ABSTRACT

We show that the probability of risk parity beating any other portfolio is more than 50 percent. We also prove that if portfolio performance is measured by Sharpe ratio, risk parity is the only maximin portfolio when (1) all assets’ future Sharpe ratios are greater than an unknown constant and all correlations are less than another constant, or (2) when the sum of all assets’ future Sharpe ratios is greater than some constant. If portfolio performance is measured by expected return, risk parity is the only minimax portfolio when the sum of assets’ Sharpe ratios is greater than a constant.
We formulate and prove several novel optimal properties of risk parity portfolios. We show under general conditions that the probability of risk parity beating any other portfolio is more than 50 percent.

If the performance of a portfolio is measured by its expected return, we prove with a game-theoretic approach that risk parity is the only minimax portfolio if the sum of assets’ future Sharpe ratios is greater than a constant. In general, we find a minimax portfolio if the assets’ Sharpe ratios satisfy some linear inequality.

If the performance of a portfolio is measured by its Sharpe ratio, we prove that risk parity is optimal in the maximin sense: under some natural assumptions, it will do better than any other portfolio construction method under the worst possible combination of true expected returns. We prove this under two scenarios: the first assumes that all assets’ future Sharpe ratios are greater than some positive unknown constant and all correlations are less than another unknown constant, while the second only assumes that the sum of all assets’ future Sharpe ratio is greater than some unknown constant. In each case, we provide explicit formulas and show that risk parity is the unique maximin portfolio.

Finally, we empirically examine historical performance for the two main asset classes and find conformance with our theoretical results. Risk parity tends to outperform both the tangency portfolio formed from too much knowledge and the equally weighted portfolio formed from too little. In short, risk parity represents a “sweet spot” of knowledge where ignoring the knowledge of past average returns but not past volatilities helps to focus more deeply on risk allocation and typically leads to better performance.

One of the most important problems a portfolio manager faces is finding the right weights for his portfolio’s assets. A major theoretical development for the solution to this problem was made by Arthur D. Roy [1952]. He answered the following question: if we know the first two moments of returns, namely their expected returns and their covariance matrix, what asset weights would maximize the mean-volatility ratio of the portfolio? We will call such portfolios tangency portfolios because the line drawn from the risk free rate will have the highest Sharpe ratio, and be tangent to, these portfolios.¹

Portfolio managers have long recognized a major problem with the tangency portfolio: the methodology required the knowledge of future first and second moments of asset returns, and it is extremely difficult to
estimate those, especially the first moments. Merton [1980] is the classic paper showing that estimating expected returns requires a longer time period while estimating variance requires finer observations of returns.

Even worse, with accumulated knowledge, it became clear that in some important cases, the weights proposed by the tangency approach were difficult to reconcile with the intuition and experience of portfolio managers. Even Markowitz himself didn’t follow this methodology when constructing his own portfolio. According to Zweig [2009], he simply invested 50/50 in stocks and bonds. Further, the tangency weights are fragile to the assumptions and can change wildly (Britten-Jones [1999]).

Risk parity (RP) is an alternative portfolio construction approach that allocates capital to each asset inversely proportional to its future expected volatility. While it appears to take no account of expected returns, it subtly does: it requires its assets to have a positive expected return; otherwise a short position with the same volatility would be preferred.

Risk parity has historically tended to outperform tangency and other standard portfolio allocation methods and several explanations for its success have been advanced. Chaves, Hsu, Li, and Shakernia [2011] among others compared risk parity with other more standard methods. Asl and Etula [2012] discuss risk parity and similar portfolio construction strategies from the perspective of robust optimization; building on Scherer [2007], Meucci [2007], and Ceria and Stubbs [2006], they consider the standard errors of the expected return estimations as the sole source of uncertainty, and show that in such cases, portfolios similar to but different from risk parity would be optimal. By contrast, we consider two more general cases that depend only on mild conditions on future asset Sharpe ratios to show that pure risk parity would be uniquely optimal.

Asness, Frazzini, and Pedersen [2012] show that leverage aversion can lead to excess returns to a risk parity portfolio, and they document RP’s historical and sustained outperformance. Here, we show that even if leverage aversion did not apply, risk parity would still beat any other portfolio on average, under the precise conditions we provide.

DeMiguel, Garlappi, and Uppal [2009] explore the equally weighted portfolio strategy that often beats tangency as well. However, we show the general conditions under which risk parity would beat any portfolio,
including the equally weighted one. This gives us an additional interesting intermediate result: while the tangency portfolio tries to use all available information and the equally weighted portfolio seems to use none of the available information, risk parity uses some but not all of the available information, and beats them both, as well as any other portfolio.

The question of why risk parity works can be thought of as a battleground in the larger war between seemingly ad-hoc heuristics-based approaches and traditional optimization approaches to finance in general and portfolio management specifically. By exploring this arena in detail, we aim to shed light on the larger question.

The term “heuristics” generally means “rule of thumb.” It is used in behavioral sciences in a predominately pejorative sense when compared to unattainable perfect rationality. However, in computational discussions, heuristics are simple but crucial algorithms that substantially improve performance. In the context of boundedly rational investor behavior, Gigerenzer [2012] argues that particular heuristics are “ecological” in the sense that they can be helpful in particular circumstances, and are neither universally good nor universally bad. Goldstein and Gigerenzer [2009] show that fast and frugal heuristics can make better predictions than more complex and more knowledge-intensive rules.

In this context, we argue that risk parity, as a fast and frugal heuristic, tends to outperform the more complex and more knowledge-intensive mean-variance approach. It also tends to outperform the overly simply and nearly entirely knowledge-independent equally weighting approach.

