

Distributionally Robust Risk Parity Portfolios with Wasserstein Distances

Mohammad Rafay Kalim

Contents

1	Introduction	2
2	Background	3
2.1	Modern Portfolio Theory	3
2.2	Risk Parity	4
2.3	Robust Optimization	5
2.3.1	Robust MVO and Risk Parity	6
2.3.2	Distributionally Robust Risk Parity	7
2.3.3	The Wasserstein Distance	7
2.4	Differentiable Optimization Layers	7
3	My Contribution	8
3.1	Wasserstein Distance Dynamically Robust RP	8
3.2	End-To-End δ Selection	8
4	Next Steps	9

1 Introduction

The only certainty in world of investments is the lack of certainty in information. There is a large body of research in the field of portfolio optimization with goal of ingesting uncertain, estimated data to make optimal investment decisions.

Mean Variance Optimization (MVO), pioneered by Markowitz in 1952, was the first such attempt to create mathematically optimal investment decisions [?]. This techniques aims to balance the tradeoff between risk and reward, subject to the investor’s needs. Contrary to MVO, Risk Parity (RP) portfolios have the objective of ensuring an equal risk contribution from each asset, with no explicit reward goal [?]. The uncertainty of the risk and reward measures introduce significant errors making the output of RP sensitive to input data. Recent advances in Distributionally Robust Optimization have led to Distributionally Robust Risk Parity (DRRP), which introduces an uncertainty set around the probability distribution of historical observations used as inputs to RP [1]. This allows the RP model to be less sensitive to errors in the input estimates. This thesis will introduce a new version of DRRP, called DRRP-W (for Wasserstein) which assumes a uniform probability distribution for our observed historical returns, but introduces an uncertainty set around the return observations themselves using the Wasserstein Distance. Computational experiments will be run to compare this new model against DRRP, Robust MVO and Robust Risk Parity. Further, this thesis will propose an end-to-end optimization approach following [?] to optimize the choice of δ , the size of the uncertainty set.

This report is structured as follows: Section 2 contains an overview of existing literature and preliminaries to begin exploring the DRRP-W problem. Section 3 provides an update on progress made so far, and Section 4 outlines next steps and concludes.

2 Background

2.1 Modern Portfolio Theory

Many portfolio optimization techniques, such as Mean Variance Optimization (MVO) introduced by Markowitz in Modern Portfolio Theory, focus on the tradeoff between risk and reward [?]. The goal of MVO and all other portfolio optimization techniques is to obtain $\phi \in \mathbb{R}^n$, a vector which contains the optimal portfolio allocations for each of the n assets where ‘optimal’ is defined by the user’s objective. Further, this also requires the estimation of μ , the reward measure, and Σ , the risk measure. Once we have a risk and reward measure, we can formulate an optimization problem to obtain ϕ . For example, the risk-return tradeoff (the objective) variation of MVO can be formulated as

$$\begin{aligned} \min_{\phi} \quad & \phi^T \Sigma \phi - \lambda \mu^T \phi \\ \text{s.t.} \quad & \mathbf{1}^T \phi = 1, \\ & (\phi \geq 0) \end{aligned} \tag{1}$$

This can be thought of as minimizing our portfolio risk, as measured by the variance of the portfolio returns ($\phi^T \Sigma \phi$) while maximizing our portfolio reward ($\lambda \mu^T \phi$). λ here is a user-defined hyperparameter which quantifies the risk-aversion of the user by amplifying the reward measure. A high value for λ means that the user places a heavy (negative) weight on a high return, and thus have a high risk tolerance due to the fact that this is a minimization problem.

