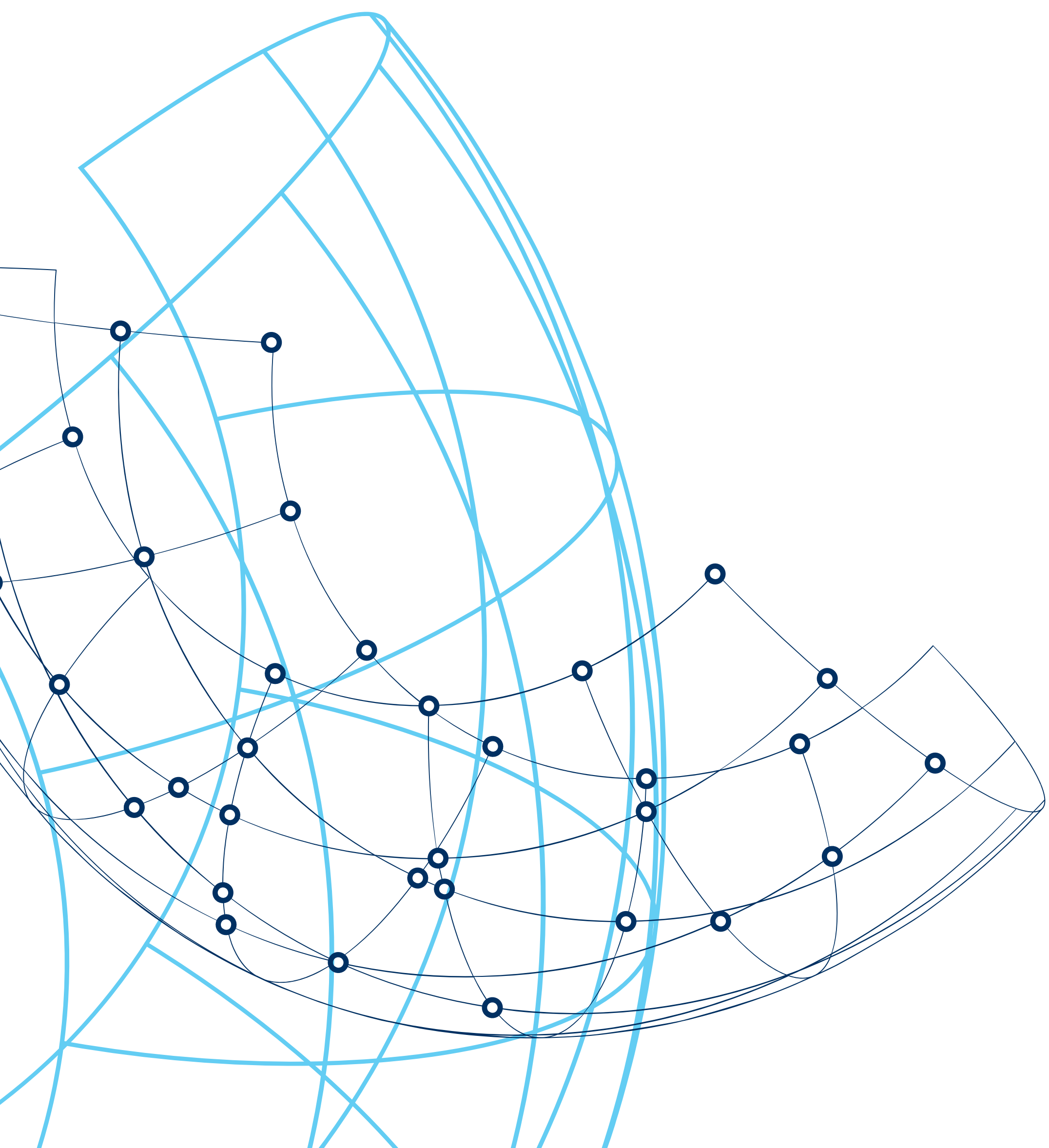
# Quantum-inspired hierarchical risk parity



White Paper

# Quantum-inspired hierarchical risk parity

Elham Alipour, Clemens Adolphs, Arman Zaribafiyan, and Maxwell Rounds

## Abstract

We present a quantum-inspired approach to portfolio optimization that is based on an optimization problem that can be solved using a quantum annealer. The proposed algorithm utilizes a hierarchical clustering tree that is based on the covariance matrix of the asset returns. We use real market data to benchmark our approach against other common portfolio optimization methods and demonstrate its strong performance in terms of a variety of risk measures and lower susceptibility to inaccuracies in the input data.

