

Risk Parity Optimality

GREGG S. FISHER, PHILIP Z. MAYMIN,
AND ZAKHAR G. MAYMIN

GREGG S. FISHER
is chief investment officer
of Gerstein Fisher in
New York, NY.
gfisher@gersteinfisher.com

PHILIP Z. MAYMIN
is an assistant professor
of finance and risk engi-
neering at NYU-Poly-
technic Institute in
New York, NY.
philip@maymin.com

ZAKHAR G. MAYMIN
is head of research at
Gerstein Fisher Research
Center and a member of
the investment strategy
group at Gerstein Fisher
in New York, NY.
zmaymin@gersteinfisher.com

One of the most important problems a portfolio manager faces is finding the right weights for portfolio assets. Arthur D. Roy [1952] made a major theoretical development for the solution to this problem by answering the following question: if we know the first two moments of returns—their expected returns and their covariance matrix—what asset weights would maximize the portfolio's mean-volatility ratio? 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 requires knowing the 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; 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 portfolio managers' intuition and experience. 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 in inverse proportion to its future expected volatility. Although 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 et al. [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 et al. [2012] show that leverage aversion can lead to excess returns in a risk parity portfolio, and they document risk parity's historical and sustained out-performance. Here we show that, even if leverage aversion did not apply, risk parity would still beat any other portfolio, on average and under the precise conditions we provide. Thus, our explanation may be viewed as more fundamental.

DeMiguel et al. [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 predominantly 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 knowledge-intensive rules.

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

Of course, risk parity's outperformance is not ubiquitous. Indeed, during 2012, because of bonds' lack-

luster performance, tangency actually beat risk parity. That makes the main questions of this article 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 article 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: $X^T = (X_1, \dots, X_n)$ such that $E(X) = \mu$ and $\text{Var}(X) = \Sigma$, where $\mu^T = (\mu_1, \dots, \mu_n)$, and $\Sigma = \{\sigma_{i,j}\}, i, j = 1, \dots, n$.

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 τ be the assets' Sharpe ratios: $\tau^T = \{\tau_i = \frac{\mu_i}{\sigma_i}, i = 1, \dots, n\}$.

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 Λ_x is the diagonal matrix with vector x on its diagonal. So $[(\sigma)^{-1}]^T \Sigma \sigma^{-1} = \mathbf{1}^T R \mathbf{1}$, where $\mathbf{1}$ is a column vector of ones.

The Sharpe ratio of a portfolio with weights $W^T = (w_1, \dots, w_n)$ 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

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

and the optimal weights w^* could be any weights that are proportional to $w^* \sim \Sigma^{-1} \mu$.

With the normalizing condition $\mathbf{1}^T w = 1$, the optimal weights—the weights of the tangency portfolio—are as follows:

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

For the equally weighted 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 : $v^T = (v_1, \dots, v_n)$ are (by definition) inversely proportional to the asset volatilities:

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

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

$$v_i = \frac{\sigma_i^{-1}}{\sum \sigma_i^{-1}}, \quad i = 1, \dots, n, v = \frac{\sigma^{-1}}{1^T \sigma^{-1}}$$

And its Sharpe ratio is

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

Let us define the equal risk contribution (ERC) portfolio. The volatility of a portfolio with weights u : $u^T = (u_1, \dots, u_n)$ 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

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

Therefore the portfolio's risk (volatility) 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:

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

Two additional constraints are usually enforced: the normalizing constraint, $\sum u_i = 1$, and the constraint forbidding short-selling, $0 \leq u_i \leq 1, i = 1, \dots, n$.

These definitions are not universally accepted. Sometimes equal risk contribution portfolios are called risk parity portfolios, and what we define as a risk parity portfolio is sometimes called a naïve risk parity portfolio.

Actually, it would be more exact and specific to call an RP portfolio a volatility parity portfolio and an ERC portfolio a beta parity portfolio. Here is why (see Maillard et al. [2010]): Denote the covariance between the i th asset and the portfolio by $\sigma_{ip} = \text{cov}(X_i, \Sigma_j [u_j X_j]) = \Sigma_j u_j \sigma_{ij}$. Then $\sigma_i(u) = \frac{u_i \sigma_{ip}}{\sigma(u)}$. By definition, the beta of asset i with the portfolio is $\beta_i = \frac{\sigma_{ip}}{\sigma^2(u)}$. We know that, for the ERC portfolio, $\sigma_i(u) = \frac{\sigma(u)}{n}$ for all $i = 1, \dots, n$. 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, but using betas instead of volatilities.

