The Curse of Knowledge: When and Why Risk Parity Beats Tangency

Gregg S. Fisher, Philip Z. Maymin, Zakhar G. Maymin*

*contact author: Zak Maymin, Gerstein Fisher Research Center, zmaymin@gersteinfisher.com

Abstract

We formulate and prove several novel optimal properties of risk parity portfolios. We present and analyze closed form solutions for the exact condition between the realized and the assumed future parameters under which the risk parity portfolio, formed by ignoring historical average returns, beats the tangency portfolio formed in the standard mean-variance framework. We show under general conditions that the probability of risk parity beating tangency is more than 50 percent. We also 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 the maximin properties of risk parity portfolio under two scenarios. One such scenario assumes that all assets’ future Sharpe ratios are greater than some positive unknown constant and all correlations are less than another unknown constant. Another scenario 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. In short, ignoring the knowledge of past average returns to focus more deeply on risk allocation typically leads to better performance.
1. Introduction

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 realized that there was one 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 the 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 tended to outperform tangency historically 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 tangency on average, under the precise conditions we provide.

Risk parity versus tangency 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. One candidate for the cognitive bias preventing all investors from pursuing risk parity might be due to what Camerer, Loewenstein, and Weber (1989) refer to as the “curse of knowledge,” a term they attribute to Robin Hogarth, where they demonstrate economic and market transactions in which agents with more knowledge are unable to ignore that knowledge, even when doing so would be to their advantage.

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: what is the condition on future parameters that makes one approach preferable to the other? Are there any conditions under which risk parity approach is optimal in some sense? Can we estimate the probability that risk parity will outperform? This paper addresses all these questions and more in a novel and general theoretical framework, with supporting empirical results.
2. Tangency Portfolio

The tangency portfolio has the highest Sharpe ratio for $n$ assets having random excess returns:

\[ X^T = (X_1, \ldots, X_n) \]

such that

\[ E(X) = \mu \text{ and } \text{Var}(X) = \Sigma \]

where

\[ \mu^T = (\mu_1, \ldots, \mu_n) \quad \text{and} \quad \Sigma = \{\sigma_{i,j}\}, i, j = 1, \ldots, n. \]

We write $X^T$, the transpose of $X$, to emphasize that we normally define new vectors as column vectors. So $X$ is a column vector and $X^T$ is a row vector.

The tangency portfolio with weights $w^T = (w_1, \ldots, w_n)$ maximizes the Sharpe ratio:

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

Instead of the usual solution of this optimization problem using a Lagrange multiplier, we will solve it using a linear algebraic and geometric framework that will be useful for us in later sections.

Because $\Sigma$ is a positive-definite symmetric matrix, by the spectral decomposition theorem it can be written as:

\[ \Sigma = Q \Gamma Q^T \]
where \( Q \) is an orthonormal matrix, \( Q^T = Q^{-1} \). The columns of \( Q \) are eigenvectors of length 1 of \( \Sigma \) and \( \Gamma \) is the diagonal matrix of the corresponding positive eigenvalues of \( \Sigma \).

Let us introduce new variables \( a^T = w^T Q = (a_1, ..., a_n) \) as the coordinates of our vector of portfolio weights \( w \) in the new orthonormal basis given by the eigenvectors of \( \Sigma \).

By the definition of \( a, w = Qa \) and \( w^T = a^T Q^T \). Therefore:

\[
SR = \frac{a^T Q^T \mu}{\sqrt{a^T Q^T \Sigma Qa}} = \frac{a^T Q^T \mu}{\sqrt{a^T \Gamma a}}
\]

Define \( b^T \equiv a^T \Gamma^{-1/2} \). The diagonal matrix \( \Gamma^{-1/2} \) is well-defined because \( \Sigma \) is positive definite. So \( a^T = b^T \Gamma^{-1/2} \). It is easy to see that \( \Gamma^{-1/2} Q^T = \Sigma^{-1/2} \), and therefore \( w^T = b^T \Gamma^{-1/2} Q^T = b^T \Sigma^{-1/2} \), and:

\[
SR = \frac{b^T \Gamma^{-1/2} Q^T \mu}{\sqrt{b^T b}} = \frac{b^T \Sigma^{-1/2} \mu}{\sqrt{b^T b}}
\]

The optimum weights stay optimal if we multiply them by any positive constant. So we can assume for the moment that \( b \) lies on a unit sphere, \( b^T b = 1 \). Then to achieve the maximum in the dot product of two vectors in the numerator of the last expression for \( SR \), \( b \) must be parallel to \( \Sigma^{-1/2} \mu \), \( b^T = \mu^T \Sigma^{-1/2} \), where \( C > 0 \) is a constant found from the condition that the length of \( b \) is 1, namely \( C^2 (\mu^T \Sigma^{-1/2} \Sigma^{-1/2} \mu) = 1 \). So:

\[
C = \frac{1}{\sqrt{\mu^T \Sigma^{-1} \mu}}
\]

