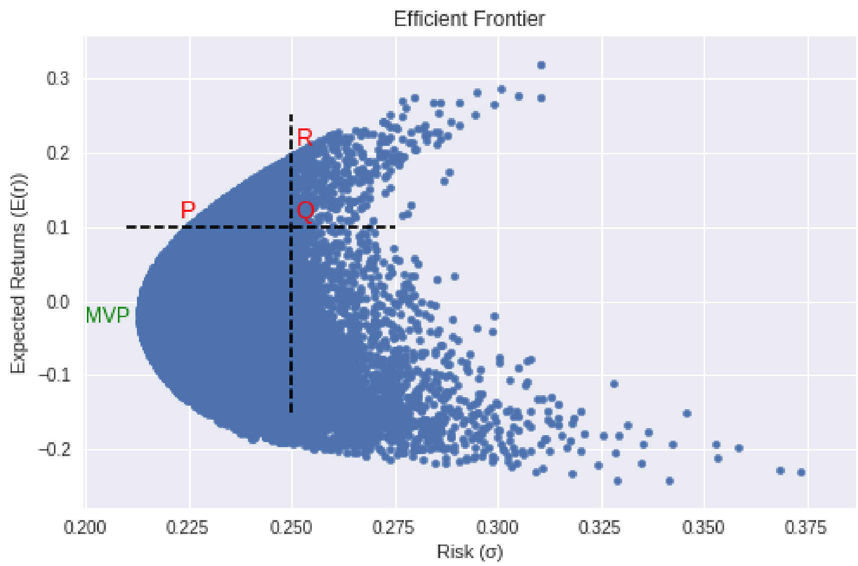

Figure 1.1: 50000 simulated portfolios from 5 Johannesburg Stock Exchange (JSE) securities: BHP Billiton plc, Naspers Ltd, MTN Group Ltd and Sanlam Ltd, Anglo American Platinum Ltd.

Using data obtained from Yahoo Finance spanning two years from January 2014 to January 2016 of five securities of the JSE, along with the use of a helpful python code written by Bernard Brenyah (Medium), we've simulated fifty thousand portfolios in an effort to illustrate the efficient frontier. The bullet-shaped curve formed to the left of this opportunity set represents the efficient frontier upon which the efficient portfolios lie. These portfolios are efficient simply because if, for example, an investor expecting a return on investment of at least 10% would rather bet his money on the portfolio that lies at P because it has the least amount of risk than any other portfolio in that 10% expected return class, say portfolio Q for instance. Conversely an investor that could stomach a level of risk of 25% would rather bet his money on portfolio R than any other portfolio on that same risk line because it offers the greatest level of expected return in that risk class, once again better than portfolio Q. The left most portfolio is called the minimum variance portfolio (MVP), as the name suggests this portfolio carries the least risk. Unfortunately in this particular case it also has a negative expected return, which is not always the case.

1.1.2 Return and Risk of a portfolio. Consider n assets with rates of return $r_1, r_2, \ldots, r_n$ with means $\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_n$. To form a portfolio one assigns weights w_i to each asset in proportion to its purchase cost (Jordaan, 2018). Then the portfolio return is

$$r = \sum_{i=1}^{n} w_i r_i, \tag{1.1.1}$$

where the rate of return of an individual asset (r_i) is defined as

$$r_i = lnS_i - lnS_0. \tag{1.1.2}$$

The share price S is a random process (usually on the stock market) and so the rate of return (r) of an asset is also random which is why we call $\bar{r}$ the expected rate of return. The expected portfolio return is then given by the weighted sum of each asset's **expected return**

$$E(r) = \bar{r} = \sum_{i=1}^{n} w_i \bar{r}_i. \tag{1.1.3}$$

The higher the expected return $\bar{r}_i$ of an individual asset, the higher the reward associated with that asset, but, unfortunately, the higher the risk associated with that asset as well.

The **variance** measures the volatility or degree of **risk** associated with an asset. This spread from the mean is defined by

$$\begin{aligned} var(r) = \sigma^2 &= E[(r - \bar{r})^2] & (1.1.4)\\ &= E[r^2 - 2r\bar{r} + \bar{r}^2] & (1.1.5)\\ &= E[r^2] - 2\bar{r}\bar{r} + \bar{r}^2 & (1.1.6)\\ &= E[r^2] - \bar{r}^2. & (1.1.7) \end{aligned}$$

Substituting equation (1.1.1) and equation (1.1.3) into (1.1.7):

$$\begin{aligned} \sigma^2 &= E[(\sum_{i=1}^{n} w_i r_i)^2] - (\sum_{i=1}^{n} w_i \bar{r}_i)^2 & (1.1.8)\\ &= E[(\sum_i w_i r_i)(\sum_j w_j r_j)] - (\sum_i w_i \bar{r}_i)(\sum_j w_j \bar{r}_j) & (1.1.9)\\ &= E[\sum_{i,j=1}^{n} w_i w_j r_i r_j] - \sum_{i,j=1}^{n} w_i w_j \bar{r}_i \bar{r}_j & (1.1.10)\\ &= \sum_{i,j=1}^{n} w_i w_j E[r_i r_j] - \sum_{i,j=1}^{n} w_i w_j \bar{r}_i \bar{r}_j & (1.1.11)\\ &= \sum_{i,j=1}^{n} w_i w_j (E[r_i r_j] - \bar{r}_i \bar{r}_j). & (1.1.12) \end{aligned}$$

Where

$$E[r_i r_j] - \bar{r}_i \bar{r}_j = \sigma_{ij} \tag{1.1.13}$$

is the covariance between assets i and j and therefore

$$\sigma^2 = \sum_{i,j=1}^{n} w_i w_j \sigma_{ij} \quad [\sigma_{ii} = \sigma_i^2]. \tag{1.1.14}$$

A rational portfolio manager wishes to minimize this risk.

1.1.3 Matrix inversion problem. As a portfolio manager you want to invest in the portfolios that lie on the efficient frontier. These portfolios are those that either have the greatest expected return for a given level of risk or, offer the lowest amount of risk for a fixed level of expected return. Let's say we want to obtain these portfolios using the latter method. If we keep the expected return fixed then what we want are the optimal weights that minimize the risk (or variance) of the portfolio (Francis and Kim, 2013). What we have at hand is an optimization problem where we:

$$\text{minimize} \quad \sigma_p^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij} \tag{1.1.15}$$

subject to three constraints:

$$\sum_{i=1}^{N} w_i E(r_i) = E(r_p) \tag{1.1.16}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{1.1.17}$$

$$\text{where } w_i \geq 0. \tag{1.1.18}$$

The first constraint (1.1.16) represents the fixed, desired level of expected return while the second constraint (1.1.18) requires that the optimal weights that fulfill optimization problem (1.1.15) sum to 1. These two constraints are called Lagrangian constraints because they are equality constraints and as such need to be multiplied by Lagrange multipliers (λ and γ) and added to equation (1.1.15) to form a Lagrangian objective function for the risk-minimization problem.

$$\text{Minimize } \mathcal{L} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij} + \lambda[E(r_p) - \sum_{i=1}^{N} w_i E(r_i)] + \gamma[1 - \sum_{i=1}^{N} w_i] \tag{1.1.19}$$

The minimum-variance portfolio is found by setting partial derivatives $\frac{\partial \mathcal{L}}{\partial w_i}$ for $i = 1, \ldots, N$, $\frac{\partial \mathcal{L}}{\partial \lambda}$ and $\frac{\partial \mathcal{L}}{\partial \gamma}$ all equal to zero. This results in the following $N+2$ linear equations.