Of course, risk parity’s outperformance is not ubiquitous. Indeed, during 2012, because of the lackluster performance of bonds, tangency actually beat risk parity. That makes the main questions of this paper especially timely: are there conditions under which the risk parity approach is optimal in some sense? Can we estimate the probability that risk parity will outperform? This paper addresses these questions and more in a novel and general theoretical framework, with supporting empirical results.
RISK PARITY, EQUAL RISK CONTRIBUTION, EQUAL WEIGHT, AND TANGENCY PORTFOLIOS

Let \( X_T \) be a vector of random excess returns of \( n \) assets: \( X^T = (X_1, \ldots, X_n) \) such that \( \mathbb{E}(X) = \mu \) and \( \text{Var}(X) = \Sigma \), where \( \mu^T = (\mu_1, \ldots, \mu_n) \) and \( \Sigma = \{\sigma_{i,j}\}, i, j = 1, \ldots, n \).

We write \( X^T \), the transpose of \( X \), to emphasize that we normally define new vectors as column vectors.

Thus, \( X \) is a column vector and \( X^T \) is a row vector.

Let \( \tau \) be the assets’ Sharpe ratios: \( \tau^T = \{\tau_i = \frac{\mu_i}{\sigma_i}, i = 1, \ldots, n\} \)

We’ll use the fact that \( \Sigma = \Lambda_\sigma R \Lambda_\sigma \) and \( \Sigma^{-1} = \Lambda_\sigma^{-1} R^{-1} \Lambda_\sigma^{-1} \) where \( R \) is the correlation matrix and \( \Lambda_x \) is the diagonal matrix with vector \( x \) on its diagonal. So \( (\sigma^{-1})^T \Sigma \sigma^{-1} = 1^T R 1 \), where \( 1 \) is a column vector of ones.

The Sharpe ratio of portfolio with weights \( w^T = (w_1, \ldots, w_n) \) is

\[
SR = SR(w, \mu, \Sigma) = \frac{w^T \mu}{\sqrt{w^T \Sigma w}}
\]

We can rewrite this as

\[
SR = SR(w, \tau, R) = \frac{w^T \tau}{\sqrt{w^T R w}}
\]

It is well known that the maximum of the Sharpe ratio over all possible weights is

\[
\max_{w \in \mathbb{R}^n} SR(w, \mu, \Sigma) = SR(w^*, \mu, \Sigma) = \sqrt{\mu^T \Sigma^{-1} \mu}
\]

And the optimal weights \( w^* \) could be any weights that are proportional to \( w^* \sim \Sigma^{-1} \mu \)

With the normalizing condition \( 1^T w = 1 \), the optimal weights, the weights of the tangency portfolio, are:

\[
w^* = \frac{\Sigma^{-1} \mu}{1^T \cdot (\Sigma^{-1} \mu)}
\]

For the equal weight portfolio, of course, the weights are simply:

\[
w^* = \frac{1^T}{n}
\]
This is the same portfolio as the tangency portfolio in the case of uncorrelated assets with identical Sharpe ratios.

The risk parity (RP) weights $v: v^T = (v_1, ..., v_n)$ are by definition inversely proportional to the asset volatilities:

$$v \sim \sigma^{-1} \equiv \left( \frac{1}{\sigma_1}, ..., \frac{1}{\sigma_n} \right)^T, \sigma_i = \sqrt{\sigma_{i,i}}, i = 1, ..., n$$

Taking into account the normalizing constraint $\sum_i v_i = 1$, we have:

$$v_i = \frac{\sigma_i^{-1}}{\sum_i \sigma_i^{-1}}, i = 1, ..., n, v = \frac{\sigma^{-1}}{1^T \sigma^{-1}}$$

And its Sharpe ratio is:

$$SR = \frac{(\sigma^{-1})^T \mu}{\sqrt{(\sigma^{-1})^T \Sigma \sigma^{-1}}}$$

Let us define the equal risk contribution (ERC) portfolio. The volatility of a portfolio with weights $u$:

$$u^T = (u_1, ..., u_n)$$

is:

$$\sigma(u) = \sqrt{u^T \Sigma u} = \sqrt{\sum_i u_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} u_i u_j \sigma_{ij}}$$

Define the risk contribution of asset $i$ as:

$$\sigma_i(u) \equiv u_i \frac{\partial \sigma(u)}{\partial u_i} = u_i \frac{\sigma_i^2 + \sum_{j \neq i} u_j \sigma_{ij}}{\sigma(u)}$$

Therefore the risk (volatility) of the portfolio can be presented as the sum of its asset risks:

$$\sigma(u) = \sum_{i=1}^n \sigma_i(u)$$

The equal risk contribution portfolio is defined by requiring that all assets’ risks are equal:

$$\sigma_i(u) = \frac{\sigma(u)}{n}, i = 1, ..., n$$

Two additional constraints are usually enforced, namely, the normalizing constraint: $\sum u_i = 1$ and the no-short-selling constraint: $0 \leq u_i \leq 1, i = 1, ..., n$. 

Note that these definitions are not universally accepted. Sometimes equal risk contribution portfolios are called risk parity portfolio, and what we define as the risk parity portfolio are sometimes called naïve risk parity portfolios.

Actually, it would have been more exact and more specific to call an RP portfolio a volatility parity portfolio and an ERC portfolio a beta parity portfolio. Here is the logic why (see Maillard, Thierry and Teiletche [2010]). Denote the covariance between the \( i \)th asset and the portfolio by \( \sigma_{ip} = \text{cov}(X_i, \sum_j u_j X_j) = \sum_j u_j \sigma_{ij} \). Then \( \sigma_i(u) = u_i\sigma_{ip}/\sigma(u) \). By definition, the beta of asset \( i \) with the portfolio is \( \beta_i = \sigma_{ip}/\sigma^2(u) \). We know that for the ERC portfolio \( \sigma_i(u) = \sigma(u)/n \) for all \( i = 1, ..., n \). Therefore:

\[
u_i = \frac{\sigma_i(u)\sigma(u)}{\sigma_{ip}} = \frac{\beta_i^{-1}}{n} = \frac{\beta_i^{-1}}{\sum \beta_i^{-1}}\]

This is the same formula as for RP, only using betas instead of volatilities.

