On the Optimality of Long–Short Strategies

Bruce I. Jacobs, Kenneth N. Levy, and David Starer

We consider the optimality of portfolios not subject to short-selling constraints and derive conditions that a universe of securities must satisfy for an optimal active portfolio to be dollar neutral or beta neutral. We find that following the common practice of constraining long–short portfolios to have zero net holdings or zero betas is generally suboptimal. Only under specific unlikely conditions will such constrained portfolios optimize an investor's utility function. We also derive precise formulas for optimally equitizing an active long–short portfolio using exposure to a benchmark security. The relative sizes of the active and benchmark exposures depend on the investor's desired residual risk relative to the residual risk of a typical portfolio and on the expected risk-adjusted excess return of a minimum-variance active portfolio. We demonstrate that optimal portfolios demand the use of integrated optimizations.

The construction and management of long–short portfolios are complicated tasks involving assumptions and actions that may seem counterintuitive to the investor unfamiliar with shorting. Despite attempts by Jacobs and Levy (1996b, 1997) to clarify the issues, many practitioners—even some of the most experienced—have been beguiled by an assemblage of myths and misconceptions. With long–short strategy becoming an increasingly important component of institutional portfolios,[1] some of the more egregious misunderstandings must be purged from the collective psyche of the investment community.

One myth that many practitioners evidently believe (see, for example, Michaud 1993 and Arnott and Leinweber 1994) is that an optimal long–short portfolio can be constructed by blending a short-only portfolio with an independently generated long-only portfolio.[2] Adherents to this belief tend to characterize the overall portfolio in terms of the excess returns of, and correlation between, the two constituent portfolios. One of the reasons such an approach is suboptimal (see Jacobs and Levy 1995) is that it fails to use the correlations between the *individual* (long and short) securities to achieve an overall reduction in variance.

Another myth is that a long–short portfolio represents a separate asset class. This misconception is common. For example, Brush (1997) described a technique for optimally blending a long–short portfolio with a long-only portfolio to achieve an overall portfolio that has a greater Sharpe ratio than either of its constituent portfolios. In so doing, Brush implicitly assigned long–short and long-only portfolios to different asset classes. Although this blending approach appears to acknowledge the benefits of long–short investment, it misses the points that a long–short portfolio does not belong to a separate asset class and that combining a long–short portfolio with a long-only portfolio produces (in the aggregate) only a single portfolio! The optimal weights of that single portfolio should be obtained from an integrated optimization. The important question is not how one should allocate capital between a long-only portfolio and a long–short portfolio but, rather, how one should blend active positions (long and short) with a benchmark security in an integrated optimization.

In addition to falling victim to such myths, some practitioners have followed common practices that may not be optimal. For example, they often seek to constrain their portfolios to be neutral with respect to some factor (that is, to be independent of, or insensitive to, that factor).[3] In particular, they often constrain their portfolios to be dollar neutral by committing the same amount of capital to their long holdings as they commit to their short holdings. In so doing, in a naive sense, they set their net market exposure to zero. Another constraint often imposed is that of beta neutrality, in which

Bruce I. Jacobs and Kenneth N. Levy, CFA, are principals and David Starer is a senior quantitative analyst at Jacobs Levy Equity Management.

the manager constrains the portfolio to have a beta of zero. Such a beta-neutral portfolio is theoretically insensitive to market movements.

The manager may apply neutrality constraints voluntarily or because the client requires them. But although valid taxation, accounting, or behavioral reasons may exist for imposing such constraints, there are generally no pressing financial reasons for doing so.[4] On the contrary, imposing them may actually prevent managers from fully using their insights to produce optimal portfolios. A general principle of optimization is that constrained solutions do not offer the same level of utility as unconstrained solutions unless, by some fortunate coincidence, the optimum lies within the feasible region dictated by the constraints. Given that neutrality is often imposed, we consider here the conditions under which this coincidence can occur. That is, we set out to find the conditions under which dollar-neutral or beta-neutral portfolios are optimal.

When Treynor and Black (1973) discussed similar issues in a classic paper, they posed the following question: "Where practical is it desirable to so balance a portfolio between long positions in securities considered underpriced and short positions in securities considered overpriced that market risk is completely eliminated?" (p. 66). This article tackles Treynor and Black's question and extends the analysis to the following:

1. Under what conditions will a net holding of zero (i.e., dollar neutrality) be optimal for a long–short portfolio?
2. Under what conditions will the combined optimal holdings in a long–short portfolio be beta neutral?
3. How should one optimally equitize a long–short portfolio? In particular, under what conditions will dollar neutrality or beta neutrality be optimal for the active portion of an equitized long–short portfolio?