$$\frac{\partial \mathcal{L}}{\partial w_1} = w_1\sigma_{11} + w_2\sigma_{12} + \ldots + w_N\sigma_{1N} - \lambda E(r_1) - \gamma = 0 \tag{1.1.20}$$

$$\frac{\partial \mathcal{L}}{\partial w_2} = w_1\sigma_{21} + w_2\sigma_{22} + \ldots + w_N\sigma_{2N} - \lambda E(r_2) - \gamma = 0 \tag{1.1.21}$$

$$\vdots \tag{1.1.22}$$

$$\frac{\partial \mathcal{L}}{\partial w_N} = w_1\sigma_{N1} + w_2\sigma_{N2} + \ldots + w_N\sigma_{NN} - \lambda E(r_1) - \gamma = 0 \tag{1.1.23}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = w_1 E(r_1) + w_2 E(r_2) + \ldots + w_N E(r_N) - E(r_p) = 0 \tag{1.1.24}$$

$$\frac{\partial \mathcal{L}}{\partial \gamma} = w_1 + w_2 + w_N - 1 = 0 \tag{1.1.25}$$

The above system of linear equations can be expressed as a Jacobian matrix equation:

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1N} & E(r_1) & 1 \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2N} & E(r_2) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \dots & \sigma_{NN} & E(r_N) & 1 \\ E(r_1) & E(r_2) & \dots & E(r_N) & 0 & 0 \\ 1 & 1 & \dots & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ -\lambda \\ -\gamma \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ E(r_p) \\ 1 \end{bmatrix}$$

which in matrix notation can be represented as

$$\mathbf{Ax} = \mathbf{b} \tag{1.1.26}$$

where $\mathbf{A}$ is the $(N+2) \times (N+2)$ coefficient matrix, $\mathbf{x}$ is the $(N+2) \times 1$ vector of weights and $\mathbf{b}$ is the $(N+2) \times 1$ vector of constants. We wish to solve for $\mathbf{x}$ and find the optimal weights in the above equation. Using matrix notation we may commence as follows:

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \tag{1.1.27}$$
$$\mathbf{Ix} = \mathbf{A}^{-1}\mathbf{b} \tag{1.1.28}$$
$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}. \tag{1.1.29}$$
$$\tag{1.1.30}$$

The solution will result in the optimal N weights which minimize the variance of the portfolio and the values for the Lagrangian multipliers λ and γ.

The mean-variance optimization method requires the input of the estimations of expected returns and covariance matrix for each asset (Chaves et al., 2010). These estimations can be at times entries of a large matrix $\mathbf{A}$ in equation (1.1.26). Given the stochastic nature of the stock markets, the estimations of these entries proves to be very difficult, in particular those of the expected returns (Lopez de Prado, 2016), this means that one small change in an input expected return for one asset may result in a totally different portfolio (Mankert, 2006). A further issue is that the estimates are formulated using historical data which can make them too noisy to be useful ((Chaves et al., 2010) and (Michaud, 2014)). This leads to large errors when the positive-definite matrix $\mathbf{A}$ is to be inverted to obtain the weights of vector $\mathbf{x}$ in equation (1.1.30). (Lopez de Prado, 2016) also states that another reason for the instability of quadratic optimizer (like Markowitz's Critical Line Algorithm (CLA)) the vector space of matrix $\mathbf{A}$ is modelled as a complete (fully connected) graph (see Figure (1.2)), where every node is a potential substitute of another. What this means in portfolio asset allocation is that the weights are allowed to vary in unintented ways. If we take an example of a portfolio containing the top 40 securities from the JSE then we have a 40×40 covariance matrix with 1600 variance-covariance entries. Figure 1.2 shows how the complete (fully connected) graph would look for just any 15 companies. To try overcome this sensitivity problem academics such as (Black and Litterman, 1992) created more robust asset allocation models to improve on Markowitz's work using bayesian priors, (Clarke et al., 2002) added more constraints (specifically short selling constraints) in their attempt to solve the problem, and (Ledoit and Wolf, 2003) used regularization methods.

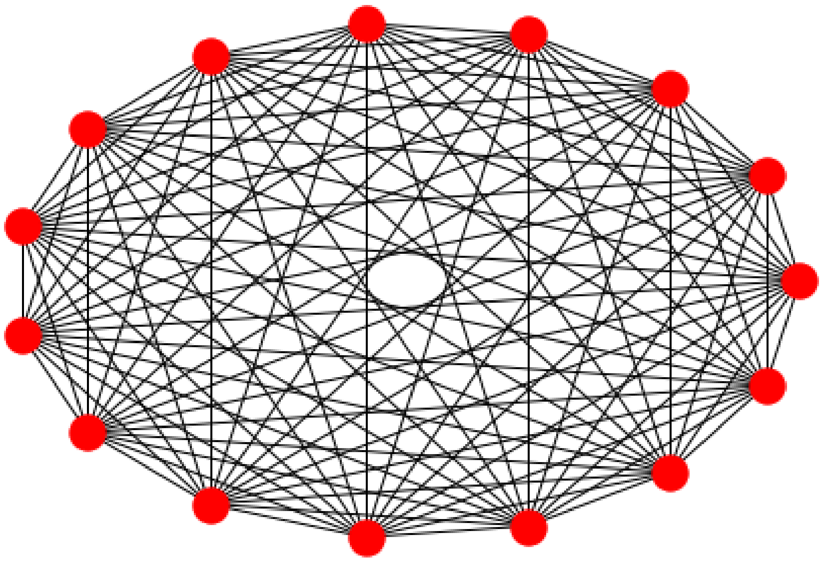

Figure 1.2: **Markowitz's CLA:** A model depiction of the relationship between 15 index companies with each company is a potential substitute of another . The red nodes represent the 15 companies while the 105 edges represents the unique correlations between them.

2. Instability Reduction Techniques

This chapter presents some advanced mathematical methods that attempt to solve problems that arises when attempting to obtain the optimal weights $\mathbf{x}$ that minimize the variance in (1.1.26). The usual estimation matrix used in Markowitz's mean-variance optimization problem is derived by replacing expected returns with their sample means (Michaud, 2014). These sample means, (Michaud, 2014) says, are "suboptimal because they ignore the inherent multivariate nature of the problem" and thus "more powerful statistical estimation techniques are required". We will discuss some of these classical statistical estimation techniques before introducing the more recent machine learning asset allocation model used in this paper.

2.1 Black-Litterman Model

(Black and Litterman, 1992) created an asset allocation model that uses Bayesian approach to infer the expected return of assets in a portfolio. The Bayesian approach is used to infer the probability distribution of expected returns (since they are random themselves) using the Capital Asset Pricing Model (CAPM) and additional views as priors (He and Litterman, 2002). Extensive work has been done on this work and for more literature one can refer to (blacklitterman.org) on the subject matter. We now mathematically formulate the Black-Litterman asset allocation model showing how the optimal weights of the unconstrained optimal portfolio are obtained using matrix notation. For an N-asset market consisting of equities, bonds, currencies and other assets, the random returns of these assets are given by a normal distribution

$$r \sim N(\bar{\mu}, \bar{\mathbf{V}}) \tag{2.1.1}$$

where $\bar{\mu}$ is the mean expected returns vector and $\bar{\mathbf{V}}$ is the covariance matrix. The Bayesian approach using CAPM and additional views as priors is used in inferring the values of the aforementioned parameters, (He and Litterman, 2002) provide a detailed explanation of this. Thus the Markowitz optimization problem then becomes:

$$\text{Maximize } \mathbf{w}^T \mu - \frac{\phi}{2} \mathbf{w}^T \mathbf{V} \mathbf{w} \tag{2.1.2}$$

where ϕ is the risk aversion parameter. The first order condition yields the vector of optimal weights

$$\mathbf{w}^* = \frac{1}{\phi} \mathbf{V}^{-1} \mu \tag{2.1.3}$$

2.2 Ledoit-Wolf Approach

The instability issue of the covariance matrix can be tackled using regularization techniques after formulating the global minimum variance portfolio as a linear regression problem (Colyer, 2018). One of these regularization techniques is called the shrinkage method where the sample covariance matrix $\hat{\mathbf{V}}$ is modified by a new estimate matrix $\tilde{\mathbf{V}}$ in order to take into account some of its uncertainties (Colyer, 2018). The advantage of the sample covariance matrix $\hat{\mathbf{V}}$ is that it's an unbiased estimator of $\mathbf{V}$, but unfortunately converges very slow when the number of assets (N) is large ((Roncalli, 2013) and (Ledoit and Wolf, 2003)). There also exists a highly structured estimator $\hat{\mathbf{\Lambda}}$ that converges much more quickly but it is biased. What (Ledoit and Wolf, 2003) proposed is to compromise and replace the covariance matrix $\mathbf{V}$ by a weighted average of the two estimators $\hat{\mathbf{V}}$ and $\hat{\mathbf{\Lambda}}$ to form a more efficient estimator