It is important to notice that in a very important general parameter case, the RP portfolio is the same as the ERC portfolio: namely, Maillard, Thierry and Teiletche [2010] proved that ERC becomes a RP portfolio when the correlations among all assets are the same. In particular, for \( n = 2 \), the ERC portfolio is the RP portfolio. Exact formulas for the weights of ERC portfolio are not known in the general case. Chaves, Hsu, Li, and Shakernia [2012] analyze algorithms for computing those weights.

**GAME THEORY FRAMEWORK**

Because game theory is not often used in portfolio theory, let’s review some basic concepts of the game theory to clarify our approach. Let’s define a 2-player zero sum game. Two players are playing a game. The goal of the game for each player is to maximize his payoff. Two abstract sets, A and B, are known to each player. A is called the set of strategies or actions or decisions of player 1 and B the set of strategies of player 2. The strategies are also called pure strategies to distinguish from mixed strategies which are randomized pure strategies. Of course any pure strategy is a mixed strategy concentrated in one decision. Both players also know the payoff function \( \phi(a,b), a \in A, b \in B \).
The game is played as follows. Player 1 chooses $a \in A$ and Player 2 chooses $b \in B$ simultaneously, each unaware of the choice of the other. Then their choices are made known and player 1 receives $\phi(a, b)$ and player 2 receives $-\phi(a, b)$. The total is zero, which is why it is called a zero-sum game. So $\phi(a, b)$ is a gain for player 1 and a loss for player 2. Player 1 wants to maximize the payoff and player 2 wants to minimize it.

In our case, player 1 is a portfolio manager who wants to find the portfolio weights of $n$ assets, such that his performance is the best under the worst possible action of the market. Player 2 is the market. $A$ is a set of portfolio weights available to the portfolio manager, $B$ is a set of parameters of distribution of assets excess returns from which the market “chooses” parameters to “hurt” the fund manager the most, to make the performance of the fund manager as bad as possible. The performance of the fund manager is measured either by expected return or by Sharpe ratio.

A game is called a matrix game if sets $A$ and $B$ are finite. Let’s look at an example of a two-player zero-sum matrix game. Assume $A = \{a_1, a_2\}, B = \{b_1, b_2\}$ and the payoff function for the first player is defined by the following table:

<table>
<thead>
<tr>
<th></th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$a_2$</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Let’s find $V_1$, the maximin gain of the game for player 1. If player 1 chooses $a_1$, then player 2 can harmfully choose $b_1$, and player 1 will receive $\phi(a_1, b_1) = 1$. If player 1 chooses $a_2$, then player 2 can harmfully choose $b_1$ again, and player 1 will receive $\phi(a_2, b_1) = 3$. Thus $V_1 = 3$. In general, $V_1$ is defined as:

$$V_1 = \max_{a \in A} \min_{b \in B} \phi(a, b)$$

In the same way, we can find $V_2$, the minimax loss of the game for player 2; in this case, $V_2 = 3$. In general, $V_2$ is defined as:

$$V_2 = \min_{b \in B} \max_{a \in A} \phi(a, b)$$

It could be shown that $V_1 \leq V_2$ for any $A, B$, and $\phi$, because for every fixed $a, b$

$$\min_{b \in B} \phi(a, b) \leq \phi(a, b) \leq \max_{a \in A} \phi(a, b)$$
Then taking the max on the left and the min on the right doesn’t change the inequality.

If $V_1 = V_2$, then $V = V_1$ is called the value of the game:

$$ V \equiv \max_{a \in A} \min_{b \in B} \phi(a, b) = \min_{b \in B} \max_{a \in A} \phi(a, b) $$

When the above equality holds, it is said that the game has a solution. Then we can find the value of the game and at least one optimal strategy for each player. Theorems establishing under what conditions games have values are called the Minimax Theorems.

Does the game always have a solution at least for 2x2 strategies?

It turns out that a matrix game always has a solution among pure strategies if the matrix has a saddle point, i.e. the matrix of payoffs has at least one element that is the minimum in its row and the maximum in its column. In our example, the matrix has a saddle point in row 2 and column 1, namely the value 3. But the following matrix doesn’t have a saddle point:

$$
\begin{array}{cc}
b_1 & b_2 \\
a_1 & 1 & 3 \\
a_2 & 4 & 2 \\
\end{array}
$$

It is easy to see that in this game $V_1 = 2$ and $V_2 = 3$, so there is no solution among pure strategies.

However, for a matrix game, a solution always exist among mixed strategies. This is the famous result of von Neumann [1928]. The solution, the mixed strategies of player 1 and player 2, is the Nash equilibrium following Nash [1951] who generalized von Neumann result for non-zero-sum games. At a Nash equilibrium, each player is making the best decision he can, taking into account the decision of the other.

In our second example, there exists a solution among mixed strategies for this game with the value $V = 2$, when player 1 chooses $a_1$ with probability $\frac{1}{2}$ and player 2 chooses $b_1$ with probability $\frac{1}{4}$.

Normally, the minimax property is attributed to the optimal strategy of player 1 and player 2 in games with a Nash equilibrium, when the game has a solution. When the game is analyzed from the point of view of player 1 only, the strategy that maximizes his minimum possible payoff is called maximin.
MINIMAX PROPERTY OF RISK PARITY AND OTHER PORTFOLIOS

Suppose that the variance-covariance matrix Σ of the n assets’ excess returns is known, but the vector of expected values μ (and therefore τ) is not known. We only know that the μ (or, equivalently, τ) belongs to a known set of vectors. We want to find the minimax portfolio in returns: the portfolio whose expected value is the greatest among the worst possible vectors μ.