In a very important general parameter case, the RP portfolio is the same as the ERC portfolio. Maillard et al. [2010] proved that ERC becomes an 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 ERC portfolio's weights are not known in the general case. Chaves et al. [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 game theory concepts to clarify our approach. Let's define a two-player, zero-sum game. Two players are playing a game; the object is to maximize payoff. Each player knows two abstract sets: A and B. Set A is player 1's set of strategies (or actions or decisions), and set B is player 2's strategies. The strategies are also called pure strategies, to distinguish them 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 $\phi(a, b)$, $a \in A$, $b \in B$.

To play the game, player 1 chooses $a \in A$ and player 2 chooses $b \in B$ simultaneously, each unaware of the other's choice. Then their choices are revealed. Player 1 receives $\phi(a, b)$ and player 2 receives $-\phi(a, b)$. The total is zero, which is why this 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; 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 portfolio performance is the best under the market's worst possible action. 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 that will "hurt" the fund manager's performance the most. The manager's performance is measured either by expected return or by Sharpe ratio.

A game is 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 first player's payoff function is defined by the following table:

	b_1	b_2
a_1	1	2
a_2	3	4

Find V_1 , player 1's maximin total gain for the game. 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 ϕ , because for every fixed a , b

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

Taking the maximum on the left and the minimum on the right doesn't change the inequality.

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

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

when this equality holds, it is said that the game has a solution, allowing us to find the game's value 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 2×2 strategies?

A matrix game always has a solution among pure strategies if the matrix has a saddle point, that is, the payoff matrix 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: the value 3. But the following matrix doesn't have a saddle point:

	b_1	b_2
a_1	1	3
a_2	4	2

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

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

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 players' optimal strategy in games with a Nash equi-

librium, when the game has a solution. When the game is analyzed only from the first player's point of view, the strategy that maximizes that player's 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 portfolios without short sales for a set of assets, when a linear inequality constrains their expected returns. We will see that in two natural special cases, the minimax portfolio is the equally weighted or risk parity portfolio.

We start by finding the portfolio that has the best return under the assets' worst distributional assumptions. Let Ω^n be the set of all possible normalized portfolios without short sales:

$$\Omega^n \stackrel{\text{def}}{=} \{w : \Sigma w_i = 1, 0 \leq w_i \leq 1, i = 1, \dots, n\}$$

and let the set of all possible assets' expected values be constrained by a set \mathbb{M}_ϵ^n , which is a set of non-negative vectors above a hyperplane:

$$\mathbb{M}_\epsilon^n \stackrel{\text{def}}{=} \left\{ \mu : \frac{\mu_1}{\epsilon_1} + \dots + \frac{\mu_n}{\epsilon_n} \geq 1, \quad \mu_i \geq 0, \quad i = 1, \dots, n \right\}$$

Then there exists a portfolio with weights $w^* \in \Omega^n$ and returns $\mu^* \in \mathbb{M}_\epsilon^n$ such that

$$\begin{aligned} V &\stackrel{\text{def}}{=} \max_{w \in \Omega^n} \min_{\mu \in \mathbb{M}_\epsilon^n} E_\mu(w^T X) = \min_{\mu \in \mathbb{M}_\epsilon^n} \max_{w \in \Omega^n} E_\mu(w^T X) = E_{\mu^*}(\llbracket(w^*)^T X\rrbracket) \end{aligned} \quad (1)$$

and

$$V = \frac{1}{1^T (\llbracket \epsilon \rrbracket)^{-1}}$$

The portfolio

$$\llbracket(w^*)^T\rrbracket = \frac{\llbracket(\epsilon)\rrbracket^{-1}}{1^T (\llbracket \epsilon \rrbracket)^{-1}} \quad (2)$$

is the only minimax portfolio, such that

$$V = \min_{\mu \in \mathbb{M}_\epsilon^n} E_\mu(\llbracket(w^*)^T X\rrbracket)$$

and the vector μ^*

$$\llbracket(\mu^*)^T\rrbracket = \frac{1^T}{1^T (\epsilon)^{-1}} \quad (3)$$

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

$$V = \max_{w \in \Omega^n} E_{\mu^*}(w^T X) \quad (4)$$

The proof is in Appendix A.

We have also proved more generally that any portfolio is a minimax portfolio for a set of constrained, expected values \mathbb{M}_ϵ^n , if (as shown by Equation (2)) the vector ϵ is inversely proportional to the weights. By analyzing any manager's portfolio, we can make a statement about that manager's view of future expected returns.

Consider the following two important cases.

Minimax Property of Expected Value of Equally Weighted Portfolio

If the sum of assets' non-negative expected returns is greater than a certain (unknown) value, then the equally weighted portfolio is the only minimax portfolio among all portfolios without short sales. 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 equally weighted portfolio: the portfolio that has the greatest expected value under the worst possible scenario.