This article is essentially divided into two parts. The first part considers an active portfolio (which we define as one that has no explicit benchmark holding), and the second part considers an equitized active portfolio (which we define as one that consists of the active portfolio combined with an explicit exposure to the benchmark security). The first part is concerned mainly with risk and return in an absolute sense, and the second part is concerned mainly with risk and return in a relative sense.

Within this framework, we first consider the optimality of dollar and beta neutrality in active long–short portfolios. We then reconsider dollar neutrality and beta neutrality in portfolios designed to minimize residual risk and in portfolios designed to maximize return subject to a constraint on residual risk. Finally, we extend the analysis to consider equitized portfolios. We examine the optimality of dollar neutrality and beta neutrality for the active portion of an equitized long–short portfolio, and we show how optimal exposure to the benchmark security should be computed.

Portfolio Construction and Problem Formulation

In answering the first two questions posed in the introduction, we assume that the investor has solved the usual expected utility maximization problem and that the solution permits shorting. We determine what properties the universe of investment opportunities should possess for the portfolio resulting from the maximization problem to be dollar neutral or beta neutral. To answer the third question, we set up an integrated criterion function and examine its properties.

We will be concerned mainly with variations of the utility function favored by Markowitz (1952) and Sharpe (1991):

$$U = r_P - \frac{1}{2\tau}\sigma_P^2, \tag{1}$$

where r_P is the expected return on the investor's portfolio, σ_P^2 is the variance of the return, and τ is the investor's risk tolerance. For mathematical convenience, we have included a factor of one-half in the utility function. This utility function can be considered an approximation to the investor's expected utility in the sense of von Neumann and Morgenstern (1944). As Sharpe (1991) pointed out, if the investor has a negative exponential utility function over wealth and if returns are jointly normally distributed, then the approximation will be exact. Moreover, Levy and Markowitz (1979) showed that the approximation is good even if the investor has a more general utility function or if returns are not jointly normally distributed or both.

Assume that, in seeking to maximize the utility function in Equation 1, the investor has an available capital of K dollars and has acquired n_i shares of security $i \in \{1, 2, \ldots, N\}$. A long holding is represented by a positive number of shares, and a short holding is represented by a negative number.[5] The holding h_i in security i is the ratio of the amount invested in that security to the investor's total capital. Thus, if security i has price p_i, then $h_i = n_i p_i / K$.

In addition to the N securities, assume also that the investor may have an exposure of K_B dollars to a benchmark security. We are intentionally vague about the nature of the benchmark security to emphasize that long–short portfolios are neutral and can be transported to any asset class by use of appropriate overlays. Thus, the benchmark security may be an equity index, a debt index, or any other instrument that the investor cares to specify. The holding of the benchmark security is $h_B = K_B/K$.[6] The investor seeks to maximize the utility function given in Equation 1 by choosing appropriate values for security holdings h_i.

Unlike the typical optimization problem for a fully invested portfolio, our utility function is not augmented with a constraint to ensure that the total holdings sum to unity. Instead, the long–short portfolio is constrained only by U.S. Federal Reserve Board Regulation T, which states that the total value of the investment should not exceed twice the investor's capital.[7] To express this constraint mathematically, we define a long set, L, and a short set, S, such that

$$L = \{i: n_i > 0\} \text{ and } S = \{i: n_i < 0\}.$$

Regulation T states that each investor must satisfy the following inequality:

$$\sum_{i \in L} n_i p_i - \sum_{i \in S} n_i p_i \le 2K.$$

This inequality need not be included explicitly in the optimization because the relative sizes of holdings are unaffected by it and all holdings can simply be scaled up or down so that it is satisfied.

Optimal Long–Short Portfolios

As discussed, many long–short investment approaches create suboptimal portfolios because they prepartition the problem. That is, they combine a long portfolio with an independently generated short portfolio, and they characterize the long–short portfolio in terms of the correlation between the two constituent portfolios. In contrast, our approach treats the portfolio as a single entity. Unlike Michaud and Arnott and Leinweber, we exploit the correlations between all of the individual securities (whether they are held long or sold short) in a single integrated optimization.

Consider first portfolios that have no explicit position in the benchmark security. Let r_i be the expected return on security i. Using matrix notation, the absolute return on the active portfolio is then

$$r_P = h^T r, \tag{2}$$

where $h = [h_1, h_2, \ldots, h_N]^T$ is a vector of holdings, $r = [r_1, r_2, \ldots, r_N]^T$ is a vector of returns, and the superscript T denotes matrix or vector transposition.