$$\tilde{\mathbf{V}}_{\alpha} = \alpha\hat{\mathbf{\Lambda}} + (1-\alpha)\hat{\mathbf{V}} \tag{2.2.1}$$

where α is the shrinkage value that lies between 0 and 1. This technique is called shrinkage because the sample matrix $\hat{\mathbf{V}}$ is shrunk towards the highly structured estimator $\hat{\mathbf{\Lambda}}$ (Ledoit and Wolf, 2003). Thus the statistical problem becomes a matter of estimating the optimal value of α in the (Ledoit and Wolf, 2003) quadratic loss function $L(\alpha)$:

$$L(\alpha) = ||\alpha\hat{\mathbf{\Lambda}} + (1-\alpha)\hat{\mathbf{V}} - \mathbf{V}||^2 \tag{2.2.2}$$

where the minimization problem is

$$\alpha^* = \text{arg min } E[L(\alpha)]. \tag{2.2.3}$$

2.3 Introducing Constraints

Even though the regularization techniques mentioned above improve the robustness of optimized portfolios, it still falls short in fully solving the covariance matrix's sensitivity issue. Regularization techniques work with the covariance matrix while the important information matrix in portfolio optimization lies in its inverse even though both matrices are strongly related in terms of eigendecomposition (Roncalli, 2013). In an unconstrained optimization problem, the optimal weights of the a portfolio can be negative. This may not be viable for some large mutual funds that are legally bound to refrain from using leverage. In addition, the Securities and Exchange Commission in the U.S prevents mutual fund managers to hold more than five percent on a single asset of a portfolio (Francis and Kim, 2013). (Clarke et al., 2002) introduced a conceptual framework along with diagnostic tools that measure the impact of constraints on the value added to a portfolio.

2.4 Hierarchical Risk Parity (HRP)

All the preceding techniques mentioned fail in fully diagnosing the sensitivity inherent in a ill-conditioned covariance matrix. In this section we explore a machine learning method (combination of hierarchical agglomerative clustering algorithm and risk parity asset allocation) called Hierarchical Risk Parity (HRP) created by (Lopez de Prado, 2016) which bypasses the need to invert a positive-definite matrix. What he realized is that covariance/correlation matrix lack the notion of hierarchy. But complex systems such as financial markets do in fact exhibit a structure that can be organized into a hierarchical manner (Simon, 1962). The mathematical blueprint of HRP was created by methods used to model physical systems by physicists like (Mantegna, 1998) and (Tumminello et al.). As mentioned before, HRP is a combination of hierarchical clustering analysis and risk parity but before combining the two entities, let's determine what each independently contributes to the HRP algorithm.

2.4.1 Risk Parity. Since the financial crisis in 2008 the markets have seen a tough time in recovery. This has led to a rise in Risk Parity as the choice of asset allocation. First pioneered by the founder and chairman of Bridgewater Associates back in the 90's for the All Weather Fund, risk parity has previously fared much better to economic shocks in comparison to the classic minimum variance and equally weighted portfolios (Roncalli, 2013) and (IPE Magazine)). The main idea behind the heuristic asset allocation strategy is to build a balanced portfolio such that risk contribution is the same for different asset classes (Roncalli, 2013). A typical Risk Parity portfolio carries assets such as equities, credit, interest rates and commodities in contrast to a classic portfolio which typically holds a 60/40 equity/bond ratio. The problem with the latter portfolio is that its risk is dominated by equity market risk because stock market volatility is much larger than bond market volatility (Chaves et al., 2012) and as such will only perform well when the markets are performing well. Risk Parity, instead, attempts to create a portfolio that performs well no matter the economic weather by allocating the same volatility risk budget to each asset class (Investopedia).

There is no official definition for Risk Parity due to practitioners having varying definitions of risk contribution for asset classes (Chaves et al., 2010). The risk parity is one of three budgeting methods in asset allocation. In this dissertation we use the most widely used Risk Parity approach where the risk budget is weighted by inverse asset class volatility.

(Roncalli, 2013) defines a long-only risk budgeting portfolio as:

$$RC_{w_i} = b_i RC_w \tag{2.4.1}$$

$$b_i > 0 \tag{2.4.2}$$

$$\sum_{i=1}^{N} b_i = 1 \tag{2.4.3}$$

$$w_i \geq 0 \tag{2.4.4}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{2.4.5}$$

where the risk contribution of the i^{th} asset RC_{w_i} is:

$$RC_{w_i} = w_i \frac{(\mathbf{Vw})_i}{\sqrt{\mathbf{w^T V w}}}. \tag{2.4.6}$$

And since we are using inverse variance asset class volatility weighting, the risk budget is given by (Roncalli, 2013):

$$b_i = \frac{\sigma_i^{-2}}{\sum_{i=1}^{N} \sigma_i^{-2}} \tag{2.4.7}$$

The advantage of focusing on Risk parity weighting as opposed to mean-variance optimization is that we circumvent the formulation of expected returns assumptions ((Roncalli, 2013) and (Chaves et al., 2010)). The only input required in Risk Parity is that of the covariance matrix. Since there is no need for the inversion of the covariance matrix, it does not seem ill-conditioned covariance estimates effect the resulting portfolio returns (Chaves et al., 2010).

2.4.2 Hierachical clustering. (Raffinot, 2016) defined hierarchical clustering as the formation of recursive clustering with an end goal of constructing a binary tree of the data that combines similar groups of points successively . The purpose of this cluster analysis is to group or cluster entities using the data's properties. We first need to define distance between clusters that suggests their similarity or dissimilarity and determine when to stop merging these clusters. In this paper we use, possibly, the simplest and most used clustering algorithm which is the linkage-based clustering algorithm. There are three main ways to define this distance (D) between these points, namely:

1. **Single Linkage** clustering (also known as the nearest neighbour algorithm), in which the distance between clusters A and B is defined by the minimum distance between members of the two clusters:

$$D(A,B) := min\{d(x,y) : x \in A, y \in B\}. \tag{2.4.8}$$

2. **Complete Linkage** clustering (also known as furthest neighbour algorithm), in which the distance between clusters A and B is defined by the (yes, you guessed it) maximum distance members of the two clusters:

$$D(A,B) := max\{d(x,y) : x \in A, y \in B\}. \tag{2.4.9}$$

3. **Average Linkage** clustering, in which the distance between clusters A and B is defined by the average distance between all pairs:

$$D(A,B) := \frac{1}{|A||B|} \sum_{x \in A, y \in B} d(x,y). \tag{2.4.10}$$

The sequence of clustering starts off with N isolated data points, then at each step merging occurs between the two most similar groups. This happens recursively until there is a single group. Building a tree is a bottom-up approach, this process is called agglomerative clustering. The single linkage clustering method provides a Minimum Spanning Tree (MST) with N nodes and $N-1$ edges (see Figure 2.1)(Raffinot, 2016). Single Linkage is the linkage criterion of choice in the algorithm used in this paper because it runs in $\mathcal{O}(N^2)$ compared to $\mathcal{O}(N^3)$ of the other two. Visually the merging process is represented by a binary tree, called a dendrogram (see Figure (3.1)) ((Murphy, 1991) and (Ben-david, 2014)).

There are two main advantages that a tree structure has over the complete graph (Lopez de Prado, 2016):

- It is more robust because it has only $N-1$ edges connecting N nodes which which means when a node (or company) is affected, weight redistribution happens accordingly in that branch (or hierarchical level).
- The weights are distributed top-down just as portfolio managers practice in industry.