A natural question to ask is how μ and Σ are estimated. A naive method is by taking the average of observed historical data. Let $\hat{\xi} \in \mathbb{R}^{n \times T}$ be the matrix of observed historical returns, with T historical occurrences and n assets. Then, the estimate of the expected asset return for asset i is

$$\hat{\mu}_i = \mathbb{E}[\xi] = \frac{1}{T} \sum_{k=1}^T \hat{\xi}_k \tag{2}$$

Similarly, our risk measure $\Sigma \in \mathbb{R}^{n \times n}$ is known as the asset covariance matrix, where $\sigma_{i,j}$ can be computed as:

$$\sigma_{ij} = \frac{1}{T-1} \sum_{t=1}^T (\hat{\xi}_{it} - \hat{\mu}_i)(\hat{\xi}_{jt} - \hat{\mu}_j) \tag{3}$$

However, as the number of assets n increases, we can see that the number of parameters to be estimated increases by $O(n^2)$. To see this, first note that Σ has n^2 entries. Since Σ is a covariance matrix, it must be symmetric. Thus, the number of unique elements is $\frac{n(n-1)}{2}$ (from the upper right section of the matrix) and n (from the leading diagonal). Finally, $O(n + \frac{n(n-1)}{2}) = O(n^2)$. In real world asset allocation problems, portfolio managers may have to choose optimal allocations amongst thousands of stocks, which can lead to a matrix with over 1,000,000 elements to be estimated. This can be very computationally expensive.

To overcome these computational challenges, many practioners use factor models, first introduced by Fama and French [?]. Factor models are usually linear models created through linear regression with the goal of explaining the return of an asset through a set of factors. These factors are nothing but the explanatory variables (or features) in a linear regression model. Now, we can introduce some notation around factor models. Ordinary Least Squares (OLS) theory provides us a closed form solution to the regression problem. Assuming p factors, let

1. $X = (\mathbf{1} \ f)$ be the data matrix, where $f \in \mathbb{R}^{T \times p}$ is the matrix of observed factor returns

2. $\mathbf{R} \in \mathbb{R}^{T \times n}$ be a matrix of observed historical returns for all assets
3. $\boldsymbol{\alpha} \in \mathbb{R}^n$ be the vector of intercept terms for each asset
4. $\mathbf{V} \in \mathbb{R}^{p \times n}$ be the matrix of coefficients to each factor for each asset

Then, the closed form solution to find \mathbf{B} , the factor coefficients, is

$$\mathbf{B} = \begin{pmatrix} \boldsymbol{\alpha}^T \\ \mathbf{V} \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R} \quad (4)$$

However, our original problem is to get an estimate of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. To that end, we can easily see that by the definition of linear regression these can be represented as

$$\boldsymbol{\mu} = \boldsymbol{\alpha} + \mathbf{V}^T \hat{\mathbf{f}} \quad (5)$$

and

$$\boldsymbol{\Sigma} = \mathbf{V}^T \mathbf{F} \mathbf{V} + \mathbf{D} \quad (6)$$

Here, $\hat{\mathbf{f}} \in \mathbb{R}^p$ is a vector of expected factor returns, $\mathbf{F} \in \mathbb{R}^{p \times p}$ is the factor covariance matrix and $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a diagonal matrix of residual variances, obtained through regression.

These estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can now be used more efficiently with MVO to get portfolios faster, and with a smaller potential for computational errors.

It is important to note that while factor models allow optimization over a larger universe of assets and improve computational efficiency, there are several drawbacks. The most important drawback is that of estimation error, which is main the topic of research amongst the world's largest asset managers. The handling of estimation errors in the inputs to our optimization problems is the focus of this thesis.

2.2 Risk Parity

Our MVO problem, with the minimization objective function of $\boldsymbol{\phi}^T \boldsymbol{\Sigma} \boldsymbol{\phi} - \lambda \boldsymbol{\mu}^T \boldsymbol{\phi}$ was just one example of a portfolio optimization problem. The MVO formulation is ideal for most use cases where users want to balance risk and reward. However, different investors can have different objectives. One possible goal of an investor is that of risk parity. Risk parity portfolios can be thought of as equal risk contribution portfolios, wherein each asset contributes the same amount of risk towards the overall portfolio risk. A simple analogy to risk parity is the equally weighted portfolio. Here, for n assets, each asset has a weight of $\frac{1}{n}$. In this simple portfolio, we are imposing wealth parity, as opposed to risk parity.