***Keywords***: risk management, portfolio optimization, hierarchical risk parity, quantum-inspired application, quantum annealing

## 1 Introduction

A central problem in finance is the optimal allocation of capital across a pool of assets. The mathematical framework for this problem is modern portfolio theory [1], which casts the problem as a quadratic programming problem with the goal of finding the best balance of return and risk of the portfolio under appropriate constraints (also known as mean-variance optimization). A number of efficient algorithms exist to solve such problems. Yet, while these methods provide an elegant mathematical solution to the portfolio optimization problem in the academic world of perfectly estimated risks and returns, these model assumptions fail to capture important practical considerations. For instance, the portfolio optimization in most quadratic programming methods requires the inversion of the covariance matrix and is therefore susceptible to inaccuracies in the input, which can stem from the noisy and uncertain nature of market data. In addition, there is no definite choice for the parameters that go into the calculation of the covariance matrix, such as the time frame of market data to be considered. It is thus desirable to find methods that are less sensitive to small changes in the input. As a result of these shortcomings, a vast collection of complementary improvements emerged in the finance literature [2, 3, 4, 5]. One example is inverse-variance parity (IVP), a version of the popular risk parity framework, which circumvents the problem of having to invert an ill-conditioned covariance matrix by taking only the variance of each asset into account and ignoring the covariance between assets. Of course, this introduces a new problem: to construct an optimal portfolio, it is important to correctly exploit the correlation between assets. The challenge, then, is to find a way to use this information while avoiding the stability issues that come along with the inversion of the covariance matrix.

The hierarchical risk parity (HRP) method introduced in Ref. [6] is designed to address the instability issues that occur with quadratic programming. The motivation for HRP is that assets such as stocks show hierarchical structure: stocks behave differently than bonds, stocks of small pharmaceutical companies will behave differently than stocks of large public utility companies, etc. This information is contained in the correlation matrix, which is used to construct a *distance* matrix to express how dissimilar two assets are. HRP then uses a particular clustering method[1] to build a *tree* of clusters. Each asset starts in its own cluster and as one moves up the hierarchical clustering tree, pairs of clusters are merged to form a larger cluster. In order to decide which two clusters to merge, a metric based on dissimilarities between assets is used. This structure can then be used to iteratively adjust the weights of the various clusters in the asset allocation, by condensing the information contained in the covariance matrix down into the cluster variance of each cluster and rebalancing weights in inverse proportion to their cluster variance. This allows for the exploitation of the information contained in the covariance matrix without inverting it.

In this white paper, we present a quantum-inspired version of HRP and achieve results that outperform standard HRP and conventional methods such as minimum-variance optimization (MV) in terms of minimizing risk. Our quantum-inspired HRP method is hardware agnostic, providing it the advantage of being solvable using either a quantum or classical solver.

This white paper is organized as follows. In Section 2 we describe our quantum-inspired versions of HRP (QHRP). Section 3 describes the process we use to benchmark the performance of the various portfolio optimization methods. Section 4 discusses benchmarking results. Finally, we give conclusions and provide an outlook in Section 5.

## 2 Quantum-inspired hierarchical risk parity

In this section, we describe our hierarchical algorithm for portfolio construction. To address some of the pitfalls of conventional methods, QHRP utilizes the information contained in the covariance matrix without inverting it by creating a hierarchy tree from it. This algorithm tends to result in a numerically stable answer even if the covariance matrix is singular. QHRP has two main steps: 1) building a hierarchical clustering tree and 2) updating the allocation weights based on the tree.

The goal of building a hierarchical clustering tree is to recursively divide the set of assets into two clusters in a way that minimizes the loss of information that results from ignoring the correlation between the two clusters. To achieve this goal, we first find a reordering, or permutation, of assets that makes the covariance matrix as quasi-block-diagonal as possible, where *quasi-block-diagonal* refers to a matrix where larger values tend to be closer to the diagonal than smaller values. To define a quasi-block-diagonal matrix, we introduce the measure

$$W(A) = \sum_{ij} A_{ij}(i-j)^2,$$

which is a weighting of a matrix's elements based on how far away from the diagonal they are. We then look for a *permutation* $\pi$ of the columns such that $W(A)$ is minimized for a similarity matrix $A$ based on the closeness of assets. For this problem, the elements of the similarity matrix $A$ are chosen to be the absolute values of the elements of the covariance matrix, $A_{ij} = |C_{ij}|$, since both strong positive and strong negative correlations between assets represent valuable information that we want to preserve. The intuition behind this is that the correlation between two assets should not be a function of whether each asset represents a long or a short position; as such, the statistic we use for clustering should be sign-invariant. By setting the similarity matrix's elements to be the absolute values of the elements of the covariance matrix, we minimize the information loss in the bisection step that follows.