The proof follows from the minimax property of a general portfolio, if we take all ϵ_i as 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 portfolios without short sales. 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: the 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 ϵ_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 portfolio weights 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: that the market always chooses the distribution that creates the worst possible manager performance. If the portfolio manager's performance is measured by the portfolio's Sharpe ratio, how should the manager choose the portfolio?

We can't directly use the standard game's theoretical approach of mixed strategies, as we did for the minimax results, because we measure a strategy's performance by its Sharpe ratio, not by its expected value.

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 assets' expected returns or the correlations between asset returns.

Yet the manager knows something and chose assets with enough care to be reasonably certain that any asset's worst Sharpe ratio is still positive. In other words, the manager knows that each chosen asset should have a positive expected return but doesn't necessarily know which asset will perform better than the others. Further, the manager also believes that different assets are indeed different, with correlations of 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 portfolios without short sales $w \in \Omega^n$, 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

$$\Omega^n \stackrel{\text{def}}{=} \left\{ w : \sum w_i = 1, 0 \leq w_i \leq 1, i = 1, \dots, n \right\}$$

$$M1 \stackrel{\text{def}}{=} \left\{ \mu : \frac{\mu_i}{\sigma_i} \geq \epsilon > 0, i = 1, \dots, n \right\}$$

$$R1 \stackrel{\text{def}}{=} \left\{ \rho_{i,j \neq i} = \frac{\sigma_{i,j}}{\sigma_i \sigma_j} \leq \delta, 0 < \delta < 1, \rho_{i,i} = 1, i = 1, \dots, n \right\}$$

The proof is in Appendix B.

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.

$$M2 \stackrel{\text{def}}{=} \left\{ \mu : 1^T \frac{\mu}{\sigma} \sum_{i=1}^n \frac{\mu_i}{\sigma_i} \geq \epsilon > 0 \right\}$$

This parameter set describes a situation in which 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. Any particular asset may even have a negative Sharpe ratio, so long as the simple total (or 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 portfolios without short sales $w \in \Omega^n$, such that

$$\min_{\mu \in M2} SR(w^*, \mu, \Sigma) = \max_{w \in \Omega^n} \min_{\mu \in M2} SR(w, \mu, \Sigma)$$

where

$$\Omega^n \stackrel{\text{def}}{=} \left\{ w : \sum w_i = 1, 0 \leq w_i \leq 1, i = 1, \dots, n \right\}$$

$$M2 = \left\{ \mu : 1^T \frac{\mu}{\sigma} \sum_{i=1}^n \frac{\mu_i}{\sigma_i} \geq \epsilon > 0 \right\}$$

The proof is in Appendix C.

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

WHEN RISK PARITY BEATS TANGENCY BY SHARPE RATIO

Say that 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 portfolio weights based on the past expected returns μ and the past variance matrix Σ .

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

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

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

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

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 $\Sigma = \Lambda_{\sigma} R \Lambda_{\sigma}$ and $\Sigma^{-1} = \Lambda_{\sigma^{-1}} R^{-1} \Lambda_{\sigma^{-1}}$ where R is the correlation matrix and Λ_x is the diagonal matrix with vector x on its diagonal. Then $\frac{1^T}{x} \Lambda_x = 1^T$, $\Lambda_x \frac{1}{x} = 1$. So:

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

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

$$\tau \stackrel{\text{def}}{=} \left\{ \tau_i = \frac{\mu_i}{\sigma_i}, i = 1, \dots, n \right\}, t \stackrel{\text{def}}{=} \left\{ t_i = \frac{m_i}{\sigma_i}, i = 1, \dots, n \right\}$$

We already established that RP outperforms any portfolio with weights w by Sharpe if

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

If $\nu \stackrel{\text{def}}{=} w \sigma = \{v_i = w_i \sigma_i, \quad i = 1, \dots, n\}$, then the last inequality can be rewritten as

$$t \sigma^T \left(\frac{\sigma^{-1}}{\sqrt{1^T R 1}} - \frac{\frac{\nu}{\sigma}}{\sqrt{\nu^T R \nu}} \right) > 0$$

or

$$t \left(\frac{1}{\sqrt{1^T R 1}} - \frac{\nu}{\sqrt{\nu^T R \nu}} \right) > 0 \quad (5)$$

If w are the weights of the tangency portfolio, then $((\tau[\sigma])_i = \tau_i \sigma_i, i = 1, \dots, n)$.