In the subsequent section we show how HRP uses these advantages as we give a detailed study of the algorithm, developed by (Lopez de Prado, 2016), while observing how it performs on real local data. We unpack the three stages of the algorithm which are tree clustering, quasi-diagonalization and recursive bisection. The only input required is a covariance matrix of the portfolio assets that need not be inverted.

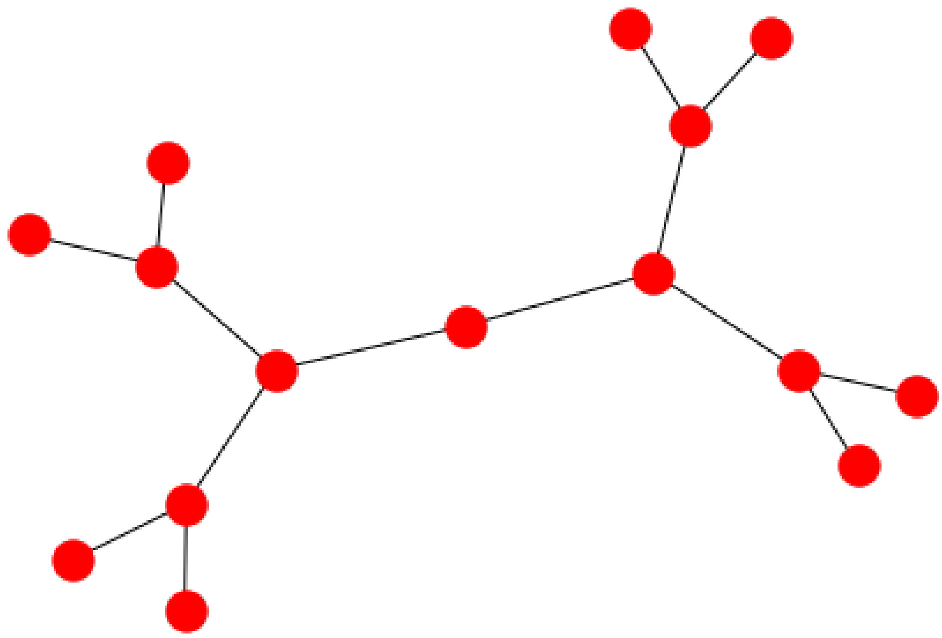

Figure 2.1: **HRP:** A tree visualization of Figure (1.2) that takes advantage of the hierarchical nature of the covariance matrix.

3. Application of HRP

The last decade has seen a high exponential growth of data in the computational realm. This means that there's been a rise in the need for efficient techniques in extracting meaningful data from such big data, particularly in the studies of astronomy, bioinformatics, finance and even social media , to name a few. Machine learning, as (Ben-david, 2014) describe it, refers to the automated detection of meaningful patterns in data. The first thing people think of in the application of machine learning into finance is the prediction of prices, volatility and rates of stock prices. Although the use of machine learning techniques in the prediction of stock prices is the most extensively studied subject in financial mathematics, financial machine learning offers more opportunities to provide insight from financial data. (YouTube-Dr. Marcos Lopez de Prado), in a recent presentation, listed these as some of those opportunities:

- Modelling non-linear relationships in a high-dimensional space
- Analyzing unstructured data (asynchronous, categorical)
- Learning complex patterns (hierarchical, non-parametric)
- Focusing on predictability over parametric adjudication
- Controlling for overfitting (early-stopping, cross-validation).

He also warns that modelling financial series is harder than driving cars and recognizing faces, and that a machine learning algorithm will always find a pattern even if there isn't one. In this dissertation we focus on the second and third opportunities as we explore the use of graph theory and machine learning in modern asset allocation (HRP) as ((Lopez de Prado, 2016) and (Raffinot, 2016)) have done in recent research.

3.1 Dataset

We present an extensive study of the (Lopez de Prado, 2016) HRP algorithm, its mathematical model along with how it works. We feed the algorithm a diversified small set of equities from South Africa's Johannesburg Stock Exchange (JSE) and report results obtained from the algorithm's exploitation of the portfolio's hierarchical structure. A covariance matrix is estimated using $N = 13$ daily closure prices, spanning four years, of local based companies from the beginning of 2014 to 2018. The thirteen randomly chosen companies are described in Table 3.3 below where we indicate each company name with its respective tick symbol along with the different industries/sectors it belongs to. Most of these companies are in the top 40 share index which means this portfolio comprises of equities only. We now introduce the three stages of the HRP algorithm using the above data. The time series data of these companies was extracted, depending on availability, from Yahoo Finance using the python module *pandas_datareader*. The three stages of HRP are tree clustering, quasi-diagonalization and recursive bisection.

Table 3.1: Discription of investigated portfolio equities

Tick	Company name	Sector(s)
1. MRP.JO	Mr Price Group Ltd	General Retailer
2. MTN.JO	MTN Group Ltd	Mobile Telecom
3. VOD.JO	Vodacom Group Ltd	Mobile Telecom
4. SLM.JO	Sanlam Ltd	Life Insurance
5. FSR.JO	FirstRand Ltd	Banking
6. NPN.JO	Naspers Ltd	Tech, Media
7. TBS.JO	Tiger Brands Ltd	Food Producer
8. SHP.JO	Shoprite Holdings Ltd	Food & Drug Retailer
9. KIO.JO	Kumba Iron Ore Ltd	Industry, Metals, Mining
10. REM.JO	Remgro Ltd	Finance, Industry, Investments
11. AME.JO	African Media Entertainment Ltd	Media
12. CFR.JO	Compagnie Financiere Richemont SA	Luxury Goods
13. SOL.JO	Sasol Ltd	Oil, Gas

3.2 Tree Clustering

First order of business is to find a degree of similarity or dissimilarity between a pair of companies A and B, using their time series data. This is done using Pearson's correlation coefficient (Mantegna, 1998):

$$\rho_{A,B} = \frac{\sigma_{A,B}}{\sigma_A \sigma_B} = \frac{\langle X_A X_B \rangle - \langle X_A \rangle \langle X_B \rangle}{\sqrt{(\langle X_A^2 \rangle - \langle X_A \rangle^2)(\langle X_B^2 \rangle - \langle X_B \rangle^2)}} \quad (3.2.1)$$

where the daily returns $X_A = ln(P_A(t)) - ln(P_A(t-1))$ and $P_A(t)$ is the adjusted closure price of stock/company A at the end of day t. An $N \times N$ correlation matrix is formed for all N companies. The correlation coefficient ρ_{AB}, by definition, lies between -1 and 1, where $\rho_{AB} = -1$ describes a pair of oppositely correlated stocks, $\rho_{AB} = 0$ shows no correlation between the pair and $\rho_{AB} = 1$ describes a pair of perfectly correlated stocks (Mantegna, 1998). Now we move on to define a distance metric that will allow us to exploit the hierarchical nature of the portfolio. Unfortunately the correlation coefficient between a pair of stocks can not be used as a distance measure because it does not satisfy the three axioms that define a metric.

3.2.1 Definition. A distance measure $d : (X_i, X_j) \subset B \to \mathbb{R} \in [0,1]$(Lopez de Prado, 2016):

$$\mathbf{D} = d_{i,j} = d[X_i, X_j] = \sqrt{\frac{1}{2}(1 - \rho_{i,j})} \quad [\rho_{i,j} = \rho[X_i, X_j]] \quad (3.2.2)$$

where X are the daily returns and B is the Cartisian product of items in $1, \ldots, i, \ldots, N$. Matrix $\mathbf{D} = d_{i,j}$ fulfills these three axioms:

1. $d_{i,j} = 0$ iff $i = j$)
2. $d_{i,j} = d_{j,i}$ (symmetry)
3. $d_{i,j} \leq d_{i,k} + d_{k,i}$ (sub-additivity)

Axiom 1 is true only if two stocks are perfectly correlated i.e when $\rho = 1$, which coincidentally lies on the diagonal of matrix of $\mathbf{D}$ as shown in Table 3.2. Axiom 2 is due to the fact that the correlation matrix, like the covariance matrix, is symmetric. It has $\rho_{ii} = 1$ along the diagonal and has $\frac{N}{2}(N-1)$ unique correlation coefficients. Equation (3.2.2), as we'll show, is equivalent to the Euclidean distance between two vectors $\mathbf{D_i}$ and $\mathbf{D_j}$ and thus making axiom 3 valid (Mantegna, 1999).