Finding the optimal permutation is an NP-hard problem that can be formulated as a quadratic unconstrained binary optimization (QUBO) problem by introducing a total of $N^2$ binary variables

$$x_{ii'} = \begin{cases} 1 & \text{if } \pi(i) = i' \\ 0 & \text{otherwise} \end{cases},$$

with the constraint that for each asset $i$ there exists exactly one position $i'$ such that $x_{ii'} = 1$, and for each position $i'$

[1] Agglomerative single-linkage clustering

there exists exactly one asset $i$ such that $x_{ii'} = 1$. The objective function $W(C)$ can then be written as

$$W(C) = \sum_{ij} \sum_{i'j'} |C_{ij}|(i' - j')^2 x_{ii'} x_{jj'},$$

and the constraint can be converted into the penalty term

$$\sum_{i} \left(1 - \sum_{i'} x_{ii'}\right)^2 + \sum_{i'} \left(1 - \sum_{i} x_{ii'}\right)^2.$$

Combining the objective function and the constraints then leads to a fully connected QUBO problem of size $N^2$, which can be solved either with a quantum annealer such as the D-Wave 2X [7] or a classical solver.

After finding the permutation of assets that minimizes the quasi-block-diagonalization objective function, we recursively split the assets into two parts in the following way: at any given step, we try out all possible split points of assets $\vec{\pi} = (\vec{\pi}_1, \vec{\pi}_2)$. This splits the similarity $A$ into a $2 \times 2$ block-matrix. We then select the split that minimizes a suitably chosen metric, such as the mean absolute value of the off-diagonal blocks' entries. As shown in Fig. 1, the result of this process is a hierarchical clustering tree with the set of all of the assets as the root and the leaves consisting of all of the assets in their own clusters of size one.