$$\frac{w}{\sqrt{w^T \Sigma w}} = \frac{\Sigma^{-1} \mu}{\sqrt{\mu^T \Sigma^{-1} \mu}} = \frac{\Lambda \sigma^{-1} R^{-1} \Lambda \sigma^{-1} (\tau \sigma)}{\sqrt{(\tau \sigma)^T \Lambda_{\sigma^{-1}} R^{-1} \Lambda_{\sigma^{-1}} (\tau \sigma)}} = \frac{\Lambda_{\sigma^{-1}} R^{-1} \tau}{\sqrt{\tau^T R^{-1} \tau}}$$

Therefore RP beats tangency in Sharpe ratio if and only if

$$t \left(\frac{1}{\sqrt{1^T R 1}} - \frac{R^{-1} \tau}{\sqrt{\tau^T R^{-1} \tau}} \right) > 0 \quad (6)$$

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.

$$R_n^+ = \{t_1 \geq 0, \dots, t_n \geq 0\}$$

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 (Equation (5)) as

$$t(e - c_R d) > 0 \quad (7)$$

where $e \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}}, d = \frac{\nu}{|\nu|}$ and $C_R \stackrel{\text{def}}{=} \frac{\sqrt{1^T R 1} |\nu|}{\sqrt{n} \sqrt{\nu^T R \nu}}$.

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 ν) from the hyperplane is longer than the distance of e from the same hyperplane.

Assume for all ν and R that $C_R \geq 1$ (we will prove this statement at the end). Then,

$$d(e - c_R d) = de - c_R < 0 \quad (8)$$

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

$$de < \frac{1}{c_R} \quad (9)$$

B. We can now assume that Equation (9) doesn't hold

$$de \geq \frac{1}{c_R} \quad (10)$$

The distance from a unit vector u to a plane passing through the origin perpendicular to a vector h is $\frac{|uh|}{|h|}$. 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 Equation (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 λ_1 , the maximum eigenvector of matrix R . According to Morrison [1967]: $\lambda_1 = 1 + (n-1)\rho$. That completes the proof.

Illustration for $n = 2$ Uncorrelated Assets

In this case, according to Inequality (6):

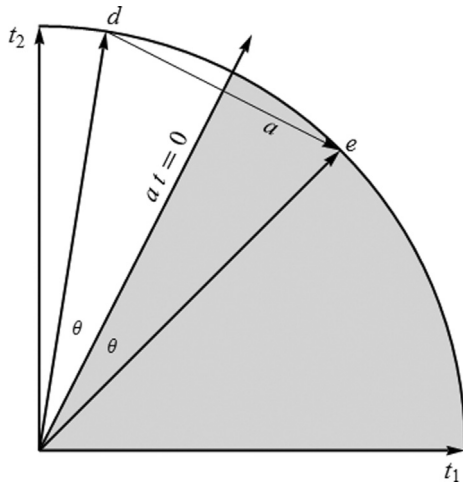
$$t \left(\frac{1}{\sqrt{2}} - \frac{\tau}{\sqrt{\tau^T \tau}} \right) > 0$$

We can depict the result geometrically, as shown in Exhibit 1.

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

EXHIBIT 1

Probability That Risk Parity Beats Tangency for Random Future Sharpe Ratios



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

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

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% in stocks and 40% in bonds. The fixed portfolio may also be viewed as an approximation to the equally weighted portfolio.

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 U.S. Aggregate Bond Index from Dimensional Fund Advisors (DFA) Returns 2.0 software, from February 1988 through October 2012.

Exhibit 2 shows these three portfolios' 24-month rolling Sharpe ratios, 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% and 37.9%, while the tangency portfolio ranged from -8.957%