Table 3.2: Distance matrix $\mathbf{D}$ of the 13 JSE Equities

0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.71	0.7	0.7	0.7	0.73	0.72	0.71	0.72	0.71	0.72	0.71	0.73
0.71	0	0.7	0.67	0.67	0.66	0.62	0.67	0.67	0.66	0.61	0.67	0.68
0.7	0.7	0	0.62	0.49	0.54	0.59	0.4	0.51	0.4	0.62	0.51	0.55
0.7	0.67	0.62	0	0.67	0.63	0.64	0.64	0.65	0.63	0.59	0.64	0.67
0.7	0.67	0.49	0.67	0	0.59	0.62	0.5	0.52	0.49	0.65	0.55	0.6
0.73	0.66	0.54	0.63	0.59	0	0.61	0.53	0.59	0.53	0.58	0.61	0.56
0.72	0.62	0.59	0.64	0.62	0.61	0	0.57	0.62	0.58	0.57	0.63	0.64
0.71	0.67	0.4	0.64	0.5	0.53	0.57	0	0.51	0.43	0.59	0.5	0.54
0.72	0.67	0.51	0.65	0.52	0.59	0.62	0.51	0	0.51	0.62	0.53	0.6
0.71	0.66	0.4	0.63	0.49	0.53	0.58	0.43	0.51	0	0.6	0.53	0.57
0.72	0.61	0.62	0.59	0.65	0.58	0.57	0.59	0.62	0.6	0	0.62	0.63
0.71	0.67	0.51	0.64	0.55	0.61	0.63	0.5	0.53	0.53	0.62	0	0.61
0.73	0.68	0.55	0.67	0.6	0.56	0.64	0.54	0.6	0.57	0.63	0.61	0

Now we are able to compute the Euclidean distance between any two column-vectors of $\mathbf{D}$.

3.2.2 Definition. $\tilde{d} : (D_i, D_j) \subset B \to \mathbb{R} \in [0, \sqrt{N}]$:

$$\tilde{d}_{i,j} = \tilde{d}[D_i, D_j] = \sqrt{\sum_{n=1}^{N}(d_{n,i} - d_{n,j})^2} \tag{3.2.3}$$

where $\tilde{d}$ is a distance defined over the entire metric space.

Then we cluster together the pair of columns (i^*, j^*) such that:

$$(i^*, j^*) = argmin(i,j)_{i \neq j}\tilde{d}_{i,j} = u[1] \tag{3.2.4}$$

What follows is to define (see (2.4.8)) the distance between the new cluster $u[1]$ and all the other single, unclustered items using single linkage criterion. This is done to update matrix $\{\tilde{d}_{i,j}\}$. So a distance between an item i of $\tilde{d}$ and the new cluster $u[1]$ is

$$\hat{d}_{i,u[1]} = min[\{\tilde{d}_{i,j}\}_{j \in u[1]}]. \tag{3.2.5}$$

Updating $\{\tilde{d}_{i,j}\}$ is then just a matter of appending it with $\hat{d}_{i,u[1]}$ followed by dropping clustered columns and rows $j \in u[1]$.

Finally we return to clustering a pair of close columns (equation (3.2.4)) and repeat recursively $N-1$ times until we've built the minimum spanning tree (MST). Basically the MST is a step by step building process that links all the elements of the set together in a graph characterized by single linkage between stocks (Mantegna, 1998) which in our case results in the dendogram below in Figure 3.1.

In our distance matrix lies the shortest 12 $(N-1)$ distinct pairwise distances observed from an original matrix of 78 $(\frac{N}{2}(N-1))$. In ascending order these distances are:

Table 3.3: Discription of investigated portfolio equities

Order	Link	Euclidean distance
1.	FSR.JO-SLM.JO	0.5644
2.	REM.JO-FSR.JO	0.5699
3.	MRP.JO-REM.JO	0.7123
4.	TBS.JO-MRP.JO	0.7354
5.	SHP.JO-TBS.JO	0.7362
6.	MTN.JO-SHP.JO	0.7827
7.	VOD.JO-MTN.JO	0.8063
8.	NPN.JO-SOL.JO	0.8095
9.	SOL.JO-VOD.JO	0.8345
10.	KIO.JO-NPN.JO	0.8491
11.	CFR.JO-KIO.JO	0.8797
12.	AME.JO-CFR.JO	1.023

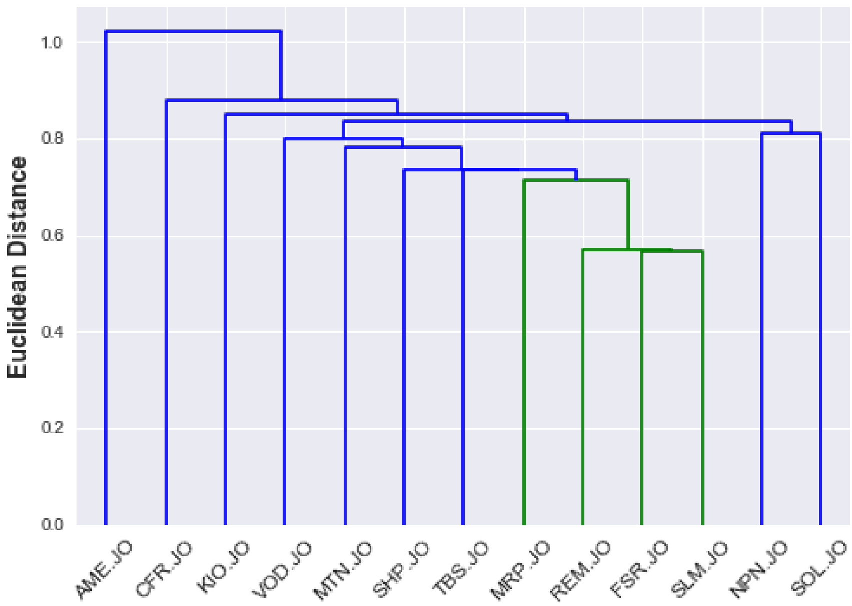

Figure 3.1: Dendogram illustrating the hierarchical structure between assets of the portfolio

These distances in Table 3.3 are then to be used to construct the minimum spanning tree (MST) connecting the the N stocks in our portfolio Mantegna (1999). The first pair of stocks that form a cluster are FSR.JO and SLM.JO with $d = 0.5644$. The next smallest distance lies between stocks REM.JO and FSR.JO with $d = 0.5699$, which means the original cluster grows into REM.JO-FSR.JO-SLM.JO. Continuing to use these minimal distances, this cluster grows until it comprises of 8 stocks (VOD.JO-MTN.JO-SHP.JO-TBS.JO-MRP.JO-REM.JO-FSR.JO-SLM.JO). The next pair of stocks with the shortest distance are NPN.JO-SOL.JO with $d = 0.8095$. At this stage, the MST has grown to have two separate clusters VOD.JO-MTN.JO-SHP.JO-TBS.JO-MRP.JO-REM.JO-FSR.JO-SLM.JO and NPN.JO-SOL.JO. These two clusters are then linked using the shortest distance $d = 0.8345$ between

SOL.JO and VOD.JO. The next stocks to join this larger cluster are KIO.JO, CFR.JO and AME.JO respectively in an ordered fashion until we have a MST graph of a set of the whole N stocks with $N-1$ links between them. This MST now presents the hierarchical organization of the stocks in our portfolio as seen in Figure 3.1.

(Lopez de Prado, 2016) eluded to the fact that stocks could be grouped together using various properties such as liquidity, size, industry, and region. Within a certain similar group, stocks would then compete for allocation. What is clearly observable with the links between the stocks and the dendogram illustration 3.1 is that some companies in similar industries in some cases are grouped closer together. The branch REM.JO-FSR.JO-SLM.JO, all financial institutions, is an example of this. The close proximity of the telecom companies VOD.JO-MTN.JO or food retailers SHP.JO-TBS.JO further emphasizes this point.