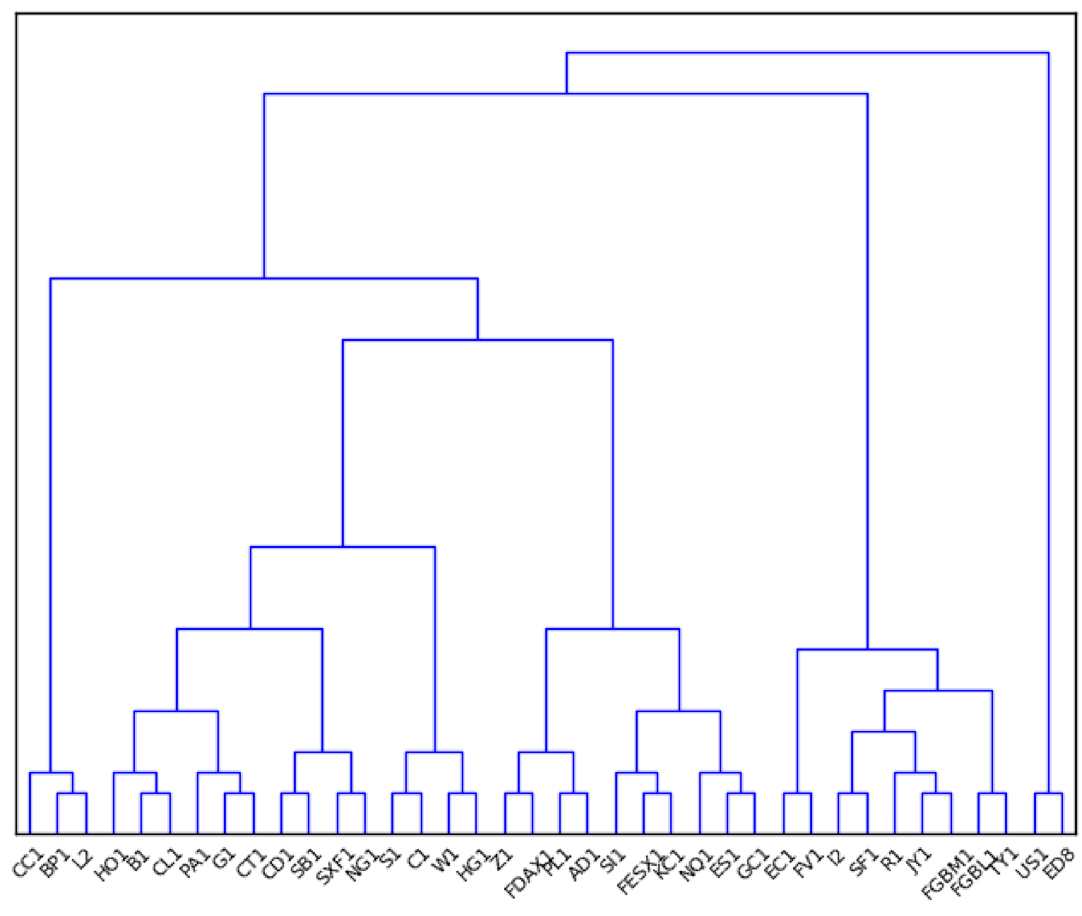



Figure 1: Example hierarchical clustering tree from 2014 for the CTA dataset.

The second step of QHRP is to recursively update the allocation weights based on the clustering tree obtained in the previous step. We begin by assigning a unit weight to all of the assets. Then, as one moves down the clustering tree, at each level of the tree, the weights of assets in each cluster are rescaled in inverse proportion to the cluster's variance. The variance of a cluster is defined as

$$V = w_i C_{ij} w_j,$$

where $V$ is the variance of the cluster, $C_{ij}$ is the covariance between constituents of the cluster, and $w_i = C_{ii}^{-1}/\text{tr}(\text{diag}(C)^{-1})$. Note that this algorithm preserves the general conditions for portfolio optimization such as the weight of each asset being non-negative[2] and smaller than one, and also that all of the weights sum up to one.

As mentioned, the number of binary variables in this QUBO formulation is $N^2$, with $N$ being the number of assets. Since current quantum annealers can only solve this problem for very small-sized problems (six assets at most), our experiment uses a heuristic method (bandwidth reduction) [8] that approximates the solving of this QUBO problem and is suitable for larger sizes. We look forward to implementing this problem on a quantum annealer as the technology continues to mature.

## 3 Benchmarking

In this section, we describe the methods used to benchmark various portfolio optimization methods. We backtest against real historic price data of suitably chosen asset universes and use bootstrapping to perform statistical analysis.

### 3.1 Asset universes

To highlight the ability of our method to exploit the hierarchical structure of assets, we focus our benchmarking on a universe of thirty-eight diverse futures contracts, including stocks and bonds from different countries, as well as commodities such as oil, wheat, and gold. We call this the CTA portfolio, because it is a fairly typical portfolio for a commodity trading advisor (CTA) to manage.

Some additional testing is done on the thirty stocks that make up the Dow Jones Industrial Average (DJIA). These are all well-established, large-cap stocks of US companies. They are all much more strongly correlated than the CTA assets, and thus allow us to compare the performance of our method in the two extremes with respect to the heterogeneity of the asset universe. There is one caveat for the DJIA: its composition is not constant over time. Stocks enter and leave the DJIA and this must be taken into account during benchmarking to avoid *survivorship bias*: during each rebalancing step, only those stocks that make up the DJIA *at that date* are allowed to make up the portfolio going forward from that date.

### 3.2 Bootstrapping

In performing a backtest, we are not interested only in the results that are based on the single real-world price history of the assets but also of statistical descriptors such as the standard deviation, hereafter referred to as volatility, since this expresses how reliable the results are and how sensitive the method is to small deviations in the input data. Since it is not possible to know the exact distribution from which the returns are drawn, we use bootstrapping [9]. This is the process of obtaining a probability distribution from an observed sample.

In essence, the historical returns form a time series

$$\vec{x}_1, \vec{x}_2, \vec{x}_3, \ldots, \vec{x}_N,$$

where each $\vec{x}_t$ contains the logarithmic daily returns of all the assets at time step $t$. To generate a *bootstrapped* sample out of this time series, one picks $N$ random variables $r_i$ sampled uniformly from $\{1, \ldots, N\}$, with replacement. That is, there may exist some $i, j$ for which $r_i = r_j$. The bootstrapped sample is then the time series

$$\vec{x}_{r_1}, \vec{x}_{r_2}, \ldots, \vec{x}_{r_N}.$$

This process is then repeated several times (typically 1000 or 10,000). It can be shown that the bootstrapped samples share the statistical properties of the original sample [9]. In particular, correlations between groups of stocks are

[2] Negative weights would indicate short-selling an asset, which we do not consider here, although this is a potential area for future research.

preserved. The original sample plus the bootstrapped samples then provide an empirical probability distribution that can be used to compute statistics of interest.