We will find that any portfolio is a minimax portfolio among all no short sales portfolios for a set of assets when their expected returns are constrained by a linear inequality. We will see that in two natural special cases, the minimax portfolio is the equal weight or risk parity portfolio.

Let us start with finding the portfolio that has the best return under the assets’ worst distributional assumptions. Let Ω^n be the set of all possible normalized, no short sales portfolios:

Ω^n ≡ {w: Σ w_i = 1, 0 ≤ w_i ≤ 1, i = 1, ..., n}

and let the set of all possible assets’ expected values be constrained by a set M^ε, which is a set of non-negative vectors above a hyperplane:

M^n ε ≡ {μ: μ_1/ε_1 + ... + μ_n/ε_n ≥ 1, μ_i ≥ 0, i = 1, ..., n}

Then there exists a portfolio with weights w^* ∈ Ω^n and returns μ^* ∈ M^n ε such that

V ≡ max_{w ∈ Ω^n} min_{μ ∈ M^n ε} E_μ (w^T X) = min_{μ ∈ M^n ε} max_{w ∈ Ω^n} E_μ (w^T X) = E_{μ^*}((w^*)^T X) \tag{1}

and

V = \frac{1}{1^T (ε)^{-1}}

The portfolio

(w^*)^T = \frac{(ε)^{-1}}{1^T (ε)^{-1}} \tag{2}

is the only minimax portfolio, so that

V = min_{μ ∈ M^n ε} E_μ((w^*)^T X)

and the vector μ^*
\[(\mu^*)^T = \frac{1^T}{1^T(\varepsilon)^{-1}}\]  

is the only minimax vector of the assets returns, so that

\[V = \max_{\varepsilon \in \mathbb{E}} E_{\mu^*}(w^T X)\]

**Proof.** Because we want to find the portfolio performing the best under the worst conditions, and knowing that assets expected values are non-negative, we can redefine \(\mathbb{E}^n\) without loss of generality as

\[\mathbb{E}^n = \{\mu: \frac{\mu_1}{\varepsilon_1} + \cdots + \frac{\mu_n}{\varepsilon_n} = 1, \mu_i \geq 0, i = 1, \ldots, n\}\]

Consider now a zero-sum two players game in which player 1 is a portfolio manager whose set of strategies is \(A = \{a_1, \ldots, a_n\}\). Strategy \(a_i\) means investing the entire capital of $1 in asset \(i, i = 1, \ldots, n\). Player 2 is the market whose set of strategies is \(B = \{b_1, \ldots, b_n\}\). Strategy \(b_i\) means asset \(i\) has expected return \(\varepsilon_i\) and the rest of the assets have expected return 0. Obviously such a vector of assets expected values belong to \(\mathbb{E}^n\). Let us define the payoff of this game as:

\[\phi(a_i, b_j) = \begin{cases} \varepsilon_i, & i = j, i, j = 1, \ldots, n \\ 0, & i \neq j, i, j = 1, \ldots, n \end{cases}\]

As a matrix game, this game has a solution \(V\). Let

\[w = \{w_1, \ldots, w_n\}, w_1 + \cdots + w_n = 1, w_i \geq 0, i = 1, \ldots, n\]

and

\[\frac{\mu}{\varepsilon} = \left(\frac{\mu_1}{\varepsilon_1}, \ldots, \frac{\mu_n}{\varepsilon_n}\right), \frac{\mu_1}{\varepsilon_1} + \cdots + \frac{\mu_n}{\varepsilon_n} = 1, \mu_i \geq 0, i = 1, \ldots, n\]

be arbitrary mixed strategies and \(w^* = \{w_1^*, \ldots, w_n^*\}\) and \(\mu^*/\varepsilon = \{\frac{\tau_1}{\varepsilon_1}, \ldots, \frac{\tau_n}{\varepsilon_n}\}\) be the minimax mixed strategies for player 1 and player 2, respectively. Then

\[V = \sum_{i=1}^{n} w_i^*(\frac{\mu_i^*}{\varepsilon_i}) \varepsilon_i = \sum_{i=1}^{n} w_i^* \mu_i^*\]

That proves (1).
The payoff of the minimax mixed strategy of player 1 is at least $V$, regardless of the strategy player 2 chooses, so for any pure strategy $b_i$

$$w_i^* \epsilon_i \geq V, \quad i = 1, \ldots, n \tag{4}$$

Therefore

$$V \leq \sum_{i=1}^{n} w_i^* \epsilon_i (\mu_i^* / \epsilon_i) = \sum_{i=1}^{n} w_i^* \mu_i^* = V$$

which means that (4) are equalities:

$$w_i^* = V / \epsilon_i, \quad i = 1, \ldots, n$$

And the value of the game is

$$V = \frac{1}{\frac{1}{\epsilon_1} + \cdots + \frac{1}{\epsilon_n}}$$

Similarly, analyzing the game from player 2’s point of view, we can prove that

$$\frac{\mu_i^*}{\epsilon_i} = \frac{V}{\epsilon_i}, \quad i = 1, \ldots, n$$

or

$$\mu_i^* = V \frac{1}{\frac{1}{\epsilon_1} + \cdots + \frac{1}{\epsilon_n}}, \quad i = 1, \ldots, n$$

establishing (3). That finishes the proof.

We have also proved more generally that any portfolio is a minimax portfolio for a set of constrained expected values $\mathbb{M}_{\epsilon}^n$, if, as shown by (2), the vector $\epsilon$ is chosen inversely proportional to the weights. This means that by analyzing the portfolio of any portfolio manager, we can make a statement about his view of future expected returns.