EXHIBIT 2

Historical Performance Comparison of Three Portfolio Construction Methods

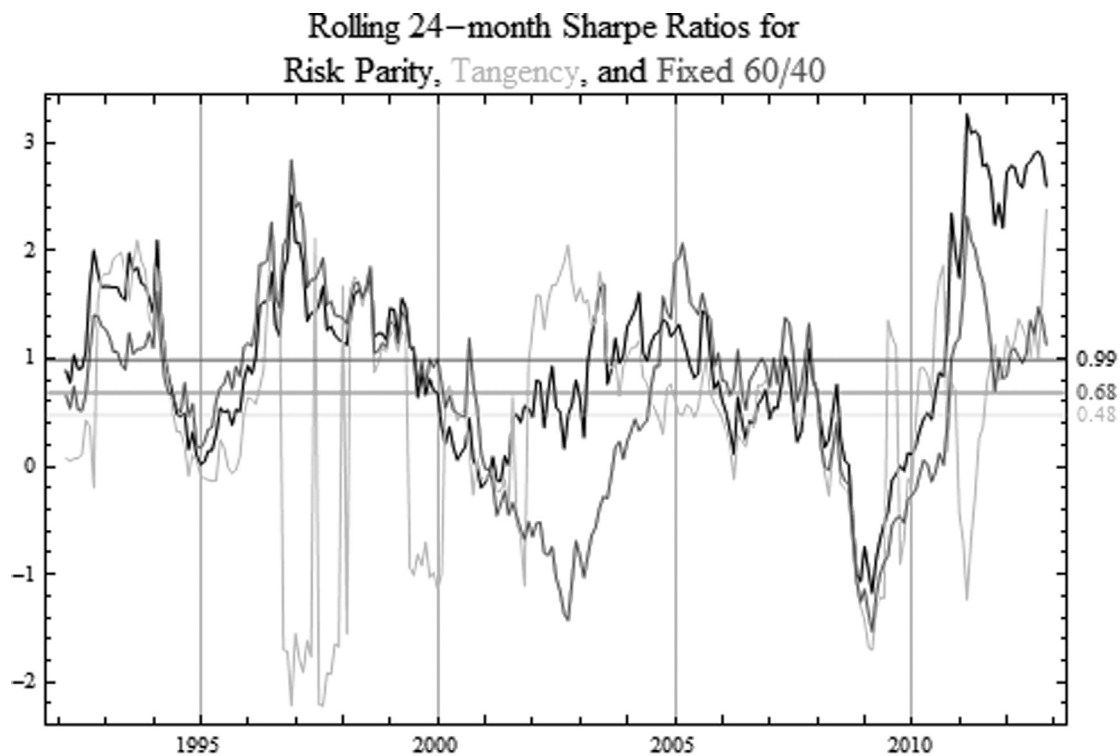
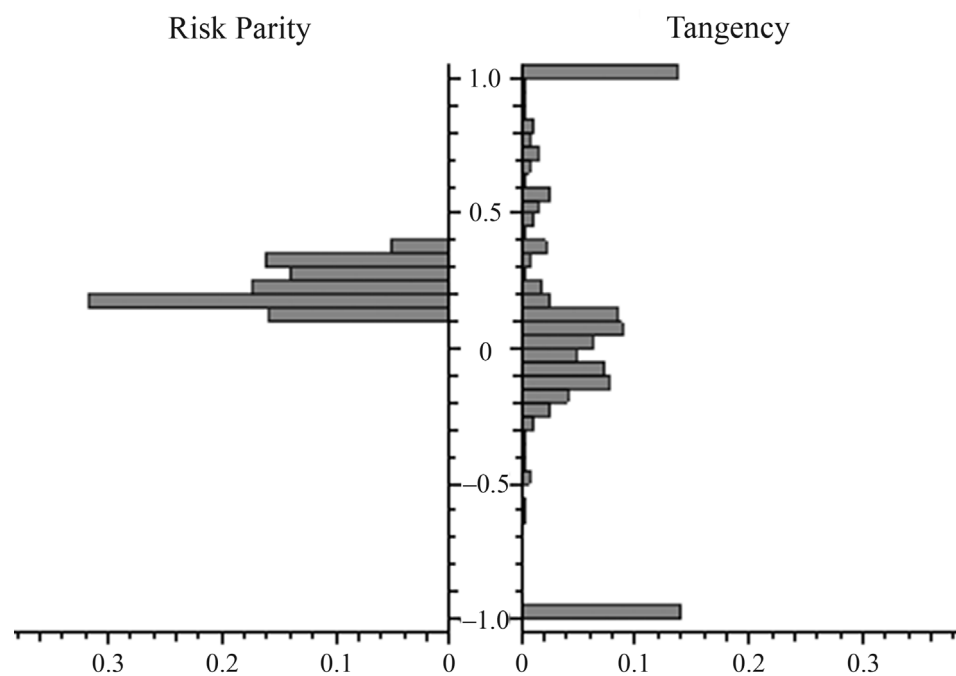


EXHIBIT 3

Paired Histogram of Equity Weightings



to 2.644%; the exhibit shows the clipped distribution, with all weights below -1 or above $+1$ reflected in those final bars.

To test our theoretical framework's implications, we can examine the sensitivity of the performance of the risk parity and tangency portfolios to the underlying assets' performance. Exhibit 4 separately plots the Sharpe ratio of each of the two portfolios, 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. We overlay the best-fit regression line and compute all Sharpe ratios for the same time periods, on a rolling 10-month basis.

Consider the first column in Exhibit 4, showing the relationship 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 stocks' simultaneous performance.