We numerically tested the performance of different portfolio optimization methods on the historical returns of the DJIA and CTA indices for the period 2005–2016. The benchmarking works as follows:

1. The *rebalancing date* is the first business day of each month.[3]
2. At each rebalancing date, we look back at a one-year rolling window of daily returns.
3. We perform bootstrapping as described above on this window of daily logarithmic returns to generate 1000 additional time series of logarithmic returns.
4. The portfolio optimization method of choice is applied to all of these time series, both original and bootstrapped.
5. For each of the time series, we obtain a set of weight allocations for the assets that—on the rebalancing date—make up the index.
6. The performance of each portfolio, whether based on the real returns or the bootstrapped returns, is recorded for the month following the rebalancing date, up to the last business day of that month. Note: no bootstrapping takes place on these out-of-sample returns. It is not necessary here, as these returns are used only to evaluate portfolio performance and are not in themselves an input to generating the portfolio.
7. The process is repeated for each of the rebalancing dates until we have daily returns for each portfolio over the entire observation period.
8. The daily returns are then used to compute a number of risk and performance measures, such as annualized volatility. These are described in detail in the next section.
9. When computing the risk measures, the same measure is computed for each of the time series, based on both the real data and the bootstrapped samples. This results in a large sample for the measure, from which we can calculate statistical descriptors. In particular, we want to show error bars for each quantity we report, and using the standard deviation over all of the bootstrapped samples allows us to do so.

## 4 Results

We benchmarked the performance of QHRP against the following portfolio optimization methods: minimum variance (MV, using the open source implementation provided in Ref. [10]), inverse-variance portfolio (IVP), and Hierarchical Risk Parity (HRP). Note that for QHRP, we approximate the solving of the objective function using a heuristic method that is presented in Ref. [8].

First we present the results of benchmarking on the CTA portfolio data. The annualized volatility of portfolio returns is one of the most common risk measures. In Fig. 2, the annualized volatility for CTA is plotted as a function of time for different methods. As can be seen from this plot, QHRP outperforms all of the conventional methods tested on this dataset in almost all years. In particular, the annualized volatility of MV is almost three times the annualized volatility of QHRP in almost every year. Note that even though MV minimizes the volatility in-sample, its out-of-sample performance is rather poor due to the noisy nature of financial data. In addition, the high condition number of the covariance matrices for this particular asset pool (on the order of $10^4$–$10^5$) makes MV very unstable, as it uses the inverse of the covariance matrix. Moreover, the performance of IVP and HRP are, on average, $34\%$ and $16\%$ worse than QHRP, respectively. This suggests that QHRP preserves more information from the covariance matrix and can therefore provide portfolios with lower risk. We also looked at semi-variance as the risk measure. Semi-variance is the expectation of the squared deviation from the mean from below and is used to quantify downside risk. The results of benchmarking for this risk measure were similar to the annualized volatility and further demonstrate the advantage of QHRP.

[3] CTA contains assets from various markets with different holiday calendars; days for which there are no returns are discarded.

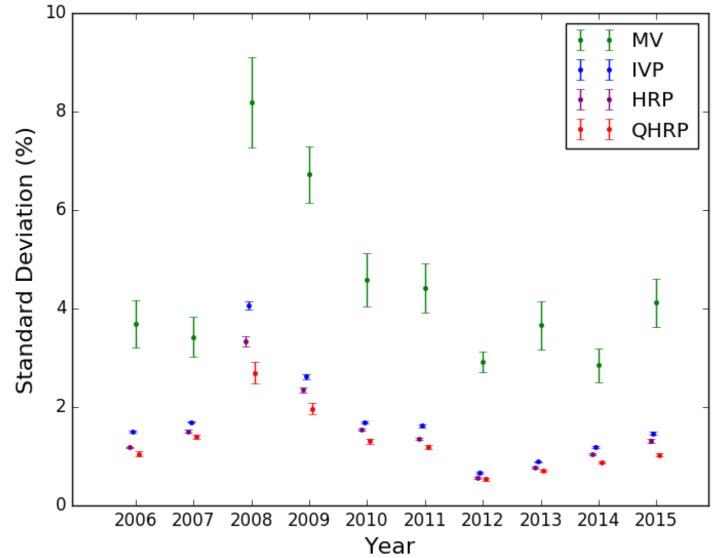



Figure 2: The annualized volatility of the CTA returns as a function of time for MV, IVP, HRP, and QHRP.