Let’s consider two important cases.

**Minimax property of expected value of equal weight portfolio**

If the sum of assets’ non-negative expected returns is greater than a certain (unknown) value, then the equal weight portfolio is the only minimax portfolio among all no-short-sales portfolios.
In other words, if the portfolio manager knows that the sum of all non-negative expected returns is greater than a certain (unknown) constant, then, regardless of the constant, the minimax portfolio is the equal weight portfolio: this portfolio that will have the greatest expected value under the worst possible scenario.

The proof follows from the minimax property of a general portfolio if we take all $\epsilon_i$ equal to each other.

**Minimax property of expected value of risk parity portfolio**

If the sum of assets non-negative expected Sharpe ratios is greater than a certain (unknown) constant, then the risk parity portfolio is the only minimax portfolio among all no-short-sales portfolios.

In other words, if the portfolio manager knows that the sum of all non-negative assets Sharpe ratios is greater than a certain (unknown) constant, then, regardless of the constant, the minimax portfolio is the risk parity portfolio: this portfolio that will have the greatest expected value under the worst possible scenario.

The proof follows from the minimax property of a general portfolio if we take all $\epsilon_i$ to be proportional to asset volatilities.

**MAXIMIN PROPERTIES OF RISK PARITY**

In this section we will establish two maximin properties of risk parity.

In both cases we fix a certain set of parameters and show that the minimum Sharpe ratio of the RP portfolio on this set is greater than the minimum Sharpe ratio on the same set of any other portfolio.

We look at portfolio manager activity as a two-stage game. In stage one, the portfolio manager chooses the weights of his portfolio from some fixed set of weights. In stage two, the market chooses the parameters of the assets’ excess return distribution from some other set. We can assume the worst possible scenario for the portfolio manager, namely that the market always chooses the distribution that makes the performance of the portfolio manager worst. If the portfolio manager’s performance is measured by his portfolio’s Sharpe ratio, how should he choose his portfolio?

We can’t directly use the matrix games as we did for the minimax results because we measure the performance of a strategy by its Sharpe ratio, not by its expected value. So the standard game theoretical approach of mixed strategies cannot be directly applied.
Each asset’s Sharpe ratio is positive and all correlations are less than one

Let’s assume again that the portfolio manager knows the asset volatilities but does not know either the asset expected returns or the correlations between asset returns.

Yet he knows something. He chose assets with enough care so that he is reasonably certain that the worst Sharpe ratio of any asset is still positive. In other words, all he knows about the chosen assets is that the expected return of each should be positive, but he doesn’t necessarily know which would perform better than the others. Further, he also believes that different assets are indeed different, with correlations less than one.

We want to prove the following statement: the risk parity portfolio with weights \( w^* = \frac{\sigma^{-1}}{1^T \sigma^{-1}} \) is the only maximin portfolio with respect to the Sharpe ratio \( SR \),

\[
SR = SR(w, \mu, \sigma, R) = \frac{w^T \mu}{\sqrt{(w \sigma)^T R (w \sigma)}},
\]

among all no short sales portfolio \( w \in \Omega^n \) such that

\[
\min_{\mu \in M1, R \in R1} F(w^*, \mu, \sigma, R) = \max_{w \in \Omega^n} \min_{\mu \in M1, R \in R1} F(w, \mu, \sigma, R)
\]

where

\[
\Omega^n \triangleq \{w: \sum w_i = 1, 0 \leq w_i \leq 1, i = 1, ..., n\}
\]

\[
M1 \triangleq \{\mu: \frac{\mu_i}{\sigma_i} \geq \epsilon > 0, i = 1, ..., n\}
\]

\[
R1 \triangleq \{\rho_{i,j} \neq i = \frac{\sigma_{i,j}}{\sigma_i \sigma_j} \leq \delta, 0 < \delta < 1, \rho_{i,i} = 1, i = 1, ..., n\}
\]

Proof

Introducing new variables \( a = (a_i = w_i \sigma, i = 1, ..., n) \), we can rewrite the Sharpe ratio as:

\[
SR = \frac{a^T \mu}{\sqrt{a^T R a}}
\]

Because all \( a_i \)'s are non-negative, the Sharpe ratio achieves its smallest possible value when the numerator is as small as possible and the denominator is as large as possible:
where D is a correlation matrix with all non-main-diagonal correlations equal to \( \delta \).

To finish our proof we need the following statement: if the Sharpe ratios of all assets are equal and their correlations are all equal, then the risk parity portfolio is the tangency portfolio. Maillard, Thierry and Teiletche [2010] proved this statement. A different proof was offered by Kaya and Lee [2012]. We’ll give here yet another, simpler proof.

We know that the weights of the tangency portfolio are proportional to \( D^{-1}1 \). In order to prove that this portfolio is the risk parity portfolio we need to show is that \( D^{-1}1 \) is a product of a constant times \( 1 \). If correlations are equal, row sums of \( D \) are equal,

\[
D1 = k1
\]

for some constant \( k \). Thus \( 1 = kD^{-1}1 \), which proves the result.

Actually, because ours is a no short-sales portfolio, we needed a slightly different statement: if the Sharpe ratios of all assets are equal and positive and their correlations are all equal and greater than zero, then the risk parity portfolio is the portfolio with the highest Sharpe ratio among all no-short-sales portfolios and it is equal to the tangency portfolio. The proof is similar to the previous statement; we simply add that, because correlations are positive, the constant \( k \) is positive, which means that the tangency portfolio has all weights positive, which confirms that it is a no-short-sales portfolio.

Analysis of the proof shows that the risk parity is the only maximin portfolio.

