Department of Economics- FEA/USP

Saving Markowitz: A Risk Parity approach based on the Cauchy Interlacing Theorem

Fernando Fernandes

Rodrigo De-Losso

Rogerio Oliveira

Angelo J. D. Soto

Pedro Delano Cavalcanti

Gabriel M. S. Campos

WORKING PAPER SERIES Nº 2020-13

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2020-13

Saving Markowitz: A Risk Parity approach based on the Cauchy Interlacing Theorem

Fernando Fernandes (fernando.fernandes@boanota.credit)

Rodrigo De-Losso (delosso@usp.br)

Rogerio Oliveira (roliveira@constanciainvest.com.br)

Angelo J. D. Soto (diazsoto@ime.usp.br)

Pedro Delano Cavalcanti (pedelano@gmail.com)

Gabriel M. S. Campos (gabrielmscampos@gmail.com)

Research Group: NEFIN

Abstract

It is well known that Markowitz Portfolio Optimization often leads to unreasonable and unbalanced portfolios that are optimal in-sample but perform very poorly out-of-sample. There is a strong relationship between these poor returns and the fact that covariance matrices that are used within the Markowitz framework are degenerated and ill-posed, leading to unstable results by inverting them, as a consequence of very small eigenvalues.

In this paper we circumvent this problem in two steps: the enhancement of traditional risk parity techniques, which consider only volatility, aiming to avoid matrix inversions (including the widespread Naive Risk Parity model) within the Markowitz framework; the preservation of the correlation structure, as much as possible, aiming to isolate a "healthy" portion of the correlation matrix, that can be inverted without generating unstable and risky portfolios, aiming to rescue the original Markowitz framework, by means of using the Cauchy Interlacing Theorem.

Using Brazilian and US market data, we show that the discussed framework enables one to build portfolios that outperform the traditional and the newest risk parity techniques.

Keywords: Markowitz, Cauchy Interlacing Theorem, NRP, CIRP.

JEL Codes: C38, C61, G11, G17.

Saving Markowitz: A Risk Parity approach based on the Cauchy Interlacing Theorem

Fernando Fernandes Neto
Rogerio de Deus Oliveira
Rodrigo de Losso da Silveira Bueno
Angelo Jonathan Diaz Soto
Pedro Delano Cavalcanti
Gabriel Moreira da Silva Campos

São Paulo

May 2020

Abstract

It is well known that Markowitz Portfolio Optimization often leads to unreasonable and unbalanced portfolios that are optimal in-sample but perform very poorly out-of-sample. There is a strong relationship between these poor returns and the fact that covariance matrices that are used within the Markowitz framework are degenerated and ill-posed, leading to unstable results by inverting them, as a consequence of very small eigenvalues.

In this paper we circumvent this problem in two steps: the enhancement of traditional risk parity techniques, which consider only volatility, aiming to avoid matrix inversions (including the widespread Naive Risk Parity model) within the Markowitz framework; the preservation of the correlation structure, as much as possible, aiming to isolate a "healthy" portion of the correlation matrix, that can be inverted without generating unstable and risky portfolios, aiming to rescue the original Markowitz framework, by means of using the Cauchy Interlacing Theorem.

Using Brazilian and US market data, we show that the discussed framework enables one to build portfolios that outperform the traditional and the newest risk parity techniques.

1. Introduction

Among portfolio managers and academics, it is not new that optimal mean variance portfolios are often extreme and perform poorly out of sample, sometimes even worse than Naive equal weight allocations, see [1].

A simple solution to build diversified portfolios that perform decently out of sample, and which is currently used by various practitioners to manage hundreds of billions of dollars, is called naive risk parity (NRP), which assigns the weight of each asset as proportional to the inverse of the volatility.

It is easy to show that NRP allocation is consistent with Markowitz optimality when all Sharpe ratios and correlations are the same. These implicit assumptions in NRP avoid the two main problems associated with Markowitz: inaccuracy of forecasted returns and inverting ill-behaved correlation matrices that maybe unstable to invert.

Our contribution in this paper is two-fold. First we significantly improve the out of sample performance of NRP but averaging it with a novelty orthogonal portfolio that we name asymptotically global minimum variance risk parity (MVRP). Because NPR maybe considered an "aggressive" portfolio (lemma 3.1), we call this improved allocation balanced risk parity (BRP).

The second contribution is to pursue an innovative approach to circumvent problematic sample correlation matrices. Using Cauchy Interlacing theorem we find and preserve the integrity of a large portion of the sample correlation matrix which is "well behave", in a sense to be made precise, and allow shrinkage only on the "problematic" part of the correlation structure.

We build a Markowitz optimal portfolio with this procedure, name it Cauchy Interlacing Risk Parity (CIRP), and show that it has substantially improved out-of-sample performance over NRP, MVRP, BRP and other diversified allocations that have been proposed as competitors in the literature.

To move towards more realistic assumptions about Sharpes and correlations, we estimate expected returns using Bayes-Stein shrinkage (see [2]), where the sample means are interpolated towards a common value. However, this technique also requires the inversion of the covariance matrix, just like portfolio optimization does, which enhances the importance of avoiding ill-behaved correlation matrices.

The literature suggests another shrinkage technique (see [3]) where the correlation matrix actually used is the interpolation between the sample matrix and equal correlations for all assets. These authors provide a formula for the optimal interpolation size. It is also important to notice that NRP may be seen as an extreme version of this matrix shrinkage, where most information contained in the correlation structure is fully collapsed to its mean and, hence, a lot of diversification opportunities are lost.

2. Dataset Description and the Backtesting Procedure

In this work, for the sake of simplicity, but yet, aiming a minimal level of completeness, we have used two datasets to study the effects of risk parity approaches that are discussed along this paper: US market data and Brazilian market data.

For the Brazilian market data, we have selected a window span that comprises 2002 until 2020, where a moving window of 60 months is used to compute all calculations required to backtest the performance of the procedures aforementioned, using a full dataset with all available stocks prices ($>$ 500 companies). Hence, we include in the list of investible assets all the 65 stocks in the Bovespa index but also several off index as well taking into account liquidity thresholds. As of March 2020, the number of eligible investible stocks was 200, but this number fluctuated with liquidity within our historical dataset.

For the US market data, we have built a dataset comprising the largest US companies data, within a frame that spans from 1994 until 2020, encompassing more than 1200 public companies. As a possible major drawback of this second dataset as a result of its size, we have chosen to shrink the dataset size, by picking up the top thirty percent companies with the highest returns in the last 60 months, also within a moving window of the same length. Hence, we avoid any computational problem regarding processing time, given the fact that, for this first paper, we have not yet optimized our computational procedure for a very large matrix, such as a potential correlation matrix with 1200 columns / rows.

Furthermore, all companies within each window frame of 60 months which had more than ten percent of prices that had an "not available" were excluded at each backtest epoch. This approach aims to fix two potential issues: avoid to introduce noisy observations on the covariance matrix as a consequence of substituting not available data by zeroes; and fixing for illiquidity effects (it is straightforward to notice that assets which are more often tradable, are more liquid).

All portfolios we build and strategies we develop in this paper are only rebalanced on the last business day of each month and are held constant until the next rebalance day. Some portfolios will be long only but others will be long-short, but the aggregate stock allocations always need to add up to 100%.

It is also important to mention that, in the case of Brazilian data, Ibovespa is a very concentrated index with the top 5 companies adding to almost 50% of its total value. We aim to build diversified portfolios with optimality rational using this broader set of stocks with the goal of having better risk adjusted performance than the index.

When covariances are known but there is uncertainty about expected returns, previous work (see [4]) have shown that the optimal portfolio under the most conservative set of expected returns within the uncertainty bounds converges to the global minimum variance portfolio when the level of uncertainty is arbitrarily high.

As the table below shows, the underperformance of the Markowitz Portfolio that uses sample data for both expected returns and covariances is dismal, with yearly average historical returning -84.76% versus 7.85% for Ibovespa. This is unfortunate but typical. Many other works have verified that Markowitz portfolios, which are optimal in-sample, perform very poorly out-of-sample.