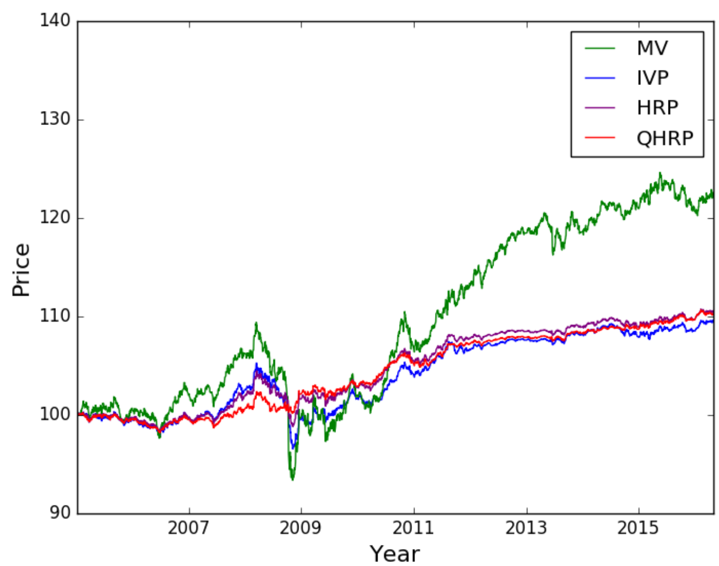



Figure 3: The average price of portfolios provided by MV, IVP, HRP, and QHRP as a function of time on the CTA index. The initial price has been normalized to 100.

Another interesting risk measure to look into is the maximum drawdown. Maximum drawdown measures the largest drop in an investment's value from its historical peak over a given period. In other words, it represents the realized worst-case scenario from the perspective of an investor who invested near the market top. In Fig. 3, we plot the average price of portfolios constructed by different algorithms over all bootstrapped samples for CTA. The initial price is normalized to 100. This figure shows that even though the MV portfolio has appreciated more in recent years, it is overall much more volatile than the QHRP portfolio. Specifically, the MV portfolio shows a significant maximum drawdown in 2008, the year of the global financial crisis. We present the overall maximum drawdown of all of the portfolio optimization methods in Table 1.

The last risk measure that we looked at is the Sharpe ratio, which is a measure for calculating risk-adjusted returns. In other words, it measures the expected return per unit of risk. In Fig. 4, the annualized Sharpe ratio for the CTA portfolio of different methods is plotted. As can be seen for this risk measure, there is no one method that outperforms the others in all the years. However, QHRP proves advantageous when there is a shock in the system. For example the QHRP portfolio is the only one with a positive annualized Sharpe ratio in the year 2008, where most of the assets experienced a collapse. In Table 1, we show the Sharpe ratio for different methods computed over the entire time period. For this range, QHRP clearly outperforms the other methods.

Table 1: The maximum drawdown and Sharpe ratio of portfolios constructed by MV, IVP, HRP, and QHRP methods on CTA index.

| Method | Maximum Drawdown (%) | Sharpe Ratio |
|---|---|---|
| MV | $14.9 \pm 2.7$ | $0.51 \pm 0.16$ |
| IVP | $8.3 \pm 0.4$ | $0.54 \pm 0.03$ |
| HRP | $5.0 \pm 0.6$ | $0.70 \pm 0.06$ |
| QHRP | $2.5 \pm 0.9$ | $0.82 \pm 0.13$ |

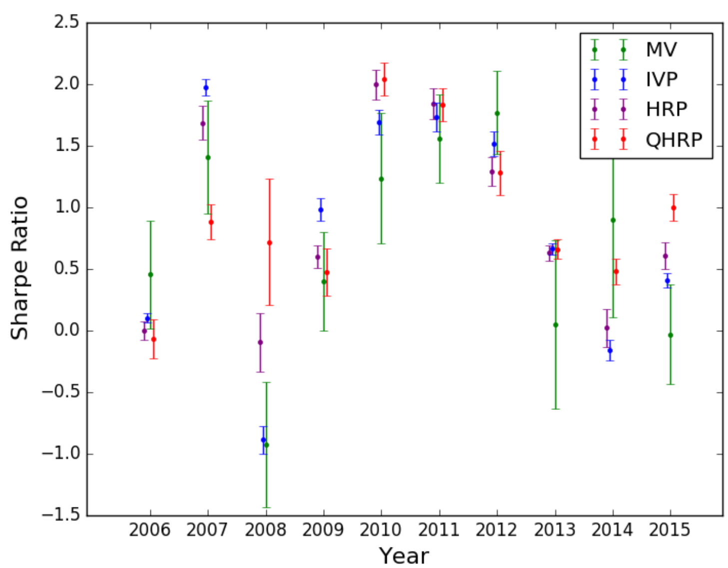



Figure 4: The annualized Sharpe ratio of the CTA returns as a function of time for MV, IVP, HRP, and QHRP.