The sum of all assets’ Sharpe ratios is positive

Let us prove that the risk parity is a maximin portfolio in Sharpe ratio when the sum of the Sharpe ratios of all of the assets is greater than some positive constant.

\[
M2 = \{ \mu: 1^T \frac{\mu}{\sigma} = \sum_{i=1}^{n} \frac{\mu_i}{\sigma_i} \geq \epsilon > 0 \}
\]
This parameter set describes a situation when the portfolio manager is reasonably certain that in the worst case the total sum of all assets’ Sharpe ratios cannot be less than some positive constant. In this case, any particular asset may even have a negative Sharpe ratio, so long as the simple total (or, equivalently, average) across all assets is still positive.

We want to prove the following statement: the risk parity portfolio with weights \( w^* = \frac{\sigma^{-1}}{1^{T}\sigma^{-1}} \) is the only maximin with respect to Sharpe ratio SR,

\[
SR = SR(w, \mu, \Sigma) = \frac{w^{T}\mu}{\sqrt{w^{T}\Sigma w}}
\]

among all no short sales portfolio \( w \in \Omega^n \) such that

\[
\min_{\mu \in M^2} SR(w^*, \mu, \Sigma) = \max_{w \in \Omega^n} \min_{\mu \in M^2} SR(w, \mu, \Sigma)
\]

where

\[
\Omega^n \overset{def}{=} \{ w: \sum w_i = 1, 0 \leq w_i \leq 1, i = 1, ..., n \}
\]

\[
M^2 = \{ \mu: 1^{T}\mu = \sum_{i=1}^{n} \frac{\mu_i}{\sigma_i} \geq \epsilon > 0 \}
\]

**Proof**

In the worst case we have:

\[
\min_{\mu \in M^2} SR = \epsilon \min_{i=1,..,n} (w_i\sigma_i) \frac{1}{\sqrt{w^{T}\Sigma w}}
\]

We need to find maximum in \( w \) of the following function:

\[
f(w) = \epsilon \min_{i=1,..,n} (w_i\sigma_i) \frac{1}{\sqrt{w^{T}\Sigma w}}
\]

The optimal weights \( w^* \) for which this function achieves its maximum is the same vector on which the following function achieves its minimum:

\[
\frac{1}{f(w)^2} = \frac{1}{\epsilon^2} \frac{w^{T}\Sigma w}{\left( \min_{i=1,..,n} (w_i\sigma_i) \right)^2} \geq \frac{1}{\epsilon^2} \sum_{i,j=1}^{n} \frac{\sigma_{i,j}}{\sigma_i \sigma_j}
\]
The last inequality holds because
\[
\left( \min_{i=1,\ldots,n} w_i \sigma_i \right)^2 \leq (w_j \sigma_j)(w_k \sigma_k), \quad j, k = 1, \ldots, n
\]

And therefore:
\[
f(w) \leq \epsilon \frac{1}{\sqrt{\sum_{i=1}^{n} \frac{\sigma_{i,j}}{\sigma_i \sigma_j}}}
\]

For the RP portfolio $P_{w^*}$ with weights $w^* \{w_i^* = c/\sigma_i, i = 1, \ldots, n\}$ where $c = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i}}$, we have:
\[
\min_{\mu \in \mathbb{R}^2} \frac{E_{\mu}(P_{w^*})}{\sqrt{\text{Var}(P_{w^*})}} = \min_{\mu \in \mathbb{R}^2} \frac{\sum_{i=1}^{n} \frac{\mu_i}{\sigma_i}}{\sqrt{(w^*)^T \Sigma (w^*)}} = f(w^*) = \epsilon \frac{1}{\sqrt{\sum_{i,j=1}^{n} \frac{\sigma_{i,j}}{\sigma_i \sigma_j}}}
\]

which shows that $w^*$ is in fact the value for which function $f(w)$ achieves its maximum. Analysis of the proof shows that the risk parity is the only maximin portfolio.

**WHEN RISK PARITY BEATS TANGENCY BY SHARPE RATIO**

Say weights $w$ outperform weights $v$ for a given $m$ and $S$ if they result in a higher Sharpe ratio:
\[
SR(w; m, S) = \frac{w^T m}{\sqrt{w^T S w}} \geq SR(v; m, S) = \frac{v^T m}{\sqrt{v^T S v}}
\]

where $m$ are the assets’ future expected returns, $S$ is the assets’ future variance matrix, and $w$ and $v$ are portfolios weights based on the past expected returns $\mu$ and the past variance matrix $\Sigma$.

Taking $w$ as the weights for the risk parity portfolio, and $v$ as the weights for the tangency portfolio, we see that risk parity outperforms tangency if and only if:
\[
m^T \left( \frac{\sigma^{-1}}{\sqrt{(\sigma^{-1})^T S \sigma^{-1}}} - \frac{\Sigma^{-1} \mu}{\sqrt{\mu^T \Sigma^{-1} S \Sigma^{-1} \mu}} \right) > 0
\]

This defines an $n$-dimensional hyperplane for the vectors $m$. This hyperplane passes through the origin and is perpendicular to the vector:
\[
\left( \frac{\sigma^{-1}}{\sqrt{(\sigma^{-1})^T S \sigma^{-1}}} - \frac{\Sigma^{-1} \mu}{\sqrt{\mu^T \Sigma^{-1} S \Sigma^{-1} \mu}} \right)
\]
The future returns do not depend on the future variance matrix and therefore risk parity beats tangency in expected returns if and only if:

\[ m^T \left( \frac{\sigma^{-1}}{1^T \sigma^{-1}} - \frac{\Sigma^{-1} \mu}{1^T \Sigma^{-1} \mu} \right) > 0 \]

**Case when the future variance matrix is equal to the past variance matrix**

If the future variance matrix is equal to the past, then risk parity beats tangency if and only if:

\[ m^T \left( \frac{\sigma^{-1}}{\sqrt{(\sigma^{-1})^T \Sigma \sigma^{-1}}} - \frac{\Sigma^{-1} \mu}{\sqrt{\mu^T \Sigma^{-1} \mu}} \right) > 0 \]

Let us simplify the general expression for the difference in Sharpe ratios between RP and any arbitrary portfolio, if the future variance matrix is equal to the past. We will use the fact that \( \Sigma = \Lambda_x R \Lambda_x \) and \( \Sigma^{-1} = \Lambda_x^{-1} R^{-1} \Lambda_x^{-1} \) where \( R \) is the correlation matrix and \( \Lambda_x \) is the diagonal matrix with vector \( x \) on its diagonal.