3.3 Quasi-diagonalization

In this stage we want a correlation matrix that exhibits a useful property of similar stocks placed together and dissimilar stocks are placed far apart using the linkage matrix $\{\hat{d}_{i,u}\}$ that results from previous stage as we've illustrated(Lopez de Prado, 2016). The reason we want to diagonalize the correlation matrix is because the inverse-variance risk parity allocation approach we wish to use is optimal for it and we will prove this after going through how the algorithm works. In the previous stage we ended up with an $(N-1) \times 4$ linkage matrix with structure

$$Y = \{(y_{m,1}, y_{m,2}, y_{m,3}, y_{m,4})\}_{m=1,\dots,N-1} \tag{3.3.1}$$

where $(y_{m,1}, y_{m,2})$ are the stocks in the cluster, $y_{m,3}$ is the minimum distance between stocks in $y_{m,1}$ and stocks in $y_{m,2}$ and $y_{m,4}$ is the number of stocks in cluster m. Each row in this linkage matrix merges two branches of the MST into one. In the quasi-diagonalization algorithm (Lopez de Prado, 2016) replaces in $(y_{N-1,1}, y_{N-1,2})$ with their constituents recursively until no clusters remain. The algorithm results in the covariance matrix shown in Figure 3.3.

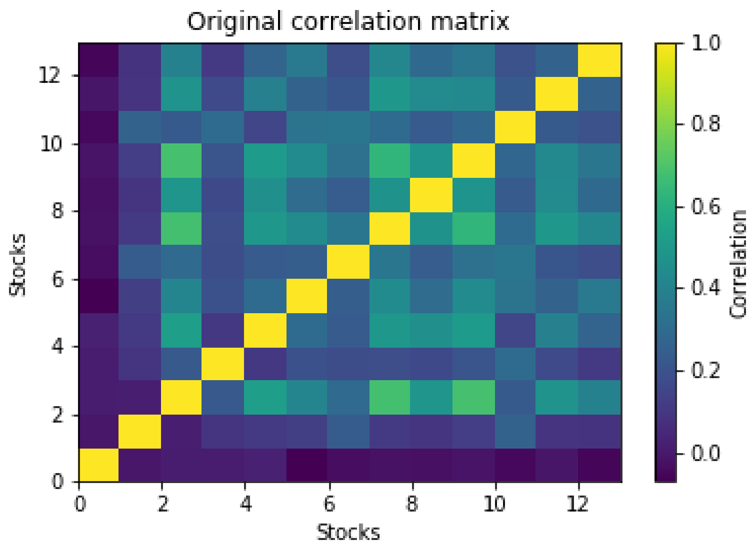

Figure 3.2: Heatmap of original correlation matrix

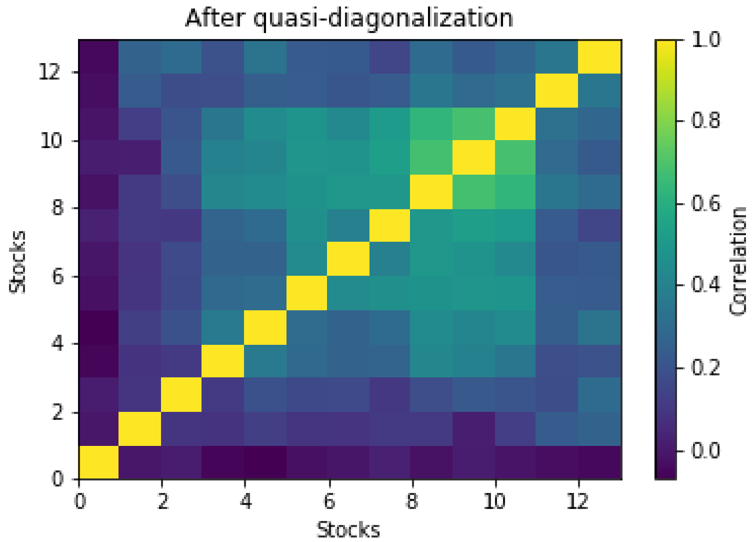

Figure 3.3: Heatmap of correlation matrix post quasi-diagonalization

The original correlation matrix was computed using the distance matrix $\mathbf{D}$ of the 13 JSE stocks as shown in the previous section. As before mentioned the diagonalized matrix 3.3 is optimal for the inverse-variance risk parity allocation approach. To show this, let's consider the standard quadratic optimization problem of size N (Lopez de Prado, 2016),

$$\text{Min}_{\mathbf{w}} \quad \mathbf{w}^T\mathbf{V}\mathbf{w} \tag{3.3.2}$$

$$\text{subject to: } \mathbf{w}^T\mathbf{1} = 1 \tag{3.3.3}$$

which has solution

$$\mathbf{w} = \frac{\mathbf{V}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1}} \tag{3.3.4}$$

for a minimum variance portfolio. If $\mathbf{V}$ is diagonal then

$$w_N = \frac{\mathbf{V}_{N,N}^{-1}}{\sum_{i=1}^N \mathbf{V}_{i,i}^{-1}} \tag{3.3.5}$$

In a two-asset $(N = 2)$ example,

$$w_1 = \frac{\mathbf{V}_{1,1}^{-1}}{\mathbf{V}_{1,1}^{-1} + \mathbf{V}_{2,2}^{-1}} \tag{3.3.6}$$

$$= 1 - \frac{\mathbf{V}_{1,1}}{\mathbf{V}_{1,1} + \mathbf{V}_{2,2}} \tag{3.3.7}$$

$$\text{and } w_2 = 1 - w_1. \tag{3.3.8}$$

This is how the weight is split in the subsequent stage.

3.4 Recursive Bisection

The goal of recursive bisection is to find the optimal weights which should be allocated to our portfolio using inverse-variance risk parity. For this we use the quasi-diagonal matrix (Lopez de Prado, 2016):

1. "Bottom-up, to define the variance of a continuous subset as the variance of an inverse-variance allocation;
2. top-down, to split allocations between adjacent subsets in inverse proportion to their aggregated variances."

The algorithm is as follows:

Algorithm 1 Recursive Bisection

1: **Initialize:**
2: $L \leftarrow \{L_0\}$ with $L_0 = \{n\}_{n=1,\dots,N}$ (Set list of items)
3: $w_n \leftarrow 1 \quad \forall n = 1,\dots,N$ (Assign unit weight to all items)
4: **while** $|L_i|! = 1, \forall L_i$ in L **do**
5: $L_i^{(1)} \cup L_i^{(2)} = L_i$, where $|L_i^{(1)}| = \text{int}[\frac{1}{2}|L_i^{(2)}|]$
6: $\tilde{w}_i^{(j)} = \text{diag}[V_i^{(j)}]^{-1}\frac{1}{\text{trace}[\text{diag}[V_i^{(j)}]^{-1}]}$ where $V_i^{(j)}$ is the covariance matrix
7: $\tilde{V}_i^{(j)} \equiv (\tilde{w}_i^{(j)})^T V_i^{(j)} \tilde{w}_i^{(j)}$ between $L_i^{(j)}$, $j = 1,2$
8: $\alpha_i = 1 - \frac{\tilde{V}_i^{(1)}}{\tilde{V}_i^{(1)} + \tilde{V}_i^{(2)}} \quad [0 \le \alpha_i \le 1]$
9: $\alpha_i * w_n \quad \forall n \in L_i^{(1)}$
10: $(1-\alpha_i) * w_n \quad \forall n \in L_i^{(2)}$

Lines 7 takes advantage of the quasi-diagonalized matrix bottom-up. This line is the definition of the variance of of the partition $L_i^{(j)}$ using the inverse weightings $\tilde{w}_i^{(j)}$ from line 6. Line 8 on the other hand takes advantage of the quasi-diagonalized matrix top-down. The weight is split in inverse proportion to the cluster's variance. The weights $w_i \ \forall i = 1,\dots,N$ are guaranteed to lie in the interval $[0,1]$ and the constraint $\sum_{i=1}^{N} = 1$ is satisfied.