	Ibovespa	Equal Weights	Markowitz	
			Equal Expect Returns (Minimum Variance Portfolio)	Sample Historical
Historical yr. Avg. Return (i)	7,85%	13,37%	12,08%	-84,76%
Volatility (ii)	23,05%	22,82%	127,61%	808,01%
Downside Volatility (iii)	16,40%	17,73%	113,96%	1115,26%
Max Drawdown (iv)	-49,58%	-41,98%	-209%	-2600%
Sharpe Ratio (i)/(ii)	0,34	0,59	0,09	-0,10
Sortino Ratio (i)/(iii)	0,48	0,75	0,11	-0,08
Calmar Ratio -(i)/(iv)	6,32	0,32	0,06	-0,03
Recovery Time (yrs) -(iv)/(i)	0,16	3,13	17,31	-30,67
Leverage (Gross Exposure)	100%	100%	2367%	6919%

The historical annual average return of the Markowitz global minimum variance portfolio (same expected returns) is 12.08%, better than the Ibovespa Index, but still below the naïve $1/N$ equal weight portfolio. And the typical risk levels as measure by volatility, downside volatility and maximum drawdown are unacceptable.

This optimal portfolio is also very leveraged, with gross exposure (the absolute sum of longs and short) equal to 2300%. In practice, this kind of leverage would be unattainable because of amount of collateral that would be required. The typical risk measure (volatility, downside volatility and maximum drawdown) are also worse than the equal weight portfolio.

Our goal in this paper is to rescue Markowitz and build optimal, diversified and balanced portfolios with good risk-reward performance in the backtest procedure.

Our first step will be to revise, reinterpret within the optimality framework and significantly improve risk parity allocations with a novelty addition. This already gets us a long way and delivers portfolios with better risk-rewards profile.

However, risk parity, even our improved version, ignores or oversimplifies the correlation structure, wasting diversification opportunities. This is consciously done in order to avoid the risk of inverting the correlation matrix, which are often quasi-ill-defined and may generate very unbalanced and unreasonable optimal portfolios.

We use the Cauchy Interlacing Theorem to make encouraging progress in dealing with the risk of facing complicated correlation matrices. The theorem allows us to find and use the "good" portion of the correlation matrix to construct better, diversified, optimal portfolio.

3. Improving Risk Parity Allocations

It is well known that optimal mean variance portfolios are often extreme and perform poorly out of sample, sometimes even worst than Naive equal weight allocations, see [1].

Naive Risk Parity (NRP) is a well established methodology that is currently adopted in the financial industry to build diversified portfolios and help practitioners managed hundreds of billions of dollars in various markets. NRP simply assigns weights that are proportional to the inverse of the volatility of each asset, that is

$$W_i^{NRP} = \frac{\frac{1}{\sigma_i}}{\sum\limits_i \left[\frac{1}{\sigma_i}\right]}.$$

In the appendix, we prove the following Lemma that lends some rationality to NRP allocations:

Lemma 3.1. *If all assets have the same Sharpe ratio and all correlations are the same, then NRP allocation is the portfolio that achieves the maximum Sharpe Ratio.*

The assumptions in lemma 3.1 about the Sharpe Ratios allows NRP to avoid the first major pitfall in Markowitz Optimizations: forecasting expected returns and dealing with challenging correlation matrices.

Nonetheless, lemma 3.1 suggests that NRP is an "aggressive" portfolio and to counter balance it, we define an alternative diversified portfolio with conservative bias, which we name *asymptotic global minimum variance risk parity* (MVRP) portfolio.

Define de average of inverse volatilities as $\frac{1}{\acute{\sigma}_N} = \frac{\sum\limits_i \left[\frac{1}{\sigma_i}\right]}{N}$, and the weights of this new portfolio as:

$$W_i^{NRP} = \frac{\frac{1}{\sigma_i} \times \left(\frac{1}{\sigma_i} - \frac{1}{\acute{\sigma}_N}\right)}{\sum\limits_i \left[\frac{1}{\sigma_i} \times \left(\frac{1}{\sigma_i} - \frac{1}{\acute{\sigma}_N}\right)\right]}$$

In the appendix we prove the following lemma:

Lemma 3.2. *If all correlations are the same and different from zero, and σ_N converges as you increase N, then MVRP allocation converges to the global minimum variance portfolio. If the correlations are all zero then the Global min variance portfolio is given by weights that are proportional to the inverse of the variance of each asset.*

To the best of our knowledge MVRP has been completely overlooked in the literature, by academics and practitioners.

We think it makes a lot of sense to consider MRRP in this context not just because it is conservative and NPR is aggressive, but also because NPR's implicit assumption of equal Sharpe ratios for all assets contrasts with the usual high uncertainty about estimation of expected returns.

To address parameter uncertainty under the Markowitz framework, [4] solved a minmax optimization and obtain an analytic solution for the most conservative optimal portfolio when all possible expected returns within some bounds of parameter uncertainty are considered. They prove that this optimal minimax portfolio gets arbitrarily close to the global minimum variance portfolio as the error uncertainty growths.

Lemma 3.3 below, which we prove in the appendix, lends strong support to the compelling idea of combining NPR and MVRP to improve risk-reward profile.

Lemma 3.3. *If all correlations are the same the NRP and MVRP portfolios are orthogonal.*

We define a balanced risk parity (BRP) allocation as the simple average between NRP and MVRP with 50% weights assigned to them. More refined or optimized weighting scheme might be possible.

Using our dataset, the out-of-sample backtest results in the table below shows that a robust 50% weighs generate BRP portfolios with risk rewards measures that outperform NRP in risk adjusted terms, with better Sharpe ratio, Sortino and Calmar ratios, for the Brazilian market data.

	NRP	MVRP	BRP	Equal Weights
Sortino	0,65	0,48	0,79	0,75
Downside Risk monthly	4,84%	4,82%	3,74%	5,13%
Downside Risk yearly	16,77%	16,69%	12,96%	17,73%
Sharpe	0,54	0,44	0,66	0,59
Volatility monthly	5,78%	5,28%	4,48%	6,59%
Volatility yearly	20,01%	18,30%	15,52%	22,82%
Returns monthly	0,87%	0,64%	0,81%	1,05%
Returns yearly	10,36%	8,00%	10,17%	13,37%
Max Drawdown	38,67%	46,25%	29,32%	40,71%
Returns March/2020	-30,80%	-16,89%	-23,85%	-32,95%

However, while repeating the same exercise for the US data, these results do not hold, suggesting that the importance of correlation is greater for this market:

	NRP	MVRP	BRP	Equal Weights
Sortino	1,0400	0,2425	0,7600	1,0705
Downside Risk monthly	3,42%	4,57%	3,29%	3,67%
Downside Risk yearly	11,83%	15,83%	11,41%	12,72%
Sharpe	0,77	0,20	0,63	0,77
Volatility monthly	4,62%	5,55%	3,99%	5,11%
Volatility yearly	16,00%	19,23%	13,82%	17,69%
Returns monthly	0,97%	0,31%	0,70%	1,07%
Returns yearly	12,30%	3,84%	8,67%	13,62%
Max Drawdown	48,75%	49,75%	43,60%	49,43%
Returns March/2020	-11,59%	-8,62%	-10,11%	-12,13%

4. Using Cauchy Interlacing Theorem to Preserve the Correlation Structure

In this section we take the difficult task of dealing with inverting the correlation matrix, which might be potentially unstable, and the main source of extreme, unbalanced optimal Markowitz portfolios.

We will sidestep the problem of forecasting the expected returns by using sample historical returns and the Bayes-Stein shrinkage estimation, please see the appendix for details. It is worth noticing that this technique also requires the inverse of covariance and correlation matrices.

When the sample correlation matrix is quasi-ill-defined, which is often the case in practical work, a solution explored in the literature is the use of Ledoit-Wolf-shrinkage (see [3]) where the authors suggest an interpolation between the sample correlation matrix and a constant correlation matrix, where the constant is the average sample correlation. The authors provide a formula for the optimal interpolation size.

In the lemmas of the last section, the equal correlations assumption necessary for NRP and MVRP to be optimal, may be interpreted as being an "extreme", 100%, shrinkage of the correlation matrix towards its average.

We believe this full shrinkage excessively deforms the correlation structure and, in most cases, destroys important information that could be used for better portfolio diversification. Even if we adopt the optimal interpolation size proposed by [3], their technique imposes a shrinkage towards the average correlation that is uniformly across the whole correlation matrix, which sees to be inefficient.