Then \( \frac{1}{x} \Lambda_x = 1^T, \Lambda_x \frac{1}{x} = 1 \). So:

\[ (\sigma^{-1})^T \Sigma \sigma^{-1} = 1^T R 1 \]

Let us use the Sharpe ratios instead of expected returns of assets:

\[ \tau \equiv \left\{ \tau_i = \frac{\mu_i}{\sigma_i}, i = 1, ..., n \right\}, t \equiv \left\{ t_i = \frac{m_i}{\sigma_i}, i = 1, ..., n \right\} \]

We already established that RP outperforms any portfolio with weights \( w \) by Sharpe if:

\[ m^T \left( \frac{w}{\sqrt{1^T R 1}} \right) > 0 \]

If \( v \equiv w \sigma = \{ v_i = w_i \sigma_i, i = 1, ..., n \} \) then the last inequality can be rewritten as:

\[ t \sigma^T \left( \frac{v/\sigma}{\sqrt{1^T R 1}} \right) > 0 \]

or:

\[ t \left( \frac{1}{\sqrt{1^T R 1}} - \frac{v}{\sqrt{v^T R v}} \right) > 0 \quad (5) \]

If \( w \) are the weights of the tangency portfolio, then \( ((t \sigma)_i = \tau_i \sigma_i, i = 1, ..., n) \)
\[
\frac{w}{\sqrt{w^T \Sigma w}} = \frac{\Sigma^{-1} \mu}{\sqrt{\mu^T \Sigma^{-1} \mu}} = \frac{\Lambda_{\sigma^{-1}} R^{-1} \Lambda_{\sigma^{-1}} (\tau \sigma)}{\sqrt{\tau \sigma^T R^{-1} \tau}}
\]

Therefore RP beats tangency in Sharpe ratio if and only if:

\[
t \left( \frac{1}{\sqrt{1^T R 1}} - \frac{R^{-1} \tau}{\sqrt{\tau R^{-1} \tau}} \right) > 0
\]

**PROBABILITY THAT RISK PARITY BEATS ANY OTHER PORTFOLIO IS GREATER THAN 50%**

Assume that all future asset variances are the same as the past and all future asset correlations are equal to a non-negative number. Assume that the directions of the assets’ future Sharpe ratios \( t \) are drawn completely randomly from the positive hyperquadrant

\[
R_n^+ = \{ t_1 \geq 0, \ldots, t_n \geq 0 \}
\]

Then we can show that the probability that risk parity beats any other portfolio with positive coefficients by Sharpe ratio is greater than 50%.

To begin, we rewrite the inequality (5) as:

\[
t (e - c_R d) > 0,
\]

where \( e = \frac{1}{\sqrt{n}} d = \frac{v}{\|v\|} \) and \( c_R = \frac{\sqrt{1^T R 1 \|v\|}}{\sqrt{n \|v\| R v}} \).

The vector \( e \) is the rotation axis of \( R_n^+ \). Therefore to prove our statement it is sufficient to prove that either: A) \( d \) and \( e \) lie on different sides of the hyperplane defined by Equation (7), or B) \( d \) and \( e \) lie on the same side of the hyperplane but the distance of \( d \) (which is a unit vector in the direction of the portfolio with weights \( v \)) from the hyperplane is longer than the distance of \( e \) from the same hyperplane.

Assume for all \( v \) and \( R \) that \( c_R \geq 1 \) (we will prove this statement at the end.) Then:

\[
d (e - c_R d) = de - c_R < 0
\]

because \( d \) and \( e \) are unit vectors.

Now, let us analyze the two cases.

A) Because of Equation (8), for \( d \) and \( e \) to lie on different sides of the hyperplane we must have:
\[ e (e - c_R d) > 0 \]

which is equivalent to:

\[ de < 1/c_R \] \hspace{1cm} (9)

B) We can now assume that (9) doesn’t hold:

\[ de \geq 1/c_R \] \hspace{1cm} (10)

The distance from a unit vector \( u \) to a plane passing through the origin perpendicular to a vector \( h \) is

\[ |uh|/|h| \]. For our hyperplane defined by Equation (7), \( h = e - c_R d \). Therefore the distance from \( d \) to the hyperplane is:

\[ |de - c_R| = c_R - de \]

because \( c_R > 1 \geq de \). The distance from \( e \) to the hyperplane is

\[ |1 - c_R de| = c_R de - 1 \]

where the last equation follows because of (10).

\( d \) is further from the plane than \( e \) if and only if:

\[ c_R - de \geq c_R de - 1 \]

which is obvious because \( d \) and \( e \) are unit vectors.

The only thing remaining to be proved is that \( c_R \geq 1 \) or:

\[ \frac{v^T R v}{v^T v} \leq \frac{1^T R 1}{n} \]

The right hand side of this inequality is equal to \( 1 + (n - 1)\rho \), where \( \rho \geq 0 \) is the correlation between any two assets, the common term in matrix \( R \). The left hand side of this inequality is the so-called Raleigh quotient and is never greater than \( \lambda_1 \), the maximum eigenvector of matrix \( R \). According to Morrison [1967, 244-245]: \( \lambda_1 = 1 + (n - 1)\rho \). That completes the proof.