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 this theoretical framework is that risk parity would be closer than ex ante tangency to ex post tangency more than half of the time. Exhibit 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 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%, and the average angle discrepancy is about 10 degrees.

EXHIBIT 4

Sharpe Ratios of Risk Parity and Tangency Portfolios Relative to Assets

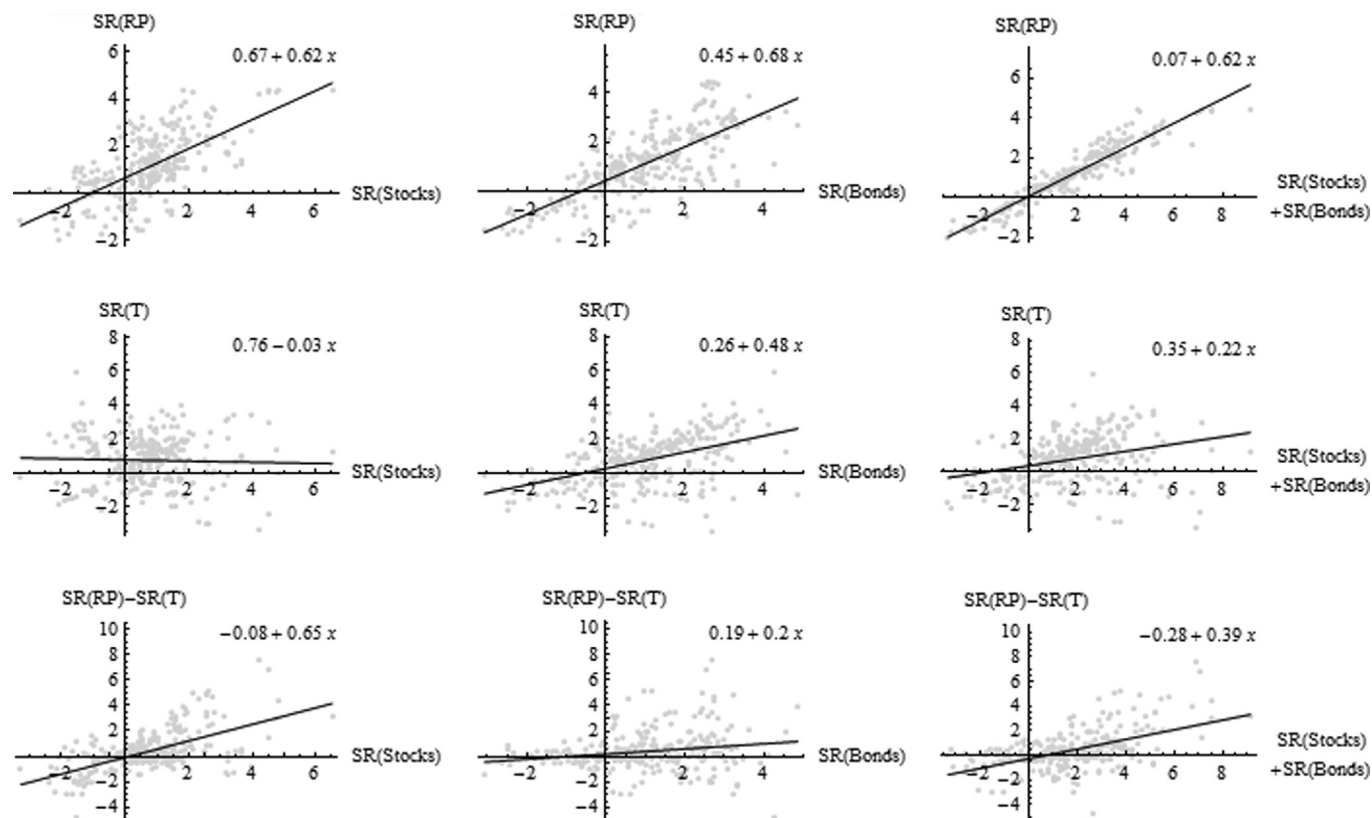
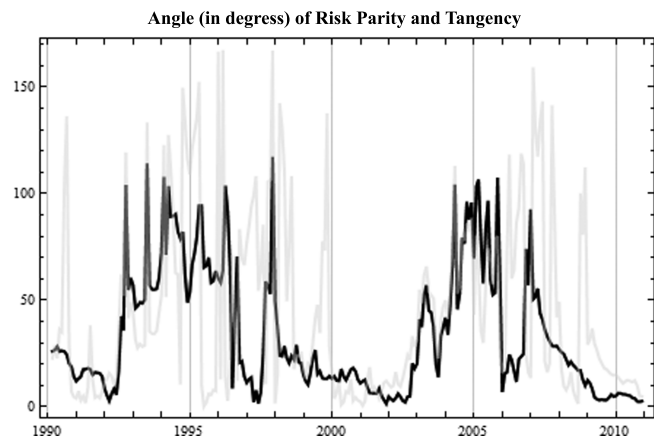