This brings us to the main contribution of this paper. We acknowledge the need to shrink some correlations to push the full matrix away from being ill-defined. However, we believe a portion of the correlation matrix is "healthy" and worth keeping intact. To help us identify this portion we will use the following theorem (see [5]):

Cauchy's Interlacing theorem: Let A hermitian matrix of order $n > 1$ and B a main submatrix from A, of order $r \leq n$. If $\lambda_1 \geq \ldots \geq \lambda_n$ are the eigenvalues of A and $\theta_1 \geq \ldots \geq \theta_r$ are the eigenvalues of B, so $\lambda_i \geq \theta_i \geq \lambda_{i+n-r}$, for $1 \leq i \leq r$.

Correlation matrices of N assets often have eigenvalues very close to zero when N is large. Sometimes, when the number of correlations to be estimated $N(N-1)/2$ is too large compared to the number of dates T available for estimation, we even run into the unfortunate situation of having negative eigenvalues, indicating the matrix is not even positive-semi definite and, hence, couldn't hope to be a proper Correlation matrix.

The problems about inverting the correlation matrix is essentially a problem of having eigenvalues too close to zero, or that the ratio of the largest eigenvalue and the smallest eigenvalue, the condition number, is too big.

The Cauchy Interlacing theorem assures us that when you remove an asset, the smallest eigenvalue increases and the largest eigenvalue decreases and, hence, the condition number decreases. Therefore, we implement a recursive procedure whereby in each interaction we remove the asset that promotes the largest decrease in the condition number.

We loop this recursive Cauchy Interlacing procedure until the correlation submatrix is "well behaved enough". To make this statement more precise, at each iteration we evaluate portfolio leverage, computed as the gross exposure, which is the absolute sum of the long and short parts of the optimal Markowitz portfolio with maximum Sharpe ratio, which should be lower than an arbitrary value.

It is worth mentioning that other stopping criteria can also be used instead of imposing a a minimum eigenvalue. Positiveness conditions for the minimum variance portfolio or other prescribed economic rationales can be used, for example.
In this specific case, the rationale for this stopping criteria is that high leverage not only tends to make the portfolio unstable out-of-sample but also requires large collateral, sometime unfeasible or with the risk of triggering margin calls in volatile markets and forced liquidation is sharp sell offs.

We believe we should not do any shrinkage to the correlation submatrix of the assets that were not removed by this recursive procedure. We keep this portion of the correlation matrix intact and this maximum Sharpe ratio optimal sub-portfolio, that uses the sample covariance submatrix and sample historical returns with Stein-shrinkage, we name it the Cauchy Interlacing Risk Parity (CIRP) allocation.

For the assets that were removed by the recursive Cauchy interlacing procedure, we use MVRP, that is, we apply extreme correlation shrinkage. We call this portfolio the restricted MVRP.

The lemma below shows that the restricted MVRP and CIPR might offer good diversification opportunity. It is a similar rationale that lead us to the same conclusion for MVRP and NRP.

Lemma 4.1. *For any CIRP asset, if its correlations with all non-CIPR are the same, then CIRP and restricted MVRP portfolios are orthogonal.*

We define the Balanced Cauchy Intelacing Risk Parity (BCIRP) as the combination of CIRP and restricted MVRP with weights equal the inverse of their volatilities.

For completeness, we have also added in the table above the Hierarchical Risk Parity (HRP) allocation, which uses clustering and graphs theory to produce robust portfolios that have been used by some practitioners as a superior alternative to NRP. Also, we have added the Maximum Decorrelated portfolio built on top of the assets selected by the CIRP algorithm. Please see [6] and [7][8] respectively for further details.

That said, in the extensive Table 2 (Appendix D) we show that a pure Maximum Sharpe portfolio with assets selected by using the CIRP procedure performs vastly better than all competitor allocations in the backtest of our dataset, for the Brazilian Data, while BCIRP performance results are encouraging, in terms of other risk measures. The overall performance is even more impressive when observing Figure 1 (Appendix D).

In terms of the US Data, in the extensive Table 1 (Appendix D) we show that a maximum decorrelated portfolio with assets selected by using the CIRP procedure performs vastly better than all competitor allocations in the backtest of our dataset, while HRP performance results are also good in terms of other risk measures. The overall performance is also even more impressive when observing Figure 2 (Appendix D), suggesting that while one is aiming to build more aggressive portfolios, a CIRP based strategy is very attractive.

5. Mathematical Implementation of CIRP

Given a matrix A, the idea is to find the least degenerate main submatrix of A in each iteration, that is, that submatrix that has the highest minimum eigenvalue, such that this minimum eigenvalue is greater than an arbitrary number.
So

$$A = \begin{bmatrix} B & X^T \\ X & Z \end{bmatrix} \tag{1}$$

With $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{(n-k)\times(n-k)}$, where $1 \leq k, X \in R^k$ and $Z \in \mathbb{R}^{k \times k}$.
We begin our analysis when k = 1, finding the least degenerate submatrix (the principal submatrix that contains the largest least eigenvalue among all possible principal submatrices). Once this submatrix has been found, we compare whether its minimum eigenvalue is greater than or equal to our stopping criteria. If it does not comply, we pass to the case $k = 2$ and so on.
Cauchy's interlacing eigenvalues theorem tells us that

$$\lambda_{min}(A) \leq \lambda_{min}(B_{n-1}) \leq \lambda_{min}(B_{n-2}) \leq \ldots \leq \lambda_{min}(B_1).$$

Therefore, it ensures that our method stops for a minimum imposed eigenvalue as a stopping criterion.

Using this method, we show that it is possible to obtain a better way of working with the covariance matrix, based on the Cauchy interlacing eigenvalues theorem. The algorithm's objective is to eliminate assets that affect the stability of the covariance matrix, in addition to maintaining the assets that preserve the maximum information content for the Markowitz application. Further details on matrix theory are provided in the Appendix B.

6. Conclusions and Future Research

We started this paper stating that it is well known that Markowitz Portfolio Optimization often leads to unreasonable and unbalanced portfolios and its possible main causes are related to the fact that covariance matrices that are used within the Markowitz framework are degenerated and ill-posed.

In order to save the traditional (and still theoretically elegant) Markowitz framework by understanding its limitations and circumventing them by making use of a simple but yet powerful mathematical framework to preserve the correlation structure and optimize results.

By making extensive use of large datasets for two different markets, we show that somehow (Maximum Decorrelated Portfolio based on a correlation matrix obtained by the CIRP method - in the case of US data; or Maximum Sharpe porfolios based on James-Stein Estimator and CIRP covariance matrix - in the case of Brazilian data) our approach outperforms the traditional existing techniques and the so modern HRP (Hierarchical Risk Parity), which is based on Machine Learning techniques.

But yet, there is a huge room for further improvements. A more efficient implementation of the CIRP algorithm is needed, in order to allow an efficient analysis of even larger correlation matrices.

Also, there is a clear path involving the research of the impacts building portfolios and diversifing risk within sub-optimal portfolios chosen using risk factors, such as "low-vol", "beta", "momentum" (which was partially explored here for the US data), and other potential factors as well.

Furthermore, other possible extensions of this work are (but not limited to) testing the performance for other markets; evaluate other stoppage criteria, such as a fixed minimum eigenvalue for the correlation matrices; how to obtain long-only portfolios within the Markowitz Framework; and, finally, extending this work to variable selection to exclude collinearity issues in a broader set of regression problems, which typically involve covariance matrices analysis and computations.

Appendix

A. Proof of Lemmas

Lemma A.1. *If all assets have the same Sharpe ratios $\acute{s} = s_i = \dfrac{\mu_i}{\sigma_i}$ and all the pair-wise correlations ρ are the same, then NRP allocation is the portfolio that achieves the maximum Sharpe Ratio.*

Proof:

We would like to maximize the portfolio Sharpe ratio $\dfrac{\mu^T w}{\sqrt{(w^T \sum w)}}$ over all possible w. First notice that this objective function is invariant to scale, so that we only need to very that the maximum is achieved for $w_i^i = \dfrac{1}{\sigma_i}$.

Taking ln(), in order to maximize $\ln(\mu^T w) - \dfrac{1}{2} \ln\left(w^T \Sigma w\right)$ we have to verify the first order conditions (F.O.C.): $\dfrac{\mu}{\mu^T w} - \dfrac{\Sigma w}{w^T \Sigma w} = 0$.

Define the correlation matrix $C(\rho) = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}$, so that Covariance matrix become $\sum = diag(\sigma).C(\rho).diag(\sigma)$, where $diag(\sigma)$ is de diagonal matrix of volatilities.