We also benchmarked the performance of QHRP against alternative approaches on the DJIA dataset. We chose DJIA as we expect MV to perform well on this particular dataset. The reason is that the covariance matrices of this dataset do not have high condition numbers (typically around a few hundred). Therefore, the inversion of these matrices does not pose instability issues. In Fig. 5, the annualized volatility for portfolios constructed on DJIA index is plotted as a function of time. Note that the performance of MV on this dataset is better than that of IVP and HRP. We believe the reason for this behaviour is that the DJIA index is more favourable for MV than for the other competitors. The covariance matrices

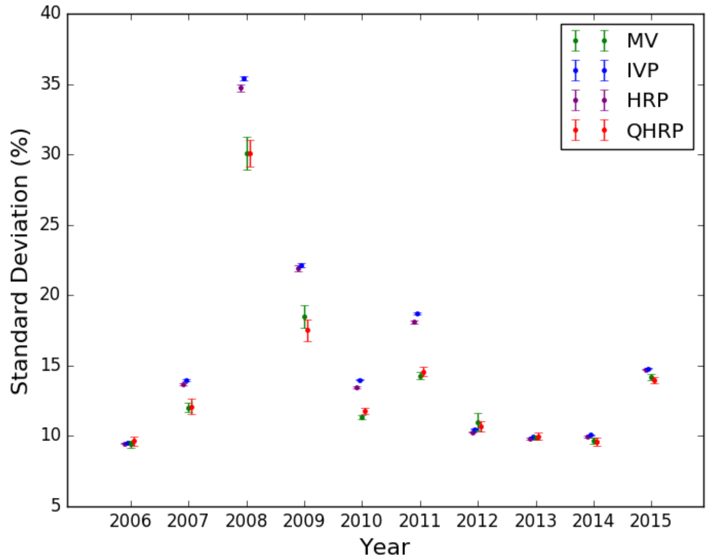



Figure 5: The annualized volatility of the DJIA returns as a function of time for MV, IVP, HRP, and QHRP.

on this dataset are not ill-conditioned and there is little hierarchical structure in the dataset, as all thirty assets in the DJIA are very large, publicly owned companies based in the US that have very high correlation amongst them. However, the performance of QHRP is similar to the performance of MV. This shows that even on a dataset with limited hierarchical structure, the performance of QHRP is reasonably good.

Note that since the solution of QHRP is obtained using a heuristic method, we can expect better performance for all of the risk measures on both datasets if we obtain the true solution to this optimization problem, using a device such as a quantum annealer.

## 5 Conclusion

Although there are many conventional methods that provide theoretically correct solutions for the portfolio optimization problem, most of them exhibit many shortcomings when dealing with real data. Minimum-variance portfolio optimization, one of the well-known methods in the literature, generally provides less reliable solutions. One of the main reasons for this is that MV requires the inversion of a covariance matrix, which can lead to numerically unstable solutions if the covariance matrix is ill-conditioned. In addition, the out-of-sample performance of this method tends to be inferior, even though it minimizes the risk in-sample. On the other end of the spectrum, traditional risk parity methods such as inverse-variance parity provide numerically stable portfolios. However, they fail to make use of the important information contained in the correlations between assets.

In this paper, we presented a new quantum-inspired approach for portfolio optimization that addresses the shortcomings of conventional methods. In particular, QHRP uses the correlations between assets to build a hierarchical structure that is based on the information contained in the covariance matrix but which does not require its inversion. Then, using this clustering tree, it recursively updates the allocation weights until we obtain a diversified portfolio. The method used in QHRP for building hierarchical clustering trees has many advantages. First, it minimizes the loss of information that is caused by ignoring the correlation between clusters, leading to better performance for the portfolio optimization. Second, it is a hardware-agnostic approach, which means that it can be solved using either a quantum annealer or a classical solver. Finally, our benchmarking results show that QHRP outperforms alternative approaches for a variety of risk measures.