In this analysis, we ignore risk-free holdings. If we were to consider them, however, they would simply result in the addition of the term $h_F r_F$ to the expression for the portfolio return.

The variance of the portfolio's absolute return is

$$\sigma_P^2 = h^T Q h, \tag{3}$$

where $Q = \mathrm{cov}(r, r^T)$ is the covariance matrix of the individual securities and is assumed to be known.

Substituting Equation 2 for the portfolio return and Equation 3 for the variance into the utility function (Equation 1), differentiating the utility with respect to holding vector h (see, for example, Magnus and Neudecker 1988), setting this derivative equal to zero, and solving for h produces the optimal weight vector

$$h = \tau Q^{-1} r. \tag{4}$$

This form is typical for the expression for an optimal portfolio, and it shows that the best mix of risky assets in an investor's portfolio depends only on the expected returns and their covariances. The investor's wealth and preferences affect only his or her demand for risky assets through a *scalar* τ that is the same for all risky assets.

As with the portfolio given by Equation 4, optimal security weights in many portfolio problems turn out to be proportional to the securities' expected returns and inversely proportional to the covariance of the returns. In addition to maximizing the utility function of Equation 1, appropriately scaled versions of Equation 4 also give the optimal portfolio weights for such problems as maximizing the Sharpe ratio (Sharpe 1994), minimizing portfolio variance while holding portfolio expected return fixed (Treynor and Black), and maximizing expected return subject to a constraint on variance.

We will find it useful to define the portfolio of Equation 4 with $\tau = 1$ as the unit-risk-tolerance active (URA) portfolio, ϕ. That is,

$$\phi \equiv Q^{-1} r.$$

The expected absolute return of this portfolio is

$$r_{URA} = r^T Q^{-1} r,$$

and the variance of this portfolio's absolute return is

$$\sigma_{URA}^2 = r^T Q^{-1} r.$$

Optimality of Dollar Neutrality. Consider now the conditions under which a portfolio would be dollar neutral. The net holding H is the sum of all the individual holdings,

$$H = \sum_{i=1}^{N} h_i = \mathbf{1}_N^T h, \tag{5}$$

where $\mathbf{1}_N$ represents an $N \times 1$ vector of ones. Substituting Equation 4 into Equation 5 leads to the following expression for the net holding:

$$H = \tau \mathbf{1}_N^T Q^{-1} r. \tag{6}$$

For the portfolio to be dollar neutral, the value of the long holdings must equal the negative of the value of the short holdings. By using the definitions of the long and short sets, this equality is expressed mathematically as

$$\sum_{i \in L} h_i = -\sum_{i \in S} h_i.$$

Equivalently, because L and S are exhaustive, the sum of the weights must be zero and the general condition for dollar neutrality is

$$H = 0. \tag{7}$$

The logical argument attached to Equation 7 must be kept clearly in mind. The condition expressed in the equation is necessary but not sufficient for an optimal portfolio to be dollar neutral. Thus, if the condition holds, the optimal portfolio must be dollar neutral. One can, however, construct a portfolio that is dollar neutral (and thus satisfies Equation 7) but not optimal.

For the specific portfolio under consideration, substituting Equation 6 into Equation 7 gives the following condition for optimal dollar neutrality:

$$\tau \mathbf{1}_N^T Q^{-1} r = 0. \tag{8}$$

This general condition for dollar neutrality can be simplified by making various assumptions about the structure of covariance matrix Q. For example, one special case arises if one subscribes to the assumptions of the constant correlation model of Elton, Gruber, and Padberg (1976), under which the elements of the covariance matrix are given by

$$q_{ij} = \begin{cases} \rho\sigma_i\sigma_j, & i \neq j \\ \sigma_i^2, & i = j \end{cases},$$

where σ_i is the standard deviation of the return of the ith security and ρ is a constant correlation factor. Equivalently, in the Elton, Gruber, and Padberg model, the covariance matrix can be written in matrix notation as

$$Q = (1-\rho)D_\sigma + \sigma\rho\sigma^T, \tag{9}$$

where D_σ is a diagonal matrix having the variances $\sigma_i^2; i = 1, \ldots, N$ along its diagonal and σ is a vector of standard deviations: $\sigma = [\sigma_1, \sigma_2, \ldots, \sigma_N]^T$. The covariance matrix as written in Equation 9 is in a convenient form for application of the *matrix inversion lemma*.