With this notation, pre-multiplying the F.O.C by $diag(\sigma)^{-1}$ and noticing that

$diag(\sigma)^{-1}.\mu = \begin{pmatrix} \frac{\mu_i}{\sigma_1} \\ \vdots \\ \frac{\mu_i}{\sigma_n} \end{pmatrix} = \acute{s} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, the F.O.C becomes $\dfrac{\acute{s}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}{\mu^T w} - \dfrac{C(\rho).diag(\sigma).w}{w^T \sum w} = 0$

To verify that this equation is valid for $w = \begin{pmatrix} \frac{1}{\sigma_1} \\ \vdots \\ \frac{1}{\sigma_n} \end{pmatrix}$, we just need to compute the pieces

(i) $\mu^T w = n\acute{s}$,

(ii) $C(\rho).diag(\sigma).w = C(\rho).\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = (1 - \rho + n\rho)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ and

(iii) $w^T \Sigma w = w^T diag(\sigma).C(\rho).diag(\sigma)w = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}' C(\rho) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = n(1 - \rho + n\rho)$

and substitute them in. This completes the proof.

Lemma A.2. *if all correlations are the same and different from zero and $\acute{\sigma}_N$ con-*

verges as you increase N, then MVRP allocation

$$W_i^{NRP} = \frac{\frac{1}{\sigma_i} \times \left(\frac{1}{\sigma_i} - \frac{1}{\acute{\sigma}_N}\right)}{\sum_i \left[\frac{1}{\sigma_i} \times \left(\frac{1}{\sigma_i} - \frac{1}{\acute{\sigma}_N}\right)\right]},$$

where $\frac{1}{\acute{\sigma_N}} = \frac{\sum_i \left[\frac{1}{\sigma_i}\right]}{N}$, *converges to the global minimum variance portfolio (GMVP). If the correlations are all zero then the GMVP is given by weights that are proportional to the inverse of the variance of each assets.*

Proof:

The minimum variance portfolio is given by $w^i = \gamma . \sum^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, with $\gamma \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \sum^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$.

The correlation matrix $C(\rho) = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}$ has the inverse

$$C(\rho) = \frac{1+(N+2)\rho}{1+(N+2)\rho-(N-1)\rho^2} \begin{pmatrix} 1 & \hat{\rho} & \dots & \hat{\rho} \\ \hat{\rho} & 1 & \dots & \hat{\rho} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\rho} & \hat{\rho} & \dots & 1 \end{pmatrix}, \text{ where } \hat{\rho} = \frac{-\rho}{1+(N+2)\rho}$$

Hence

$$w = \gamma.diag(\sigma^{-1}).C(\rho)^{-1}.diag(\sigma^{-1}) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \gamma.diag(\sigma^{-1}).C(\rho)^{-1} \begin{pmatrix} \frac{1}{\sigma_1} \\ \vdots \\ \frac{1}{\sigma_N} \end{pmatrix}$$

$$= \gamma \frac{1+(N-2)\rho}{1+(N-2)\rho-(N-1)\rho^2} diag(\sigma^{-1}) \begin{pmatrix} \frac{1-\hat{\rho}}{\sigma_1} + \hat{\rho} \sum_i \frac{1}{\sigma_1} \\ \vdots \\ \frac{1-\hat{\rho}}{\sigma_N} + \hat{\rho} \sum_i \frac{1}{\sigma_I} \end{pmatrix}$$

$$w = \gamma \frac{1+(N-2)\rho}{1+(N-2)\rho-(N-1)\rho^2} \begin{pmatrix} \frac{1}{\sigma_1} . \left(\frac{1-\hat{\rho}}{\sigma_1} + \frac{\hat{\rho} N}{\sigma_N}\right) \\ \vdots \\ \frac{1}{\sigma_N} . \left(\frac{1-\hat{\rho}}{\sigma_N} + \frac{\hat{\rho} N}{\sigma_N}\right) \end{pmatrix}$$

If $\hat{\rho} = 0$, and the i-th component of the global minimum variance portfolio is proportional to $\frac{1}{{\sigma_i}^2}$.
If $\hat{\rho} \neq 0$, then as $N \to \infty$ we have $\hat{\rho} \to \infty$ and $\hat{\rho}.N \to -1$. Therefore, in the limit, the i-th component of the global minimum variance portfolio is proportional to $\frac{1}{\sigma_i} \left(\frac{1}{i} - \frac{1}{\acute{\sigma}_N}\right)$. This completes the proof.

Lemma A.3. *If all correlations are the same the NRP and MVRP portfolios are orthogonal.*

Proof:

The covariance between NRP and MVRP is

$$w^{NRP^T} w^{MVRP} = \frac{1}{\sum_i \left(\frac{1}{\sigma_i}\right)} \begin{pmatrix} \frac{1}{\sigma_1} \\ \vdots \\ \frac{1}{\sigma_N} \end{pmatrix} diag(\sigma).C(\rho).diag(\sigma) \begin{pmatrix} \frac{1}{\sigma_1}\left(\frac{1}{\sigma_1} - \frac{1}{\acute{\sigma}_N}\right) \\ \vdots \\ \frac{1}{\sigma_N}\left(\frac{1}{\sigma_N} - \frac{N}{\acute{\sigma}_N}\right) \end{pmatrix} \frac{1}{\sum_i \left[\frac{1}{\sigma_1}\left(\frac{1}{\sigma_1} - \frac{1}{\acute{\sigma}_N}\right)\right]}$$

Simple math give us:

(i) $\begin{pmatrix} \frac{1}{\sigma_1} \\ \vdots \\ \frac{1}{\sigma_1} \end{pmatrix} .diag(\sigma) = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ and

(ii)$C(\rho)diag(\sigma) \begin{pmatrix} \frac{1}{\sigma_1}\left(\frac{1}{\sigma_1} - \frac{1}{\acute{\sigma}_N}\right) \\ \vdots \\ \frac{1}{\sigma_N}\left(\frac{1}{\sigma_N} - \frac{N}{\acute{\sigma}_N}\right) \end{pmatrix} = C(\rho) \begin{pmatrix} \frac{1}{\sigma_1} - \frac{1}{\acute{\sigma}_N} \\ \vdots \\ \frac{1}{\sigma_N} - \frac{1}{\acute{\sigma}_N} \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{\sigma_1} - \frac{1}{\acute{\sigma}_N}\right)(1-\rho) \\ \vdots \\ \left(\frac{1}{\sigma_N} - \frac{1}{\acute{\sigma}_N}\right)(1-\rho) \end{pmatrix}$

Multiplying (i) * (ii) sums up to zero.
Hence the covariance is zero, proving the claim.

Lemma A.4. *For any CIRP asset, if its correlations with all non-CIPR are the same, then CIRP and restricted MVRP portfolios are orthogonal.*

Proof:

Let P be the number of assets selected by the recursive CIRP procedure and Ψ_{CIRP} their PxP correlation matrix. Let M = N-P the number of non-CIRP stocks, or equivalently, the restricted MVRP stocks. Denote their MxM correlation matrix by $\Psi_{RESTRICTED_{MRP}}$ and the cross correlation matrix PxM between CIRP and non-CIRP assets as Ψ_{Cross}.
By assumption, each one of the P CIPR assets have the same correlation with all Non-CIPR assets, that is, $\Psi_{cross} = \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_P \end{pmatrix} \begin{pmatrix} 1 & \ldots & 1 \end{pmatrix}$ The covariance between CIRP and MVRP is

$$COV = \begin{pmatrix} W^{CIRP} \\ 0_{Mx1} \end{pmatrix} diag(\sigma) \begin{pmatrix} \Psi_{CIRP} & \Psi_{Cross} \\ \Psi_{Cross} & \Psi_{Restricted_{MRP}} \end{pmatrix} diag(\sigma) \begin{pmatrix} 0_{Px1} \\ W^{Restricted_{MRP}} \end{pmatrix}$$

However,

$$diag(\sigma) \begin{pmatrix} 0_{P\times 1} \\ W^{Restricted_{MRP}} \end{pmatrix} = diag(\sigma) \begin{pmatrix} 0_{P\times 1} \\ \frac{1}{\sigma_{P+1}}\left(\frac{1}{\sigma_{P+1}} - \frac{1}{\acute{\sigma}_{P+M}}\right) \\ \frac{1}{\sigma_{P+1}}\left(\frac{1}{\sigma_{P+M}} - \frac{1}{\acute{\sigma}_{P+M}}\right) \end{pmatrix}$$