These three stages together form the HRP algorithm. HRP solves the allocation problem in deterministic logarithmic time ($T(N) = \mathcal{O}(log_2 N)$) (Lopez de Prado, 2016).

4. Results and Discussion

4.1 Results

Using the aforedescribed Hierarchical Risk Parity algorithm, we have been able to construct a risk parity portfolio with each asset having an associated percentage allocation. Figure 4.1 represents this division of asset allocation. If, for example, an investor had a capital of say ZAR$100\ 000$ to invest in such a portfolio then HRP advises an allocation of ZAR2500 to be seeded in Mr Price stock, ZAR3000 goes towards Kumba Iron Ore, ZAR3500 to Sasol, Naspers, Sanlam, FirstRand, and MTN respectively, ZAR5500 in Shoprite, ZAR9500 in Tiger Brands, ZAR$18\ 500$ towards African Media Entertainment, ZAR$17\ 500$ towards Compangnie Financiere Richemont and the largest share of the investment goes towards Vodacom stocks with an amount of ZAR$20\ 500$. The constraint $\sum_{i=1}^{N} w_i = 1$ is not in actual fact fully satisfied, we shall discuss possible reasons for this and how we've opted to address the issue in the subsequent section. Also these values are well rounded due to rounding off towards the nearest half-a-percentage along with correcting for estimation errors.

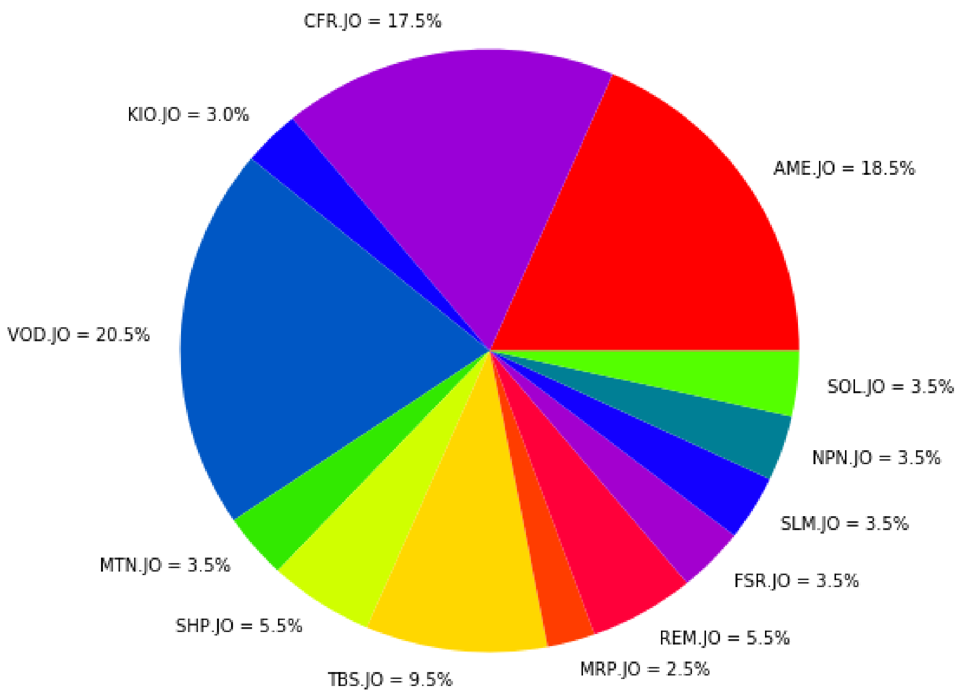

Figure 4.1: Assets allocated to a portfolio problem consisting of 13 JSE stocks using HRP

4.2 Discussion

We have managed to successfully create an optimal risk parity portfolio but, as mentioned above, it was not all that straightforward. The constraint $\sum_{i=1}^{N} w_i = 1$ was not entirely satisfied, in the first round of recursive bisection $\sum_{i=1}^{N} w_i$ amounted to a total of 0.9625. This may be due to various reasons, some which may inherently rise from error propagation within python. Other reasons may be from the weaknesses of the code, and so better assumptions and constraints need to be added to it. A combination of these two scenarios is also a possibility. This value continued to decrease in our pursuit of the individual asset allocations within each cluster by repeated recursive bisection. By the end of the iterations the total sum of individual weights amounted to $\sum_{i=1}^{N} w_i = 0.785$. As the portfolio managers we remedied the situation by first rounding up each asset allocation to the nearest percentage. This lead to a total weight of $\sum_{i=1}^{N} w_i = 0.8$. We then followed by equally spreading the outstanding weight of $1 - w_i$ across all the assets and then finally rounding off to the nearest half percent resulting in our presented allocations. This remedy does largely rely on the manager's heuristics and discretion and so is not possibly the best solution to the problem. Other options one may consider are:

- to strengthen the code's efficiency by opting for different methods of finding weights just as in (Chaves et al., 2012),
- find a different way to distribute the outstanding weight by imposing different constraints

to name a few. A question we must ask is would this HRP portfolio have a lower risk than than Markowitz CLA's portfolio containing the same assets. In-sample, the answer is probably no. But as (Lopez de Prado, 2016) so eloquently put it (and later showed), "the portfolio with minimum variance in-sample is not necessarily the one with minimum variance out-of-sample". What we seek is a portfolio that reacts well to sudden real time market shocks and (Lopez de Prado, 2016), using Monte carlo simulations, along with (Raffinot, 2016), using varying comparison measures, have been able to show that hierarchical clustering based portfolios do, in fact, achieve statistically better risk-adjusted performances. Mathematical simulations done by (Carrasco and Noumon, 2011) also show that even classical regularization methods have a higher out-of-sample performance than the sample based Markowitz portfolios.

5. Conclusion

The financial world has come a long way since Harry Markowitz first developed a model to assist portfolio managers in the asset allocation problem. Yet portfolio construction continues to be the most recurrent problem in finance. As we've showed in this paper, the problem with quadratic optimizers like Markowitz's Critical Line Algorithm is that they require the inversion of a positive-definite covariance matrix. This inversion, especially when the matrix's size grows, is prone to large errors when the covariance matrix is numerically ill-conditioned. Furthermore, a matrix of size N is modelled by a complete graph with $(\frac{N}{2}(N-1))$ edges (Lopez de Prado, 2016). With this kind of structure every node is a potential substitution of another and thus small estimation errors lead to incorrect solutions (Lopez de Prado, 2016). What Hierarchical Risk Parity manages to achieve in working around this stability issue is to use all the information in the covariance matrix without the need of its inversion or positive-definiteness. The outcome model is a tree with with N nodes and $N-1$ links which allow weights to rebalance among peers at the tree's branches.

In this paper we've managed to show how HRP exploits the hierarchical structure within covariance matrices. We've also shown HRP's flexibility and robustness while constructing an optimal portfolio. We can therefore conclude that HRP is able to compute a portfolio on an ill-generated covariance matrix.

5.1 Future Work

In this paper we've introduced and tested (Lopez de Prado, 2016) Hierarchical Risk Parity algorithm. Its optimal robustness may be dependent on other clustering methods than just single linkage which means other distance measures such as average linkage and complete linkage can be investigated and compared like (Raffinot, 2016) did. Other Risk Budgeting weighting strategies along with various weight constraints are also possible. Even the employment of Black-Litterman Bayesian approach might be a fruitful adventure. Unfortunately the out-of-sample performances using statistical simulations such as Monte Carlo or otherwise real data testing was never used in this paper which leaves room for better evaluations of HRP. With quantum computing slowly on the rise then the HRP could be tested using an extremely dimensional covariance matrix input.