To formulate risk parity as an optimization problem, we first need to decompose the portfolio risk. Here, we are once again using portfolio variance as a measure of risk.

We first introduce Euler's Theorem for positive homogenous functions. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is called a positive homogenous function of degree k if, $\forall c \geq 0$ and $\mathbf{x} \in \mathbb{R}^n$,

$$f(c \cdot \mathbf{x}) = c^k f(\mathbf{x}) \quad (7)$$

Theorem 1 (Euler's Homogenous Function Theorem) *A function f is continuously differentiable positively homogenous function of degree k if and only if*

$$k \cdot f(\mathbf{x}) = \mathbf{x}^T \nabla f(\mathbf{x})$$

Theorem 2 *The portfolio variance, $\sigma_p^2 = \boldsymbol{\phi}^T \boldsymbol{\Sigma} \boldsymbol{\phi}$, is a positive homogenous function of degree 2.*

Our risk measure is the portfolio variance. The marginal variance contribution of an asset i is

$$\frac{\partial \sigma_p^2}{\partial \phi_i} = 2(\Sigma \phi)_i$$

Where $(\Sigma \phi)_i$ is the i^{th} element of the vector $\Sigma \phi$. By using 1 and 2, we can represent σ_p^2 as

$$\sigma_p^2 = \phi^T \Sigma \phi = \frac{1}{2} \sum_{i=1}^n 2\phi(\Sigma \phi)_i = \sum_{i=1}^n \phi(\Sigma \phi)_i$$

For ease of notation, we can let $R_i = \phi(\Sigma \phi)_i$, which represents the risk contribution of an asset to the total portfolio risk. Thus, the problem of risk parity becomes finding a portfolio where $R_i = R_j \forall i, j$.

To formulate risk parity, consider the function

$$f(y) = \frac{1}{2} y^T \Sigma y - c \sum_{i=1}^n \ln(y_i)$$

for some decision variable y . Note that Σ is a PSD matrix (as it is a covariance matrix) and $\ln(\cdot)$ is strictly concave. Thus, $f(y)$ is a strictly convex function. So, we can try to find the value that minimizes f by taking the gradient and setting it to 0. To that end, we find

$$\nabla f(y) = \Sigma y - c \frac{1}{y} = 0.$$

Which further implies

$$y_i(\Sigma y)_i = R_i = c, \forall i$$

Note that since $R_i = c$ for all assets, the decision variable y holds the weights of a risk parity optimal portfolio. Note that we require that $y > 0$ due to the logarithm. Hence, the risk parity optimization problem is formulated as

$$\begin{aligned} \min_y \quad & \frac{1}{2} y^T \Sigma y - c \sum_{i=1}^n \ln(y_i) \\ \text{s.t.} \quad & y > 0 \end{aligned} \tag{8}$$

And we can find our optimal asset weights by

$$\phi_i^* = \frac{y_i^*}{\sum_{i=1}^n y_i^*}$$

2.3 Robust Optimization

In our section on **Modern Portfolio Theory** we noted that our estimates of μ and Σ are prone to large estimation errors. This is natural and to be expected, as parameters estimated from historical data will always have some degree of uncertainty. To handle this inevitable uncertainty in our estimates, we can try to make our optimization problems robust to small perturbations to the input data. This is the main idea behind the concept of robust optimization. More concretely, given a canonical linear program with m constraints and n decision variables

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A} \mathbf{x} > \mathbf{b} \end{aligned} \quad (9)$$

Robust optimization, with uncertainty on each of \mathbf{c} , \mathbf{A} and \mathbf{b} , instead solves

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A} \mathbf{x} > \mathbf{b} \forall (\mathbf{c}, \mathbf{A}, \mathbf{b}) \in \mathcal{U} \end{aligned} \quad (10)$$

Where $\mathcal{U} \in \mathbb{R}^n \times \mathbb{R}^{m \times n} \times \mathbb{R}^m$ defines the uncertainty set around each nominal estimate of $(\mathbf{c}, \mathbf{A}, \mathbf{b})$.