Where $\frac{1}{\acute{\sigma}_{P+M}} = \frac{\sum_i \left[\frac{1}{\sigma_{P+1}}\right]}{M}$ Sothe covariance becomes

$$Cov = \begin{pmatrix} W_1^{CIRP}.\sigma_1 \\ \vdots \\ W_N^{CIRP}.\sigma_N \end{pmatrix} \Psi_{Cross} \begin{pmatrix} \frac{1}{\sigma_{P+1}} - \frac{1}{\acute{\sigma}_{P+M}} \\ \vdots \\ \frac{1}{\sigma_{P+1}} - \frac{1}{\acute{\sigma}_{P+M}} \end{pmatrix}$$

$$Cov = \begin{pmatrix} W_1^{CIRP}.\sigma_1 \\ \vdots \\ W_N^{CIRP}.\sigma_N \end{pmatrix} \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_P \end{pmatrix} \begin{pmatrix} 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_{P+1}} - \frac{1}{\acute{\sigma}_{P+M}} \\ \vdots \\ \frac{1}{\sigma_{P+1}} - \frac{1}{\acute{\sigma}_{P+M}} \end{pmatrix}$$

$$Cov = \begin{pmatrix} W_1^{CIRP}.\sigma_1 \\ \vdots \\ W_N^{CIRP}.\sigma_N \end{pmatrix} \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_P \end{pmatrix} \frac{\sum_i \left[\frac{1}{\sigma_{P+1}}\right]}{M}$$

however, $\begin{pmatrix} 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_{P+1}} - \frac{1}{\acute{\sigma}_{P+M}} \\ \vdots \\ \frac{1}{\sigma_{P+1}} - \frac{1}{\acute{\sigma}_{P+M}} \end{pmatrix} = \sum_i \left[\frac{1}{\sigma_{P+i}} - \frac{1}{\acute{\sigma}_{P+M}}\right] = 0$

So Cov $= 0$, proving the Claim.

B. Proof of Cauchy Interlacing Theorem

B.1. Notions of linear algebra

We will recall certain definitions of linear algebra. For those interested in studying this topic in depth, see [9].

Definition B.1. *Let $A \in \mathbb{F}^{n \times n}(\mathbb{F} = \mathbb{R} \vee \mathbb{C}).\lambda \in \mathbb{F}$ is an eigenvalue of A if, and only if, there is a nonzero vector $v \in \mathbb{R}^n$ that satisfies the equation $Av = \lambda v$. Where v is called an eigenvector associated of A with the eigenvalue λ.*

Example B.1. *Let be*

$$A = \begin{bmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{bmatrix}$$

with eigenvalues $\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = -2$ and eigenvectores respective $v_1 = (1, 2, 1), v_2 = (1, -4, -1), v_3 = (1, -1, -1)$.

Definition B.2. *Let $A \in \mathbb{F}^{n \times n}$. The characteristic polynomial of the matrix A, denoted by $p_A(x)$, is given by $p_A(x) = det(A - xI)$.*

Example B.2. *Let A be a square matrix from the example (B.1). The characteristic polynomial is*

$$p_A(x) = det(A - xI) = det \begin{bmatrix} 1-\lambda & -1 & 4 \\ 3 & 2-\lambda & -1 \\ 2 & 1 & -1-\lambda \end{bmatrix}$$

$$= \lambda^3 - 2\lambda^2 - 5\lambda + 6 = (\lambda - 3)(\lambda - 1)(\lambda + 2).$$

Theorem B.1. *The number $\lambda \in \mathbb{F}$ is an eigenvalue of the square matrix $A \in \mathbb{F}^{n\times x}$ if, and only if, $det(\lambda I - A) = 0$. That is, if λ is the root of the characteristic polynomial $p_A(x) = det(xI - A)$. Furthermore, if A has order n, then the characteristic polynomial $p_A(x)$ has degree n.*

This can be seen in the examples (B.1) and (B.2).

Definition B.3. *A matrix $A \in \mathbb{F}^{n\times x}$ is called* **Hermitian matrix***, when $A = \overline{A}^T$. Where $\overline{A}$ denotes the conjugated matrix of A, and A^T denotes the transposed matrix of A.*

Example B.3. *The square matrices*

$$A = \begin{bmatrix} 1 & -3 \\ -3 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 1+i \\ 1-i & 1 \end{bmatrix},$$

are Hermitian matrices.

Theorem B.2. *If A is a hermitian matrix of order n, then its eigenvalues $\lambda_1, \dots, \lambda_n$ are real. In addition, A has unit eigenvectors $v_1, \dots, v_n$, two to two orthogonal, with $Av_j = \lambda$ for $j = 1, \dots, n$. Such eigenvectors form a orthonormal base of $\mathbb{R}^n$.*

If A is a hermitian matrix, $\lambda\langle v, v\rangle = \langle Av, v\rangle = \langle v, Av\rangle = \overline{\lambda}\langle v, v\rangle$. So $\lambda = \overline{\lambda}$. For matrix A in the example (B.3), their eigenvalues are $\lambda_1 = 4, \lambda_2 = -2$. For matrix $B, \lambda_1 = 1 + \sqrt{2}, \lambda_2 = \sqrt{2}$.

Observation B.1. *How the eigenvalues of a Hermitian matrix are real, so we can renamed and ordered them how $\lambda_1 \geq \ldots \geq \lambda_n$.*

Definition B.4. *Let $A \in \mathbb{F}^{n\times n}$. A square matrix B is a main submatrix of order m of A, with $m < n$, if B is obtained by the deletion of $n-m$ rows and $n-m$ columns of A.*

Example B.4. *For matrix A in the example (B.1),*

$$B = \begin{bmatrix} 1 & -1 \\ 3 & 2 \end{bmatrix}, C = \begin{bmatrix} 1 & 4 \\ 2 & -1 \end{bmatrix}, D = \begin{bmatrix} 2 \end{bmatrix}$$

are submatrices of A.

B.2. Rayleigh's principle

Let A be a hermitian matrix of order n and $\lambda_1 \geq \ldots \geq \lambda_n$ its eigenvalues, with eigenvectors $v_1, \dots, v_n$ respectively, two to two orthogonal, such that form a orthonormal base of $\mathbb{R}^n$. We have $\lambda_i = v_i^T A v_i$, for $i = 1, 2, \dots, n$.

B.2.1. Rayleigh ratio

Theorem B.3 (Rayleigh's principle)**.** *Let A be a hermitian matrix of order n and $\lambda_1 \geq \ldots \geq \lambda_n$ its eigenvalues. So*

$$\lambda_1 = \max_{\|v\|=1} v^T Av,$$

and the maximum is reached when v is a unitary eigenvector associated with the eigenvalue λ_1.

Proof: Considering $v_1, \ldots, v_n$ unitary eigenvalues two to two orthonormal of A, such that form a basis for $\mathbb{R}^n$. Therefore, for $v \in \mathbb{R}^n$ unitary ($\| v \| = 1$), we have $v = \sum_{i=0}^{n} c_i v_i$.
Thus

$$1 = \langle v, v \rangle = \left\langle \sum_{i=0}^{n} c_i v_i, \sum_{j=0}^{n} c_j v_j \right\rangle = \sum_{i=0}^{n} \sum_{j=0}^{n} c_i c_j \langle v_i, v_j \rangle$$
$$= \sum_{i=0}^{n} | c_i | .$$

Also, how

$$Av = A(\sum_{i=0}^{n} c_i v_i) = \sum_{i=0}^{n} c_i A v_i = \sum_{i=0}^{n} c_i \lambda_i v_i.$$

Thus

$$\langle Av, v \rangle = \left\langle \sum_{i=0}^{n} c_i \lambda_i v_i, \sum_{i=0}^{n} c_i v_i \right\rangle = \sum_{i=0}^{n} \lambda_i | c_i |^2 \leq \lambda_1 (| c_1 |^2 + \ldots + | c_n |^2) = \lambda_1.$$

Then, as we are working on the closed and bounded set $\{x \in \mathbb{R}^n : \| x \| = 1\}$, then it reaches a maximum, and as $\langle Av_1, v_1 \rangle = \lambda_1$. Thus

$$\lambda_1 = \max_{\|v\|=1} v^T A v.$$

This is, λ_1 is the largest value defined by the quadratic form $\langle Av, v \rangle$ in the unit sphere $\| v \| = 1$.