We benchmarked the performance of QHRP on a diverse futures index constructed from a variety of underlying assets (CTA index). The results show that QHRP outperforms all of the alternative methods tested (MV, IVP, and HRP) on this dataset for a variety of risk measures. In particular, the annualized volatilities of portfolio returns for MV, IVP, and HRP are on average $275\%$, $34\%$, and $16\%$, respectively, higher than the annualized volatility of QHRP. We also performed additional tests on the DJIA that are more favourable for MV. The reason for this is twofold: the covariance matrices on this dataset typically have low condition numbers, and so their inversion is more stable; and there is only weak hierarchical structure in the data. The results show that QHRP outperforms both IVP and HRP by a significant margin in terms of risk reduction, but its performance is the same as MV on the DJIA index. This shows that even on a dataset that is more favourable for MV, the performance of QHRP is comparable.

## Acknowledgements

We would like to thank Phil Goddard and Daniel Bezdek for providing useful comments, Marko Bucyk for editing the manuscript, and Gili Rosenberg for helpful discussions and for reviewing an earlier draft of this white paper.

## References

[1] Harry Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[2] Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621, 2003.

[3] William W. Hogan and James M. Warren. Toward the development of an equilibrium capital market model based on semivariance. *Journal of Financial and Quantitative Analysis*, 9:1–11, 1974.

[4] R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[5] Hiroshi Konno and Hiroaki Yamazaki. Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management science*, 37(5):519–531, 1991.

[6] Marcos López de Prado. Building diversified portfolios that outperform out-of-sample. *Journal of Portfolio Management*, 2016.

[7] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, et al. Quantum annealing with manufactured spins. *Nature*, 473(7346):194–198, 2011.

[8] Murat Manguoglu, Mehmet Koyutürk, Ahmed H. Sameh, and Ananth Grama. Weighted matrix ordering and parallel banded preconditioners for iterative linear system solvers. *SIAM Journal of Scientific Computing*, 32(3):1201–1216, 2010.

[9] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman and Hall/CRC, 1993.

[10] David H. Bailey and Marcos López de Prado. An open-source implementation of the critical-line algorithm for portfolio optimization. *Algorithms*, 6:169–196, 2013.

Table 2: Futures Contracts in the CTA Universe

| Ticker | Commodity | Description | Category |
|---|---|---|---|
| C1 | Corn | Corn | Agricultural |
| CC1 | Cocoa | Cocoa | Agricultural |
| CT1 | Cotton | Cotton | Agricultural |
| KC1 | Coffee | Coffee | Agricultural |
| S1 | Soybeans | Soybeans | Agricultural |
| SB1 | Sugar No. 11 | Sugar | Agricultural |
| W1 | Wheat | Wheat | Agricultural |
| FGBL1 | Euro-Bund | 10-Year German Govt Bond | Bonds |
| FGBM1 | Euro-Bobl | 5-Year German Govt Bond | Bonds |
| FV1 | 5-Year US Treasury Note | 5-Year USA Govt Bond | Bonds |
| R1 | Long Gilt | 10-Year British Govt Bond | Bonds |
| TY1 | 10-Year US Treasury Note | 10-Year USA Govt Bond | Bonds |
| US1 | 30-Year US Treasury Note | 30-Year USA Govt Bond | Bonds |
| AD1 | Australian Dollar | Australian Dollar | Currency |
| BP1 | British Pound | British Pound | Currency |
| CD1 | Canadian Dollar | Canadian Dollar | Currency |
| EC1 | Euro | Euro | Currency |
| JY1 | Japanese Yen | Japanese Yen | Currency |
| SF1 | Swiss Franc | Swiss Franc | Currency |
| B1 | Brent Crude Oil | European Oil | Energy |
| CL1 | WTI Crude Oil | USA Oil | Energy |
| G1 | Gasoil | Fuel Oil | Energy |
| HO1 | Heating Oil | Heating Oil | Energy |
| NG1 | Natural Gas | Natural Gas | Energy |
| GC1 | Gold | Gold | Metals |
| HG1 | Copper | Copper | Metals |
| PA1 | Palladium | Palladium | Metals |
| PL1 | Platinum | Platinum | Metals |
| SI1 | Silver | Silver | Metals |
| ED8 | Eurodollar | 3m US Short Term Rate | Interest Rate |
| I2 | EURIBOR | 3m European Short Term Rate | Interest Rate |
| L2 | Short Sterling | 3m British Short Term Rate | Interest Rate |
| ES1 | S&P 500 | USA Stock Index | Stocks |
| FDAX1 | DAX | German Stock Index | Stocks |
| FESX1 | Euro Stox 50 | European Stock Index | Stocks |
| NQ1 | NASDAQ 100 | USA Stock Index | Stocks |
| SXF1 | S&P/TSX 60 | Canadian Stock Index | Stocks |
| Z1 | FTSE 100 | British Stock Index | Stocks |