2.3.1 Robust MVO and Risk Parity

Let us now look at the robust counterparts of MVO and risk parity. For MVO, we consider the target return variation. The nominal MVO is

$$\begin{aligned} \min_{\boldsymbol{\phi}} \quad & \boldsymbol{\phi}^T \boldsymbol{\Sigma} \boldsymbol{\phi} \\ \text{s.t.} \quad & \boldsymbol{\mu}^T \boldsymbol{\phi} \geq R, \\ & \mathbf{1}^T \boldsymbol{\phi} = 1, \\ & (\boldsymbol{\phi} \geq 0) \end{aligned} \quad (11)$$

Here we will focus on an uncertainty set around the expected return estimate $\boldsymbol{\mu}$. The general form of \mathcal{U} is given by

$$\mathcal{U}_{\boldsymbol{\mu}} = \{x : D(\mu_i, x) \leq \delta_i, i = 1..n\} \quad (12)$$

Where $D(a, b)$ is some distance measure and δ_i is the size of our uncertainty set centered at our estimate μ_i . This now gives us two hyperparameters that we can select. First, the distance measure. For example, we can use an L1 norm to get the following uncertainty set

$$\mathcal{U}_{\boldsymbol{\mu}} = \{x : |\mu_i - x| \leq \delta_i, i = 1..n\} \quad (13)$$

Incorporating this into our nominal MVO, we get the following robust MVO formulation

$$\begin{aligned} \min_{\boldsymbol{\phi}} \quad & \boldsymbol{\phi}^T \boldsymbol{\Sigma} \boldsymbol{\phi} \\ \text{s.t.} \quad & \boldsymbol{\mu}^T \boldsymbol{\phi} - \delta^T |\boldsymbol{\phi}| \geq R, \\ & \mathbf{1}^T \boldsymbol{\phi} = 1, \\ & (\boldsymbol{\phi} \geq 0) \end{aligned} \quad (14)$$

where $\delta_i = \frac{\epsilon_1 \sigma_i}{\sqrt{T}}$.

This problem is more difficult for risk parity. Note that in the risk parity formulation, we did not have a expected return vector $\boldsymbol{\mu}$. Costa (2020) proposes a formulation for robust risk parity [1]. This robust formulation introduces an uncertainty set around the covariance matrix $\boldsymbol{\Sigma}$. The optimization problem becomes a second order cone program (SOCP) and can be written as

$$\begin{aligned}
& \min_{\boldsymbol{\phi}, \mathbf{z}, u, v, \zeta} && u - v \\
& \text{s.t.} && \mathbf{1}^T \boldsymbol{\phi} = 1, \\
& && \|\Sigma^\Delta \boldsymbol{\phi}\|_2 \leq \sqrt{n} \zeta, \\
& && \Omega \zeta \leq [\hat{\Sigma} \boldsymbol{\phi}]_i - z_i, i = 1..n, \\
& && \left\| \begin{bmatrix} 2v \\ x_i - z_i \end{bmatrix} \right\|_2 \leq x_i + z_i, i = 1..n, \\
& && \|(\hat{\Sigma} + \Sigma^\Delta)^{1/2} \boldsymbol{\phi}\|_2 \leq \sqrt{n} u, \\
& && \boldsymbol{\phi}, \mathbf{z} \geq 0, \\
& && u, v, \zeta \geq 0
\end{aligned} \tag{15}$$

The interesting constraint here is that of $\Omega \zeta \leq [\hat{\Sigma} \boldsymbol{\phi}]_i - z_i, i = 1..n$. As this constraint tightens, it reduces the size of the error term which diminishes the portfolio's exposure to assets with larger estimation error in their marginal risk contributions. Computational experiments show that this robust formulation can achieve around 6% better returns as compared to nominal risk parity portfolios.