Observation B.2. *Let A be a hermitian matrix of order n and $\lambda_1 \geq \ldots \geq \lambda_n$ its eigenvalues. Analogous to the previous theorem, we have to*

$$\lambda_n = \min_{\|v\|=1} v^T A v$$

The minimum is reached when v is the unit eigenvector associated with the eigenvalue λ_n.

Example B.5. *Let*

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \tag{2}$$

whose characteristic polynomial is given by

$$p_A(x) = det(xI - A) = x^2 - 2x - 3 = (x + 1)(x - 3).$$

With eigenvalues $\lambda_1 = 3, \lambda_2 = -1$, and orthonormal eigenvectors
$v_1 = (\sqrt{2}/2, \sqrt{2}/2)^T, v_2 = (\sqrt{2}/2, -\sqrt{2}/2)^T$.
By Rayleigh's principle we have

$$\lambda_1 = \max_{\|v\|=1} v^T A v = \max_{v_1^2 + v_2^2 = 1} v_1^2 + 4 v_1 v_2 + v_2^2.$$

Using optimization methods we obtain that $\lambda_1 = 3$, which takes as an argument the unit vector $(\sqrt{2}/2, \sqrt{2}/2)^T$.

Corollary B.1. *Let A be a hermitian matrix of order n and $\lambda_1 \geq \ldots \geq \lambda_n$ its eigenvalues. So*

$$\lambda_1 = \max_{x \neq 0} \frac{x^T A x}{x^T x}.$$

The expression $\frac{x^T Ax}{x^T x}$ is called **Rayleigh quotient**.

Proof: Since $x \neq 0$ we have $\| x \| > 0$, so we get

$$\frac{x^T A x}{x^T x} = \frac{\langle Ax, x\rangle}{\| x \|^2} = \left\langle A \frac{x}{\| x \|}, \frac{x}{\| x \|} \right\rangle = v^T A v,$$

where v is a unit vector.

Theorem B.4 (Generalized Rayleigh Principle)**.** *Let A be a hermitian matrix of order n and $\lambda_1 \geq \ldots \geq \lambda_n$ its eigenvalues. For $i \geq 2$, be $v_1, \ldots, v_{i-1}$ unitary eigenvectors two to two orthogonal, associated with eigenvalues respectively $\lambda_1, \ldots, \lambda_{i-1}$. So*

$$\lambda_i = \max_{\|v\|=1, v \perp v_1 \ldots, v_{i-1}} v^T A v$$

The maximum is reached when v is the eigenvector associated with the eigenvalue λ_i.

Proof: Considering $v_1, \ldots, v_n$ unitary eigenvalues two to two orthonormal of A, forming a basis for $\mathbb{R}^n$.
For $v \in \mathbb{R}^n$ unitary, we have to $v = c_1 v_1 + \ldots + c_{i-1} v_{i-1} + c_i v_i + \ldots + c_n v_n$.
So we have

$$1 = \| v \|^2 = \sum_{i=0}^{n} | c_i |$$

Also

$$\langle v, v_k \rangle = \langle c_1 v_1 + \ldots + c_n v_n, v_k \rangle = c_k \langle v_k, v_k \rangle = c_k, \forall k = 1, \ldots, n.$$

Also, how do we want $v \perp v_1, \ldots, v_{i-1}$, so $c_k = \langle v, v_k \rangle = 0, \forall k = 1, \ldots, i-1$.
So

$$\langle Av, v \rangle = \left\langle \sum_{i=0}^{n} c_i \lambda_i v_i, \sum_{i=0}^{n} c_i v_i \right\rangle = \lambda_i \mid c_i \mid^2 + \ldots + \lambda_n \mid c_n \mid^2 \leq \lambda_i (\mid c_i \mid^2 + \ldots + \mid c_n \mid^2) = \lambda_i.$$

Example B.6. *Let be*

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

with eigenvalues $\lambda_1 = 3, \lambda_2 = 2, \lambda_3 = 1$, and their respective eigenvectors $v_1 = (1, 0, 0)$,
$v_2 = (0, 1, 0), v_3 = (0, 0, 1)$. *By the Generalized Rayleigh Principle*

$$\lambda_2 = \max_{u_1^2 + u_2^2 + u_3^2 = 1, \langle u, v_1 \rangle = 0} (3u_1^2 + 2u_2^2 + u_3^2) = \max_{u_1^2 + u_2^2 + u_3^2 = 1, u_1 = 0} (3v_1^2 + 2v_2^2 + v_3^2)$$

$$= \max_{u_2^2 + u_3^2 = 1} (2v_2^2 + v_3^2) = 2.$$

B.3. Cauchy's interlacing eigenvalue theorem

Theorem B.5 (Courant-Fischer theorem)**.** *Let A be a hermitian matrix of order n and $\lambda_1 \geq \ldots \geq \lambda_n$ its eigenvalues. So, for $i < n$, we have*

$$\lambda_{i+1} = \min_{u_1,\ldots,u_i} \left[\max_{\|v\|=1, v\perp u_1\ldots,u_i} v^T A v \right].$$

Proof: let's define the function

$$\phi(u_1, \ldots, u_i) = \max_{\|v\|=1, v\perp u_1\ldots,u_i} v^T A v. \tag{4}$$

This is well defined because it is the maximum of a continuous function in a compact set.
Follows from Theorem B.4 that

$$\phi(v_1 \ldots, v_i) = \lambda_{i+1}.$$

Consider all vectors v_0 of the form

$$v_0 = c_1 v_1 + \ldots + c_i v_i + c_{i+1} v_{i+1}. \tag{5}$$

Of the conditions we have to

$$\langle v_0, u_k \rangle = \sum_{j=1}^{i+1} c_j \langle c_j, u_k \rangle = 0, \forall k = 1, \ldots, i.$$

This system provides equations for variables $c_1, \ldots, c_{i+1}$. As the homogeneous system has more variables than equations, we have that there are $\gamma_1 \ldots, \gamma_{i+1}$ with $\lambda_k \neq 0$, for some k, such that $c_j = \lambda \gamma_j$, for $1 \leq j \leq i+1$, is solution for every value of λ.
Considering

$$\lambda = \left(\sum_{j=1}^{i+1} \mid \gamma_j \mid^2 \right)^{-1/2}$$

we obtain a solution in the system with $\sum \mid c_j \mid = 1$.
Then the vector $v_0 = \sum c_j v_j$ satisfies all the conditions of maximization of equation (4). So we have to $\phi(u_1, \ldots, u_i) \geq \langle A v_0, v_0 \rangle$.
How

$$\langle A v_0, v_0 \rangle = \left\langle \sum_{j=1}^{i+1} c_j \lambda_j v_j, \sum_{k=1}^{i+1} c_k v_k \right\rangle = \sum_{j=1}^{i+1} \mid c_j \mid^2 \lambda_j \geq \lambda_{j+1} \sum_{j=1}^{i+1} \mid c_j \mid^2 = \lambda_j.$$

This

$$\phi(u_1, \ldots, u_i) \geq \lambda_{i+1},$$

for any set of vetores $\{u_1, \ldots, u_i\}$, and by the equation (5) we get the equality of the statement.