Acknowledgements

"*Gratitude turns what we have into [more than] enough*." - Anonymous

Firstly, I'd like to express my sincere gratitude to my supervisor Prof. Phillip Mashele, along with David Attipoe, Rock Koffi and Abdo Degoot in their combined efforts and hard work to making this an unforgettable life experience for me. I have acquired vast knowledge and skills from you all during this time, thank you for your guidance.

A great deal of appreciation is also owed to the strong friendships forged along the way, we have undoubtably become family thanks, in large part, to *DHill*. Special thanks goes out to *The Company* for their lifelong brotherhood, support and continuous inspiration on this long journey.

No words can express the deep respect and credit that I give my family for their spiritual support and love. I'm especially thankful to God for the four strongest women He's put in my life, oMavalela, oTshawana, I dedicate this to you.

And lastly I'd like to acknowledge the great work and opportunity provided to me by the AIMS community. Thank you for facilitating the African dream for the young African dreamers of this fine continent. The floor staff, to you much is owed. Thank you all, not only for feeding me, but also being family away from home that has egged me on. For being a constant reminder of the community from which I was birthed and the reason I am here.

References

Aste, T., Matteo, T. D., Strand, T., Street, G., and Centre, S. R. Risk diversification: a study of persistence with a filtered correlation-network approach. 2014.

Ben-david, S. *Understanding Machine Learning: From Theory to Algorithms*. 2014. ISBN 9781107057135.

Black, F. and Litterman, R. Global Portfolio Optimization. *Financial Analysts Journal*, 48(5):28–43, 1992. ISSN 0015-198X. doi: 10.2469/faj.v48.n5.28. URL http://www.cfapubs.org/doi/full/10.2469/faj.v48.n5.28.

blacklitterman.org. Black-literman papers. blacklitterman organization, https://blacklitterman.org/papers.html, Accessed 4 October 2018.

Carrasco, M. and Noumon, N. Optimal portfolio selection using regularization. 2011. URL http://www.admissions.american.edu/cas/economics/info-metrics/pdf/upload/Carrasco-Nov-2011-submission.pdf.

Chaves, D., Hsu, J., Li, F., and Shakernia, O. Risk Parity Portfolio vs. Other Asset Allocation Heuristic Portfolios *. pages 1–15, 2010.

Chaves, D. B., Li, F., Hsu, J. C., and Shakernia, O. Efficient Algorithms for Computing Risk Parity Portfolio Weights. (July), 2012.

Clarke, R., De Silva, H., and Thorley, S. Portfolio Constraints and the Fundamental Law of Active Management. *Financial Analysts Journal*, 58(5):48–66, 2002. ISSN 0015198X. doi: 10.2469/faj.v58.n5.2468.

Colyer, A. M. Alexandra Malindi Colyer The Markowitz mean-variance portfolio theory posits that the optimal portfolio weights can be chosen based off an efficient tradeoff between profit modeled as the mean and risk measured as the variance-covariance matrix . These va. 266(12):1–5, 2018. doi: 10.2139/ssrn.385760.3.

Davey, M. *Error-correction using Low-Density Parity-Check Codes*. Phd, University of Cambridge, 1999.

Demey, P., Maillard, S., and Roncalli, T. Risk-Based Indexation. *Ssrn*, 2010. ISSN 1556-5068. doi: 10.2139/ssrn.1582998.

Demiguel, V., Garlappi, L., and Uppal, R. Optimal Versus Naive Diversification: How Inefficient is the 1/. 2007. doi: 10.1093/rfs/hhm075.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. *Management Science*, 55(5): 798–812, 2009. ISSN 0025-1909. doi: 10.1287/mnsc.1080.0986. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1080.0986.

Economics, F. and Company, N.-h. P. No Title. 8:323–361, 1980.

Francis, J. C. and Kim, D. *"Modern Portfolio Theory"*. John Wiley Sons, Inc, 2013.

He, G. and Litterman, R. The Intuition Behind Black-Litterman Model Portfolios. *Ssrn*, 2002. ISSN 1556-5068. doi: 10.2139/ssrn.334304.

Investopedia. Risk parity. Investopedia, the Free Financial Encyclopedia, https://www.investopedia.com/terms/r/risk-parity.asp, Accessed 1 October 2018.

IPE Magazine. Risk parity-the truly balanced portfolio. Intelligence on European Pensions and Institutional Investment, https://www.ipe.com/risk-parity-the-truly-balanced-portfolio/www.ipe.com/risk-parity-the-truly-balanced-portfolio/45968.fullarticle, Accessed 2 October 2018.

Jordaan, K. Financial mathematics, 2018. Lecture Notes.

Ledoit, O. and Wolf, M. Honey, I Shrunk the Sample Covariance Matrix. *Ssrn*, pages 1–22, 2003. ISSN 0095-4918. doi: 10.2139/ssrn.433840.

Lopez de Prado, M. Building Diversified Portfolios That Outperform Out-of-Sample (Presentation Slides). *Ssrn*, 2016. ISSN 0095-4918. doi: 10.2139/ssrn.2713516.

Mankert, C. *The Black-Litterman Model - mathematical and behavioral finance approaches towards its use in practice*. 2006. ISBN 9171783970. doi: 10.3905/jpm.2007.674791.

Mantegna, R. N. Hierarchical Structure in Financial Markets. 1998. doi: 10.1007/s100510050929. URL http://arxiv.org/abs/cond-mat/9802256%0Ahttp://dx.doi.org/10.1007/s100510050929.

Mantegna, R. N. Hierarchical structure in financial markets. 197:193–194, 1999.

Markowitz, H. Portfolio Selection. *The Journal of Finance*, 7(1):77–91, 1952. ISSN 15406261. doi: 10.1111/j.1540-6261.1952.tb01525.x.

Medium. Markowitz's efficient frontier in python. Medium Scientific blog, https://medium.com/python-data/effient-frontier-in-python-34b0c3043314, Accessed 1 October 2018.

Merton, R. C. On estimating the expected return on the market: An exploratory investigation. *Topics in Catalysis*, 8(4):323–361, 1980. ISSN 0304405x. doi: 10.1016/0304-405X(80)90007-0.

Meucci, A. *"Risk and Asset Allocation"*. Springer Finance, 2005.

Michaud, R. O. The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Ssrn*, (January 1989), 2014. ISSN 0015-198X. doi: 10.2139/ssrn.2387669.

Murphy, K. P. *A probabilistic perspective*. MIT City Press, 1991. ISBN 9780262018029. doi: 10.1038/217994a0. URL http://link.springer.com/chapter/10.1007/978-94-011-3532-0_2.

Musmeci, N., Aste, T., and Di Matteo, T. Risk diversification: a study of persistence with a filtered correlation-network approach. 2014. ISSN 20557795. doi: 10.21314/JNTF.2015.005. URL http://arxiv.org/abs/1410.5621.

Norms, P. Management science. 55(5):798–812, 2009. doi: 10.1287/mnsc.1080.0986.

Raffinot, T. Hierarchical Clustering Based Asset Allocation. *Ssrn*, (September):1–22, 2016. doi: 10.2139/ssrn.2840729.

Rappaport, A. The Economics of Short-Term Performance Obsession. 61(3):65–79, 2005.

Roncalli, T. *"Introduction to Risk Parity and Budgeting"*. Chapman Hall/CRC Financial Mathematics series, 2013.

Sharpe, W. F. Capital asset prices: A theroy of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964. ISSN 00221082. doi: 10.2307/2329297.

Simon, H. A. The Architecture of Complexity. 106(6):467–482, 1962.

Tobin, J. Liquidity Preference as Behavior Towards Risk. *The Review of Economic Studies*, 25 (2):65, 1958. ISSN 00346527. doi: 10.2307/2296205. URL https://academic.oup.com/restud/article-lookup/doi/10.2307/2296205.

Tumminello, M., Lillo, F., and Mantegna, R. N. Correlation, hierarchies, and networks in financial markets. (December 2008).

YouTube-Dr. Marcos Lopez de Prado. Ten applications of machine learning|dr. marcos lopez de prado. YouTube, https://www.youtube.com/watch?v=pRurbL9-Lus&t=141s, Accessed 4 September 2018.