2.3.2 Distributionally Robust Risk Parity

2.3.3 The Wasserstein Distance

2.4 Differentiable Optimization Layers

3 My Contribution

3.1 Wasserstein Distance Dynamically Robust RP

We defined a naive estimate for $\hat{\mu}$ in equation 2. It is interesting to note that this implicitly assumes that our historical observations of returns follow a uniform distribution, as the expected value is simply the average. Costa (2022) challenges this assumption to introduce an uncertainty set around \mathbb{P} , the probability distribution of our historical observations[2]. Being a probability distribution, the general form for \mathbb{P} is given by

$$\mathbb{P} = \{p \in \mathbb{R}_+^T : \mathbf{1}^T p = 1\}$$

Costa assumes the nominal distribution is given by $p = 1/T$, and defines the ambiguity set around this nominal distribution using distance measures such as K-L divergence. The resulting DRRP problem can be solved using an algorithm that they define. However, this algorithm is computationally expensive and requires domain-specific knowledge to solve. In our work, we will take a different approach. Instead of defining an ambiguity set around \mathbb{P} , we define an ambiguity set around each observation $\hat{\xi}$.

Let us begin by defining the portfolio return and portfolio variance under an asset allocation decision ϕ . The portfolio return is given by

$$\mu_\phi = \phi^T \hat{\mu} = \phi^T \hat{\xi} \mathbf{1}/T = \mathbf{1}^T / T \hat{\xi}^T \phi$$

Where $\mathbf{1} \in \mathbb{R}^T$. Further, the portfolio variance can be written as

$$\Sigma_\phi = \phi^T \hat{\Sigma} \phi = \mathbb{E}^\mathbb{P}[(\xi^T \phi - \mathbb{E}[\xi^T \phi])^2]$$

Using the formula for variance, this can be simplified as

$$\Sigma_\phi = \mathbb{E}^\mathbb{P}[(\xi^T \phi)^2] - (\mathbb{E}^\mathbb{P}[\xi^T \phi])^2$$

With noting that $(\xi^T \phi)^2 = (\xi^T \phi)^T (\xi^T \phi) = \phi^T \xi \xi^T \phi$ and that $\phi^T \xi = \xi^T \phi$, we can re-write the above to get

$$\Sigma_\phi = \phi^T \mathbb{E}^\mathbb{P}[\xi \xi^T] \phi - \alpha^2$$

Where $\alpha = \mathbb{E}^\mathbb{P}[\phi^T \xi]$. Now, we can re-write our distributionally robust risk parity problem using these new terms to get

3.2 End-To-End δ Selection

Blanchet, Chen and Zhou (2018) explore distributionally robust mean-variance portfolios using the Wasserstein distance [?]. They are able to find a convex problem which protects portfolios against variations in μ . Their paper provides a good structure that I can follow to extend the idea to Risk Parity portfolios. They also provide an exposition on the critical choice of δ , which is based in 3 assumptions on the distribution of the underlying data. Butler and Kwon (2023) propose a novel end-to-end optimization technique for various popular portfolio optimization strategies [?]. The goal of this paper is to find an alternative to the common ‘Predict, then optimize’ workflow followed in this field, where the ‘Predict’ component is studied through factor models, and the ‘optimize’ component is studied through MVO, RP, etc. They integrate linear regression OLS optimization into MVO, to get the best regression model that maximizes the out of sample sharpe ratio. I would like to follow the steps outlined in this paper to guide the choice of δ .

4 Next Steps

It is indeed true that this work is far from over.

References

- [1] G. Costa and R. Kwon, “A robust framework for risk parity portfolios,” *Journal of Asset Management*, vol. 21, no. 5, pp. 447–466, 2020.
- [2] G. Costa and R. H. Kwon, “Data-driven distributionally robust risk parity portfolio optimization,” *Optimization Methods and Software*, vol. 37, no. 5, pp. 1876–1911, 2022.