Theorem B.6 (Cauchy's interlacing eigenvalue theorem)**.** *Let A hermitian matrix of order $n > 1$ and B be a main submatrix of order $n-1$ from A.*
If

$$\lambda_1 \geq \ldots \geq \lambda_n \text{ are the eigenvalues of } A \text{ and}$$

$$\theta_1 \geq \ldots \geq \theta_{n-1} \text{ are the eigenvalues of } B,$$

so

$$\lambda_1 \geq \theta_1 \geq \lambda_2\theta_2 \geq \ldots \geq \lambda_{n-1} \geq \theta_{n-1} \geq \lambda_n.$$

Proof: Let us first show the case where B is the main submatrix of order $n-1$ of A obtained by deleting the last row and column.

$$A = \begin{bmatrix} B & z \\ z^T & a_{nn} \end{bmatrix} \tag{6}$$

Let the vector $v \in \mathbb{R}^T$ be such that $v = (u \quad 0)^T$, so $Av = Bu$. For any vectors $x_1, \ldots, x_{i-1}$, with $i < n$, we have

$$\max_{\|v\|=1, v\perp x_1,\ldots,x_{i-1}} \langle Av, v\rangle \geq \max_{\|v\|=1, v\perp x_1,\ldots,x_{i-1},e_n} \langle Av, v\rangle, \tag{7}$$

where e_n is the canonical vector $(0, 0, \ldots, 1) \in \mathbb{R}^n$.
Taking $x_j = (w_j \quad x_j^n)^T$, with $w_j \in \mathbb{R}^n$, for $1 \leq j \leq i-1$. So

$$\max_{\|v\|=1, v\perp x_1,\ldots,x_{i-1},e_n} \langle Av, v\rangle = \max_{\|u\|=1, u\perp w_1,\ldots,w_{i-1}} \langle Bu, u\rangle.$$

Then we can define the function

$$f(x_1, \ldots, x_{i-1}) = \max_{\|v\|=1, v\perp x_1,\ldots,x_{i-1},e_n} \langle Av, v\rangle = \max_{\|u\|=1, u\perp w_1,\ldots,w_{i-1}} \langle Bw, w\rangle,$$

taking each coordinate of the vector w_j as the first $n-1$ coordinates of the vector x_j, for $j = 1, \ldots, i-1$.
As is the same function, then

$$\min_{x_1,\ldots,x_{i-1}} \left[\max_{\|v\|=1, v\perp x_1,\ldots,x_{i-1},e_n} \langle Av, v\rangle \right] = \min_{w_1,\ldots,w_{i-1}} \left[\max_{\|u\|=1, u\perp w_1,\ldots,w_{i-1}} \langle Bu, u\rangle \right] = \theta_i. \tag{8}$$

Let's note that:
For $n = 1$ we have

$$\max_{\|v\|=1, v\perp e_n} \langle Av, v\rangle = \max_{\|u\|=1} \langle Bu, u\rangle = \theta_1.$$

Using the theorem (B.5) and the equations (7), (8), we have

$$\lambda_i = \min_{x_1,\ldots,x_{i-1}} \left[\max_{\|v\|=1, v\perp x_1,\ldots,x_{i-1}} \langle Av, v\rangle \right] \geq \min_{w_1,\ldots,w_{i-1}} \left[\max_{\|u\|=1, u\perp w_1,\ldots,w_{i-1}} \langle Bu, u\rangle \right] = \theta_i.$$

The equation (8), we have

$$\theta_i = \min_{x_1,\ldots,x_{i-1}} \phi(x_1, \ldots, x_{i-1}, e_n), \tag{9}$$

where we define

$$\phi(x_1, \ldots, x_i) = \max_{\|v\|=1, v\perp x_1\ldots,x_i} v^T A v.$$

By the theorem (B.5) we have

$$\min_{x_1\dots,x_i} \phi(x_1,\dots,x_i) = \lambda_{i+1}. \tag{10}$$

In the problem (9), we consider $x_n = e_n$, then in the equation (10) we are more likely to find a minimum, thus

$$\theta_i = \min_{x_1,\dots,x_{i-1}} \phi(x_1,\dots,x_{i-1},e_n) \geq \min_{x_1\dots,x_i} \phi(x_1,\dots,x_i) = \lambda_{i+1}.$$

Example B.7. *Considering a matrix A of order n, and a submatrix B such that*

$$A = \begin{bmatrix} B & 0 \\ 0 & Z \end{bmatrix} \tag{11}$$

Let $w \neq 0$ be an eigenvector of real matrix B, with its respective eigenvalue λ_w.
So $\lambda_w = \frac{w^T B w}{w^T w}$, because $w^T B w = w^T \lambda_w w$.
The vector $w_0 = (w \;\; 0)^T$ is an eigenvector of A with its eigenvalue λ_{w_0}.
How $Bw = \lambda_w w$, so $Aw_0 = Bw = \lambda_w w = \lambda_w w_0$.
If we impose that w is the smallest eigenvector of B, not necessarily the smallest eigenvalue of A, therefore $\lambda_{min_A} \leq \lambda_w$.

Theorem B.7 (Generalized Cauchy's interlacing eigenvalue theorem)**.** *Let A hermitian matrix of order $n > 1$ and B_r be a main submatrix of order r from A whit $1 \leq r < n$.*
If

$$\lambda_1 \geq \ldots \geq \lambda_n \text{ are the eigenvalues of } A \text{ and}$$

$$\theta_1 \geq \ldots \geq \theta_r \text{ are the eigenvalues of } B,$$

so

$$\lambda_i \geq \theta_i \geq \lambda_{i+n-r},$$

for $1 \leq i \leq r$.

Proof: We proceed by induction in $n - r$.
For $n - r = 1$ we are in the case of the previous theorem.
Suppose the result is valid for $n - r - 1$ and we denote by $\gamma_1 \geq \ldots \geq \gamma_{r+1}$ the eigenvalues of the matrix B_{r+1}. By theorem (B.6) we have $\gamma_i \geq \theta_i \geq \gamma_{i+1}$.
So we have

$$\lambda_i \geq \gamma_i \geq \theta_i \geq \gamma_{i+1} \geq \lambda_{i+n-r}.$$

Thus

$$\lambda_i \geq \theta_i \geq \lambda_{i+n-r}.$$

Example B.8. *Let be*

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix} \tag{12}$$

and a submatrix

$$B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \tag{13}$$

For each matrix we obtain their respective eigenvalues of A and B.

$$\lambda_1 = 3 + \sqrt{2}, \lambda_2 = 3 - \sqrt{2}, \lambda_3 = 1$$

$$\theta_1 = 3, \theta_2 = 1$$

By the theorem we have to compare for the matrices A y B

$$3 + \sqrt{2} \geq 3 \geq 3 - \sqrt{2} \geq 1 \geq 1$$

C. Stein Shrinkage

Stein-Shrinkage (see [10]) procedure is sometimes used to estimate expected returns as interpolation between the sample historical expected returns and a fixed "grand" average return, which should reduce the estimation error. The main idea is to acknowledge the uncertainty on the parameters estimation. Therefore, a Bayes procedure is applied in order to deal with such an uncertainty. The novelty is to minimize the estimation error of the whole portfolio instead of individual assets, which is more precise.
Basically the Bayes-Stein estimator gives us the assets' expected return to be considered for optimal allocation. In that case, the estimated expected return, R^{BS}, is:

$$R^{BS} = (1 - w)\overline{R} + wr^{MV}\mathbf{1}$$

where

$\overline{R}$ is an $n \times 1$ vector of the mean sample returns;
r^{MV} is a scalar representing the expected return of the minimum variance portfolio;
$\mathbf{1}$ is nx1 vector of ones;
$w = min\left(1, \dfrac{\frac{n-2}{T}}{\left(\overline{R} + r^{MV}\right)' \sum^{-1} \left(\overline{R} + r^{MV}\right)}\right)$;
n is the number of assets;
T is the size sample;
$\sum$ is the covariance matrix of returns.

Both the R^{BS} e w result from the Bayes-Stein shrinkage procedure.