EXHIBIT 5

Vector Angle between Ex Post Tangency and Risk Parity or Ex Ante Tangency



n	$\Pr(\theta_{RP} < \theta_T)$	$\overline{\theta_{RP}}$	$\overline{\theta_T}$
12	57.66%	28.79°	39.25°
24	62.80%	34.13°	45.87°
36	80.97%	37.22°	59.12°
48	75.74%	33.93°	51.86°
60	74.72%	29.77°	56.26°

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 does 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 heuristic sweet spot, where any more or any less knowledge would seem to harm performance.

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

APPENDIX A

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

$$\mathbb{M}_\epsilon^n = \left\{ \mu : \frac{\mu_1}{\epsilon_1} + \dots + \frac{\mu_n}{\epsilon_n} = 1, \mu_i \geq 0, \quad i = 1, \dots, n \right\}$$

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

$$\phi(a_i, b_j) = \begin{cases} \epsilon_i, & i = j \\ 0, & i \neq j \end{cases}, i, j = 1, \dots, n$$

As a matrix game, this game has a solution V . Let

$$w = \{w_1, \dots, w_n\}, w_1 + \dots + w_n = 1, w_i \geq 0, i = 1, \dots, n$$

and

$$\frac{\mu}{\epsilon} = \left\{ \frac{\mu_1}{\epsilon_1}, \dots, \frac{\mu_n}{\epsilon_n} \right\}, \frac{\mu_1}{\epsilon_1} + \dots + \frac{\mu_n}{\epsilon_n} = 1, \mu_i \geq 0, \quad i = 1, \dots, n$$

be arbitrary mixed strategies and $w^* = \{w_1^*, \dots, w_n^*\}$ and $\frac{\mu^*}{\epsilon} = \left\{ \frac{\mu_1^*}{\epsilon_1}, \dots, \frac{\mu_n^*}{\epsilon_n} \right\}$ be the minimax mixed strategies for player 1 and player 2, respectively. Then,

$$V = \sum_{i=1}^n w_i^* \left[\frac{(\mu)_i^*}{\epsilon_i} \right] \epsilon_i = \sum_{i=1}^n w_i^* \mu_i^*$$

That proves Equation (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

$$w_i^* \epsilon_i \geq V, i = 1, \dots, n \quad (\text{A-1})$$

Therefore,

$$V \leq \sum_{i=1}^n w_i^* \epsilon_i \in \left[\frac{(\mu)_i^*}{\epsilon_i} \right] \epsilon_i = \sum_{i=1}^n w_i^* \mu_i^* = V$$

which means that (A-1) are equalities:

$$w_i^* = \frac{V}{\epsilon_i}, i = 1, \dots, n$$

and the value of the game is

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

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

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

or

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

establishing Equation (3). That finishes the proof.

APPENDIX B

Proof. Introducing new variables $a = (a_i = w_i \sigma_i, i = 1, \dots, n)$, 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:

$$\min_{\mu \in M1, R \in R1} SR(w, \mu, \sigma, R) = \frac{a^T \mathbf{1}}{\sqrt{a^T D a}}$$

where D is a correlation matrix with all correlations outside the main diagonal equal to δ .

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 et al. [2010] proved this statement. Kaya and Lee [2012] offered a different proof. We provide yet another, simpler proof.

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

$$D\mathbf{1} = k\mathbf{1}$$

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

Actually, because ours is a portfolio without short sales, 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 portfolios without short sales, and 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. This means that the tangency portfolio has all weights positive, which confirms that it is a portfolio without short sales.

APPENDIX C

Proof. In the worst case, we have

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

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

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

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

$$\frac{1}{f(w)^2} = \frac{1}{\epsilon^2} \frac{w^T \Sigma w}{\left(\min_{i=1, \dots, n} (w_i \sigma_i) \right)^2} \geq \frac{1}{\epsilon^2} \sum_{i,j=1}^n \frac{\sigma_{i,j}}{\sigma_i \sigma_j}$$

The last inequality holds because

$$\left(\min_{i=1, \dots, n} (w_i \sigma_i) \right)^2 \leq (w_j \sigma_j)(w_k \sigma_k), \quad j, k = 1, \dots, n$$

And therefore,

$$f(w) \leq \frac{1}{\sqrt{\sum_{i,j=1}^n \frac{\sigma_{i,j}}{\sigma_i \sigma_j}}}$$