Then the optimal Sharpe ratio \( SR^* \) is:
And the optimal weights $w$ are proportional to:

$$w \sim \Sigma^{-\frac{1}{2}}b \sim \Sigma^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}\mu = \Sigma^{-1}\mu$$

If we impose the normalizing condition $\mathbf{1}^T w = 1$, where $\mathbf{1}$ is a column vector of ones, then the optimal weights are:

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

### 3. Risk Parity and Equal Risk Contribution

The Risk Parity (RP) weights $\nu, \nu^T = (\nu_1, ..., \nu_n)$ are by definition inversely proportional to the asset volatilities:

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

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

$$\nu_i = \frac{\sigma_i^{-1}}{\sum \sigma_i^{-1}}, i = 1, ..., n, \nu = \frac{\sigma^{-1}}{\mathbf{1}^T \sigma^{-1}}$$

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:

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

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 a RP portfolio a Volatility Parity portfolio and 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 $\sigma_{ip} = \text{cov}(X_i, \sum_j u_j X_j) = \sum_j u_i \sigma_{ij}$. 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:

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

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.

4. Maximin Optimality of Risk Parity

We 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.
4.1. Each asset’s Sharpe ratio is positive and all correlations are less than one.

Let us define a set of parameters when each asset’s Sharpe ratio is greater than some positive constant and all correlations between different assets are less than some constant:

\[ M1 \equiv \frac{\mu_i}{\sigma_i} = \epsilon > 0, i = 1, ..., n \]

\[ R \equiv \left( \rho_{i,j \neq i} = \frac{\sigma_{i,j}}{\sigma_i \sigma_j} \leq \delta < 1, \rho_{i,i} = 1, i = 1, ..., n \right) \]

This parameter set describes a situation when 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 that he is reasonably certain that the worst Sharpe ratio of any asset is still positive. In other words, all he knows about chosen assets is that the expected return of each should be positive, but he doesn’t necessarily know which of them would perform better than others. Further, he also believes that different assets are indeed different, with correlations less than one.

Let’s consider the normalized portfolio with no short sales:

\[ \sum w_i = 1, 0 \leq w_i \leq 1, i = 1, ..., n \]

Then the Sharpe ratio of this portfolio is:

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

Introducing new variables \( a_i = w_i \sigma_i, i = 1, ..., n \), we can rewrite this as:
\[ SR = \frac{a^T \frac{\mu}{\sigma}}{\sqrt{a^T Ra}} \]

Obviously the Sharpe ratio achieves its smallest possible value when the numerator is as small as possible and the denominator is as large as possible:

\[ \min_{M_1} SR = \epsilon \frac{a^T 1}{\sqrt{a^T \Delta a}} \]

Where \( \Delta \) is a correlation matrix with all 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 \( \Delta^{-1} 1 \). All we need to show is that this is a product of a constant times \( 1 \). If correlations are equal, row sums of \( \Delta \) are equal,

\[ \Delta 1 = k 1 \]

for some constant \( k \). Thus \( 1 = k \Delta^{-1} 1 \), which proves the result. Analysis of the proof shows that the risk parity is the only maximin portfolio.

4.2. The sum of all assets’ Sharpe ratios is positive

Let us define a set of parameters when the sum of the Sharpe ratios of all of the assets is greater than some positive constant.
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.

Let us again consider a normalized portfolio with no short sales:

$$\sum w_i = 1, 0 \leq w_i \leq 1, i = 1, ..., n$$

This portfolio’s Sharpe ratio is:

$$SR = \frac{w^T \mu}{\sqrt{w^T \Sigma w}}$$

In the worst case we have:

$$\min_{M2} SR = \epsilon \frac{\min_{i=1,...,n} (w_i \sigma_i)}{\sqrt{w^T \Sigma w}}$$

We need to find maximum in $w$ of the following function:

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

The $w^*$ for which this function achieves its maximum is the same vector on which the following function achieves its minimum:
\[ \frac{1}{f(w)^2} = \frac{1}{\epsilon^2} \left( \min_{i=1, \ldots, n} (w_i \sigma_i) \right)^2 \geq \frac{1}{\epsilon^2} \sum_{i,j=1}^{n} \sigma_{i,j} \]

because \( \left( \min_{i=1, \ldots, n} w_i \sigma_i \right)^2 \leq (w_j \sigma_j)(w_k \sigma_k) \), \( j, k = 1, \ldots, n \). And therefore:

\[ f(w) = \epsilon \frac{1}{\sqrt{\sum_{i,j=1}^{n} \sigma_{i,j}}} \]

For the risk parity portfolio weights, \( w_i = 1/\sigma_i, i = 1, \ldots, n \), we have:

\[ \min_{m \in M} \frac{E_m(P_w)}{\sqrt{\text{Var}(P_w)}} = \min_{m \in M} \frac{\sum_{i=1}^{n} \mu_i}{\sqrt{w^T \Sigma w}} = \epsilon \frac{1}{\sqrt{\sum_{i,j=1}^{n} \sigma_{i,j}}} \]