D. Tables and Images

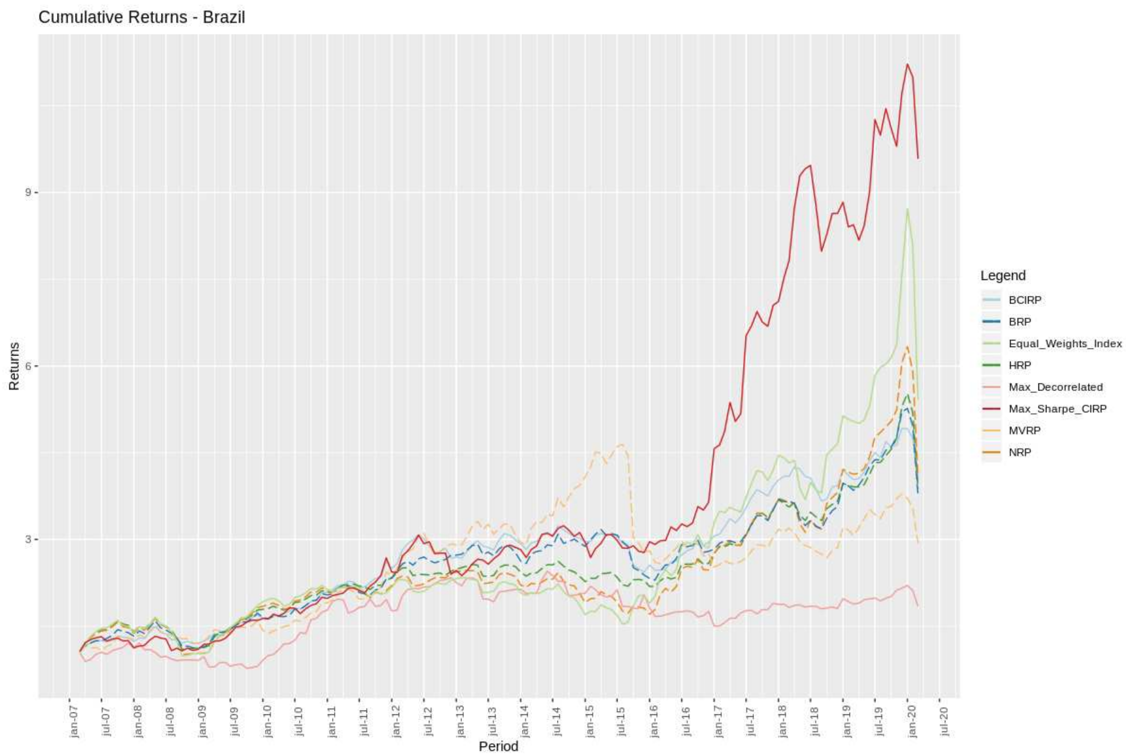

Figure 1: Cumulative Returns for Brazil

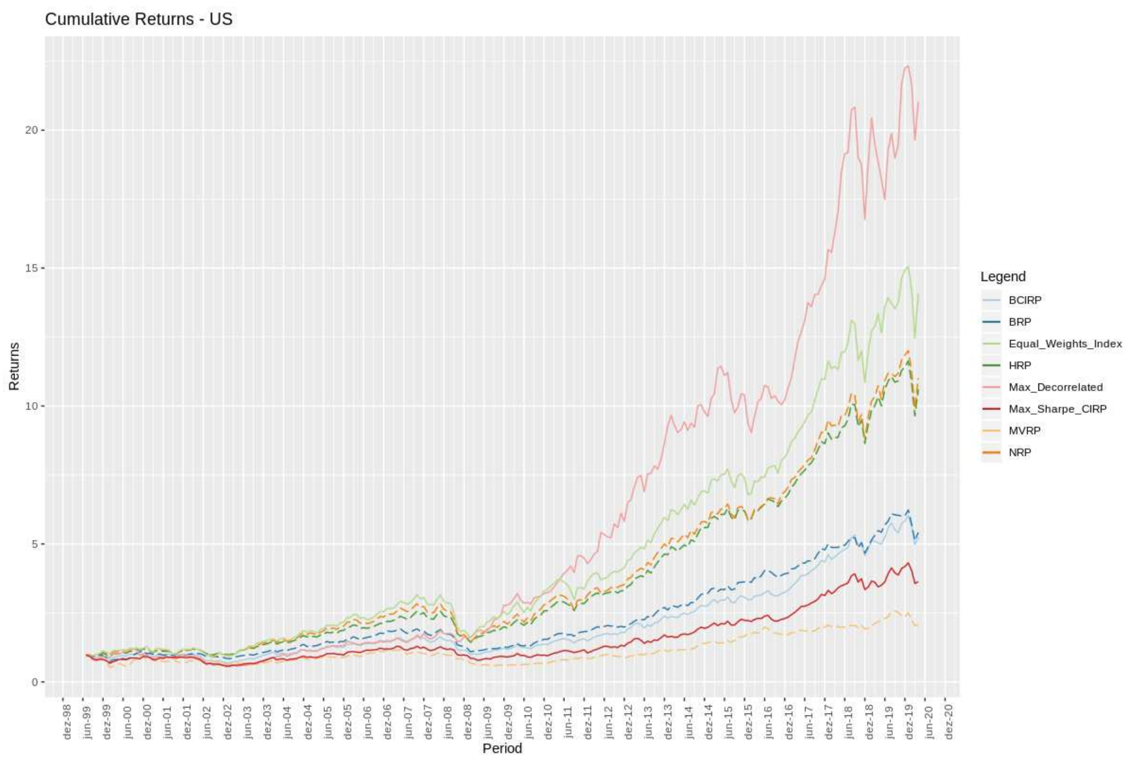

Figure 2: Cumulative Returns for US

	Markowitz + Stein	BCIRP	HRP	NRP	MVRP	BRP	Max Decorrelated (CIRP)	Equal Weights Index	Max Sharpe CIRP
Sortino	0,0000	0,7930	1,1526	1,0400	0,2425	0,7600	1,4068	1,0705	0,5685
Downside Risk monthly	474,70%	3,06%	3,04%	3,42%	4,57%	3,29%	3,27%	3,67%	3,31%
Downside Risk yearly	1644,43%	10,61%	10,54%	11,83%	15,83%	11,41%	11,34%	12,72%	11,46%
Sharpe Ratio	0	0,59	0,87	0,77	0,20	0,63	0,86	0,77	0,42
Returns montthly		0,68%	0,96%	0,97%	0,31%	0,70%	1,24%	1,07%	0,53%
Returns yearly		8,41%	12,15%	12,30%	3,84%	8,67%	15,96%	13,62%	6,52%
Volatility yearly	1292,69%	14,35%	13,94%	16,00%	19,23%	13,82%	18,55%	17,69%	15,59%
Volatility monthly	373,17%	4,14%	4,02%	4,62%	5,55%	3,99%	5,36%	5,11%	4,50%
Returns March/2020	-0,25%	-11,36%	-10,81%	-11,59%	-8,62%	-10,11%	-9,19%	-12,13%	-11,36%
Max Drawdown		39,93%	44,25%	48,75%	49,75%	43,60%	52,08%	49,43%	43,42%

Table 1: Results for US Data

	Markowitz + Stein	BCIRP	HRP	NRP	MVRP	BRP	Max Decorrelated (CIRP)	Equal Weights Index	Max Sharpe CIRP
Sortino	0	1,00	0,72	0,65	0,48	0,79	0,25	0,75	1,16
Downside Risk monthly	676,94%	3,10%	4,14%	4,84%	4,82%	3,74%	5,13%	5,12%	4,53%
Downside Risk yearly	2344,99%	10,75%	14,36%	16,77%	16,69%	12,96%	17,76%	17,73%	15,69%
Sharpe	0	0,75	0,76	0,54	0,44	0,66	0,17	0,59	0,88
Volatility monthly	671,58%	4,16%	4,46%	5,78%	5,28%	4,48%	7,72%	6,59%	6,01%
Volatility yearly	2326,42%	14,41%	15,44%	20,01%	18,30%	15,52%	26,73%	22,82%	20,82%
Returns montthly		0,86%	0,83%	0,87%	0,64%	0,81%	0,37%	1,05%	1,41%
Returns yearly		10,81%	10,36%	10,90%	8,00%	10,17%	4,47%	13,37%	18,28%
Max Drawdown	12468,97%	26,20%	32,27%	38,67%	46,25%	29,32%	53,72%	40,71%	37,53%
Returns March/2020	-56,36%	-15,54%	-25,26%	-30,80%	-16,89%	-23,85%	-12,80%	-32,95%	-12,81%

Table 2: Results for Brazilian Data

References

[1] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22(5):1915–1953, 12 2007.

[2] Philippe Jorion. Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21:279–292, 09 1986.

[3] Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.

[4] Lorenzo Garlappi, Raman Uppal, and Tan Wang. Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *The Review of Financial Studies*, 20(1):41–81, 05 2006.

[5] Andrew Clark, Qiqiang Hou, Linda Bushnell, and Radha Poovendran. Maximizing the smallest eigenvalue of a symmetric matrix: A submodular optimization approach. *Automatica*, 95:446 – 454, 2018.

[6] Marcos López de Prado. Building diversified portfolios that outperform out-of-sample. *The Journal of Portfolio Management*, 42:59–69, 2016.

[7] Yves Choueifaty and Yves Coignard. Toward maximum diversification. *Journal of Portfolio Management*, 35:40–51, 10 2008.

[8] Yves Choueifaty, Tristan Froidure, and Julien Reynier. Properties of the most diversified portfolio. *SSRN Electronic Journal*, 2, 07 2011.

[9] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2 edition, 2012.

[10] Philippe Jorion and Richard J. Sweeney. Mean reversion in real exchange rates: evidence and implications for forecasting. *Journal of International Money and Finance*, 15(4):535 – 550, 1996.