For the RP portfolio P_{w^*} with weights $w^* = \{w_i^* = \frac{c}{\sigma_i}, i = 1, \dots, n\}$ where $c = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i}}$, we have

$$\begin{aligned} \min_{\mu \in M^2} \frac{E_{\mu}(P_{w^*})}{\sqrt{\text{Var}(P_{w^*})}} &= \min_{\mu \in M^2} \frac{\sum_{i=1}^n \frac{\mu_i}{\sigma_i}}{\sqrt{(w^*)^T \Sigma (w^*)}} \\ &= f(w^*) = \frac{1}{\sqrt{\sum_{i,j=1}^n \frac{\sigma_{i,j}}{\sigma_i \sigma_j}}} \end{aligned}$$

which shows that w^* is in fact the value for which function $f(w)$ achieves its maximum.

ENDNOTE

¹Normally, 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.

REFERENCES

Asl, F., and E. Etula. "Advancing Strategic Asset Allocation in a Multi-Factor World." *The Journal of Portfolio Management*, Vol. 39, No. 1 (2012), pp. 59-66.

Asness, C., A. Frazzini, and L. Pedersen. "Leverage Aversion and Risk Parity." *Financial Analysts Journal*, Vol. 68, No. 1 (2012), pp. 47-59.

Britten-Jones, M. "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights." *Journal of Finance*, Vol. 54, No. 2 (1999), pp. 655-671.

Ceria, S., and R. Stubbs. "Incorporating Estimation Errors into Portfolio Selection: Robust Portfolio Construction." Axioma Research Paper No. 3, 2006.

Chaves, D., J. Hsu, F. Li, and O. Shakernia. "Risk Parity Portfolio versus Other Asset Allocation Heuristic Portfolios." *The Journal of Investing*, Vol. 20, No. 1 (Spring 2011), pp. 108-118.

—. "Efficient Algorithms for Computing Risk Parity Portfolio Weights." *The Journal of Investing*, Vol. 21, No. 3 (Fall 2012), pp. 150-163.

DeMiguel, V., L. Garlappi, and R. Uppal. "Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?" *Review of Financial Studies*, Vol. 22, No. 5 (2009), pp. 1915-1953.

Goldstein, D., and G. Gigerenzer. "Fast and Frugal Forecasting." *International Journal of Forecasting*, 25 (2009), pp. 760-772.

Gigerenzer, G., P. Todd, and ABC Research Group. *Ecological Rationality: Intelligence in the World*. New York: Oxford University Press, 2012.

Kaya, H., and W. Lee. "Demystifying Risk Parity." White paper, Neuberger Berman (2012).

Markowitz, H. "Portfolio Selection." *The Journal of Finance*, Vol. 7, No. 1 (1952), pp. 77-91.

Maillard, S., T. Roncalli, and J. Teiletche. "The Properties of Equally Weighted Risk Contribution Portfolios." *The Journal of Portfolio Management*, Vol. 36, No. 4 (Summer 2010), pp. 60-70.

Meucci, A. *Risk and Asset Allocation*. New York: Springer, 2007.

Merton, R. "On Estimating the Expected Return on the Market: An Exploratory Investigation." *The Journal of Financial Economics*, 8 (1980), pp. 323-361.

Morrison, D. *Multivariate Statistical Methods*. New York: McGraw-Hill, 1967.

Nash, J. "Non-Cooperative Games." *Annals of Mathematics*, Vol. 54, No. 2 (1951), pp. 286-295.

von Neumann, J. "Zur Theorie der Gesellschaftsspiele." *Mathematische Annalen*, 100 (1928), pp. 295-300.

Roy, A. "Safety First and the Holding of Assets." *Econometrica*, Vol. 20, No. 3 (1952), pp. 431-449.

Sharpe, W. "Mutual Fund Performance." *Journal of Business*, Vol. 39, No. 1 (1966), pp. 119-138.

Scherer, B. "Can Robust Portfolio Optimization Help Build Better Portfolios?" *Journal of Asset Management*, 7 (2007), pp. 374-387.

Sullivan, E. "A.D. Roy: The Forgotten Father of Portfolio Theory." In Jeff E. Biddle, Ross B. Emmett, ed. *Research in the History of Economic Thought and Methodology*. Bradford, U.K.: Emerald Group Publishing Limited, 29, (2011), pp. 73-82.

Zweig, J. "Investing Experts Urge 'Do as I Say, Not as I Do'." *Wall Street Journal*, January 3, 2009.

To order reprints of this article, please contact Dewey Palmieri at dpalmieri@iijournals.com or 212-224-3675.