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

5. 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:

\[ SR(w; m, S) = \frac{w^T m}{\sqrt{w^T \Sigma w}} \geq SR(v; m, S) = \frac{v^T m}{\sqrt{v^T \Sigma v}} \]

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

This defines an \( n \)-dimensional hyperplane for the vectors \( m \). This hyperplane passes through the origin and is perpendicular to the vector:

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

The future returns do not depend on the future variance matrix and therefore risk parity beats tangency in expected returns if and only if:

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

**5.1. 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:

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

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

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

Let us use the Sharpe ratios instead of expected returns of assets:
We already established that RP outperforms any portfolio with weights \( w \) by Sharpe if:

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

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

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

or:

\[
t \left( \frac{1}{\sqrt{1^T \mathbf{R} 1}} - \frac{\nu}{\sqrt{\nu^T \mathbf{R} \nu}} \right) > 0 \tag{5.1.1}
\]

The Sharpe ratio of the tangency portfolio is \( ((\tau \sigma)_i = \tau_i \sigma_i = \mu_i, i = 1, \ldots, n) \).

\[
\frac{\Sigma^{-1} \mu}{\sqrt{\mu^T \Sigma^{-1} \mu}} = \frac{\Lambda_{\sigma^{-1} R^{-1} \Lambda_{\sigma^{-1} (\tau \sigma)}}}{\sqrt{(\tau \sigma)^T \Lambda_{\sigma^{-1} R^{-1}} \Lambda_{\sigma^{-1} (\tau \sigma)}}} = \frac{\Lambda_{\sigma^{-1} R^{-1} \tau}}{\sqrt{\tau^T R^{-1} \tau}}
\]

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

\[
t \left( \frac{1}{\sqrt{1^T \mathbf{R} 1}} - \frac{R^{-1} \tau}{\sqrt{\tau^T R^{-1} \tau}} \right) > 0 \tag{5.1.2}
\]
6. 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 in Equation (5.1.1) as:

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

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

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 (6.1), 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 (6.2), 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} (6.3) 

B) We can now assume that (6.3) doesn’t hold:

$$ de \geq 1/c_R $$  \hspace{1cm} (6.4) 

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 (6.1), $h = e - c_R d$.

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 (6.4).

$d$ 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:
\[
\frac{v^T R v}{v^T v} \leq \frac{1^T R 1}{n} \tag{6.5}
\]

The right hand side of Equation (6.5) 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 Equation (6.5) 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.

### 6.1. Illustration for \(n = 2\) Uncorrelated Assets

In this case, according to Equation (5.1.2):

\[
t \left( \frac{1}{\sqrt{2}} - \frac{\tau}{\sqrt{\tau^T \tau}} \right) > 0
\]

We can depict the result geometrically, as shown in Figure 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:

$$P_2 = \frac{\pi}{4} + \theta$$

7. **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.
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.

Figure 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.

![Figure 2](image)

The weights for the tangency portfolio fluctuate wildly. Figure 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 figure shows the clipped distribution with all weights below -1 or above +1 reflected in those final bars.

**Figure 3**

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. Figure 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 Figure 4, showing the relation between the portfolio Sharpe ratio and the stock Sharpe ratio during the same time period. Counter to the usual intuition that tangency outperforms risk parity when equities outperform, we can 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 Figure 4.

Another implication of the theoretical framework above is that risk parity would be closer than ex ante tangency to ex post tangency more than half of the time. Figure 5 calculates the vector angle between the 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 Figure 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.
8. 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. Yet lately it has become a prominent instrument among fund managers and a central topic among academic researchers, due largely to its consistent outperformance.

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.
We formulate and prove several optimal properties of risk parity portfolios including new maximin properties under two scenarios. The maximin approach finds the portfolio that performs the best under the worst possible future distributional parameters of the assets. One such scenario assumes that all assets’ future Sharpe ratios are greater than some positive unknown constant and all correlations are less than another unknown constant. Another scenario 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.

We also show that for a positive constant correlation matrix, assuming that the assets’ future positive Sharpe ratios are chosen at random, the probability that the risk parity portfolio beats by Sharpe ratio any other portfolio is greater than 50%. This novel theoretical result has important practical applications. If the only thing that portfolio manager is able to predict about the future parameters are the ratio of assets variances, then under very general conditions, he should choose risk parity because there is a more than 50% chance that it will beat any other portfolio. Further, we find empirical support for this prediction for a wide variety of time periods when comparing the past several decades of domestic equity and bond returns.

We have thus shown not only when and why risk parity outperforms tangency, but also provided a framework and direction for showing how and when a heuristic approach ignoring some knowledge may in general beat a more traditional optimization approach utilizing all knowledge.

Future research includes adopting this approach to other areas such as risk management, dynamic asset allocation, and derivatives valuation to provide both theoretical justification and empirical validation that heuristics can consistently and strongly outperform, and that sometimes, too much knowledge can be a curse, and ought to be ignored.
References


\* 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). As an aside, in his paper Roy also introduces the concept of Value at Risk. 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.