**Illustration for \( n = 2 \) Uncorrelated Assets**

In this case, according to Inequality (6):

\[ t \left( \frac{1}{\sqrt{2}} - \frac{\tau}{\sqrt{\tau^T \tau}} \right) > 0 \]
We can depict the result geometrically, as shown in Exhibit 1.

Here $e = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, $d = \frac{\tau}{\|\tau\|}$ is a unit vector, $\tau$ is an arbitrary vector of the assets’ past Sharpe ratios from the positive quadrant, $a \equiv e - d$ by definition, and $2\theta$ is the angle between $e$ and $d$ so that $\cos(2\theta) = ed$.

We assumed that the assets’ future Sharpe ratios $t = (t_1, t_2)$ are randomly chosen from the positive quadrant of a unit circle. Then the probability that risk parity beats tangency for two assets is easily seen geometrically to be:

$$P_2 = \frac{\pi + \theta}{\frac{\pi}{2}}$$

Exhibit 1

**WHEN RISK PARITY BEATS TANGENCY EMPIRICALLY**

Consider an investor allocating between the two main asset classes: equities and bonds. The investor observes the monthly returns of both time series and compares three possible portfolios: the risk parity portfolio that invests inversely proportional to each asset’s realized volatility, the tangency portfolio that invests in the portfolio that would have had the highest *ex ante* realized Sharpe ratio, and the fixed portfolio that invests 60 percent in stocks and 40 percent in bonds. The fixed portfolio may be viewed as an approximation to the equally weighted portfolio as well.
How would the investor have performed historically under each of those three possibilities? We take the monthly total returns of the S&P 500 index from Bloomberg and the monthly total returns of the Barclays Capital US Aggregate Bond Index from Dimensional Fund Advisors (DFA) Returns 2.0 software, from February 1988 through October 2012.

Exhibit 2 shows the 24-month rolling Sharpe ratios of these three portfolios, formed using the returns from the previous 24 month period, and held for the subsequent 24 month period. Risk parity outperformed both other portfolios, averaging a 0.99 Sharpe ratio. The tangency portfolio was the worst, averaging a 0.48 Sharpe ratio. The fixed 60/40 portfolio averaged a 0.68 Sharpe ratio.

The weights for the tangency portfolio fluctuate wildly. Exhibit 3 shows a paired histogram comparing the distributions of the risk parity and tangency portfolio’s equity weighting (the fixed 60/40 portfolio was always a constant 0.60). The risk parity equity weighting was always between 12.7 percent and 37.9 percent while the tangency portfolio ranged from -8,957 percent to 2,644 percent; the exhibit shows the clipped distribution with all weights below -1 or above +1 reflected in those final bars.
To test the implications from our theoretical framework, we can examine the sensitivity of the performance of the risk parity and tangency portfolios to the performance of the underlying assets. Exhibit 4 plots the Sharpe ratio of each of the two portfolios separately, as well as the excess Sharpe ratio of the risk parity portfolio over the tangency portfolio, relative to the Sharpe ratios of the stocks and bonds separately, as well as to their sum. The best-fit regression line is overlayed. All Sharpe ratios are computed for the same time periods, on a rolling 10-month basis.

Consider the first column in Exhibit 4, showing the relation between the portfolio Sharpe ratio and the stock Sharpe ratio during the same period. Counter to the usual intuition that tangency outperforms risk parity when equities outperform, we see that empirically risk parity performs better when stocks perform better, while the performance of the tangency portfolio is essentially unrelated to the simultaneous performance of stocks.

Similarly, risk parity also has a higher sensitivity to bond performance than does tangency.

Finally, as shown earlier, the risk parity Sharpe ratio corresponds well with the sum of the asset Sharpe ratios, as can be seen in the top right graph of Exhibit 4.
Another implication of the theoretical framework above is that risk parity would be closer than \textit{ex ante} tangency to \textit{ex post} tangency more than half of the time. Exhibit 5 calculates the vector angle between the \textit{ex}
post tangency portfolio weights and the risk parity and ex ante tangency portfolio, respectively, for 24 month periods. The angle with risk parity is usually lower in the time series graph. The table accompanying Exhibit 5 shows that for periods varying from 12 months to 60 months, the risk parity angle is indeed always more likely to be lower than ex ante tangency. The average probability is about 70 percent, and the average angle discrepancy is about 10 degrees.

**CONCLUSION**

Forming risk parity portfolios does not require as much data and as many sophisticated tools as forming other portfolios, such as the tangency portfolio embraced by standard portfolio theory. But it does require more data than the equally weighted portfolio. Yet it consistently outperforms both, and lately has become a prominent instrument among fund managers and a central topic among academic researchers. Risk parity may represent a “sweet spot” of heuristics where any more or any less knowledge would seem to harm performance.

We have described the exact parametric conditions when risk parity outperforms other portfolios, including tangency. This research provides mathematical validation for portfolio managers choosing risk parity under uncertainty by formulating the exact conditions of those uncertainties and proving precise mathematical results about the superiority of risk parity portfolio under those conditions.

**REFERENCES**


\[1\] We are aware that normally the optimality result is attributed to Markowitz or Sharpe. However, the founding papers of modern portfolio theory, Markowitz [1952] and Sharpe [1966] don’t have this result while Roy [1952] does. See some discussion of Roy’s forgotten contribution in Sullivan [2011]. Markowitz [1952] appears to be the first to suggest evaluating portfolios by the relationship between their expected returns and their variances and to develop the concept of efficient portfolios.