The matrix inversion lemma (see, for example, Kailath 1980) states that for compatibly dimensioned matrixes W, X, Y, and Z,

$$[W + XYZ]^{-1} = W^{-1} - W^{-1}X \times [Y^{-1} + ZW^{-1}X]^{-1} ZW^{-1}. \tag{10}$$

Using this lemma to invert the covariance matrix in Equation 9 and substituting the result into Equation 6 for the net holding produces

$$H = \frac{\tau}{1-\rho}\left[\mathbf{1}_N^T D_\sigma^{-1} r - \frac{\rho}{1-\rho+\rho\sigma^T D_\sigma^{-1}\sigma} \times \left(\mathbf{1}_N^T D_\sigma^{-1}\sigma\right)\left(\sigma^T D_\sigma^{-1} r\right)\right]. \tag{11}$$

One can easily verify the following identities:

$$\mathbf{1}_N^T D_\sigma^{-1} r = \sum_{i=1}^{N} \left(r_i/\sigma_i^2\right),$$

$$\mathbf{1}_N^T D_\sigma^{-1} \sigma = \sum_{i=1}^{N} (1/\sigma_i),$$

$$\sigma^T D_\sigma^{-1} r = \sum_{i=1}^{N} (r_i/\sigma_i),$$

$$\sigma^T D_\sigma^{-1} \sigma = \sum_{i=1}^{N} \left(\sigma_i^2/\sigma_i^2\right) = N.$$

Thus, Equation 11 reduces to

$$H = \frac{\tau}{1-\rho}\left[\sum_{i=1}^{N} \frac{r_i}{\sigma_i^2} - a\sum_{i=1}^{N} \frac{r_i}{\sigma_i}\right], \tag{12}$$

where

$$a = \frac{\rho}{1 + N\rho - \rho} \sum_{i=1}^{N} \frac{1}{\sigma_i}.$$

Intuition concerning Equation 12 can be obtained by defining a measure of return stability, ξ_i, as the inverse of the standard deviation of the return of security i. Then, for portfolios with many securities (i.e., those with large N), the constant a is approximately equal to the average return stability. That is,

$$a = \frac{\rho}{1 + N\rho - \rho} \sum_{i=1}^{N} \frac{1}{\sigma_i} \approx \frac{1}{N} \sum_{i=1}^{N} \xi_i = \bar{\xi}.$$

Using this approximation in Equation 12 makes the net holding

$$H \approx \frac{\tau}{1-\rho}\sum_{i=1}^{N}(\xi_i - \bar{\xi})\frac{r_i}{\sigma_i}. \tag{13}$$

Thus, if the net risk-adjusted return of all securities weighted by the deviation of their stability from average is positive, the net holding should be long. Conversely, if this quantity is negative, the net holding should be short. Only under the special condition in which H in Equation 12 is equal to zero will the optimal portfolio be dollar neutral. Constraining the holding to be zero when this condition is not satisfied will produce a suboptimal portfolio.[8]

Equation 13 formalizes the simple intuitive notion that you should be net long if you expect the market as a whole to go up and net short if you expect it to go down! Importantly, however, it tells you how long or how short your net exposure should be based on your risk tolerance, your predictions of security returns and standard deviations, and your estimate of the correlation between security returns.

Equation 13 and the requirement that $H = 0$ can also be used in a normative sense. For example, because Equation 13 is independent of the individual holdings, an investor could select a universe of securities such that, based only on their expected risk-adjusted returns and return stability, the net holding of the universe as computed with Equation 13 is zero. The investor could then be confident that the portfolio formed from this universe that maximizes the utility function (Equation 1) will be dollar neutral.

More precise conditions that an optimal portfolio must satisfy to be dollar neutral can be obtained by making further assumptions about Equation 12. For example, assuming that $\rho \neq 1$ and $\tau \neq 0$ gives

$$\sum_{i=1}^{N}\frac{r_i}{\sigma_i^2} = a\sum_{i=1}^{N}\frac{r_i}{\sigma_i}. \tag{14}$$

A sufficient (but not necessary) condition for Equation 14 to hold is that both sums in the equation be zero simultaneously. Each of these sums can be regarded as a form of net risk-adjusted return that, if equal to zero, results in zero net holding being optimal. Alternatively, in the (admittedly unlikely) circumstance that all variances are equal, Equation 14 for optimal dollar neutrality is satisfied if the sum of the returns is zero. Roughly, in this case, the portfolio should have zero net holding if the average return is zero.

Optimality of Beta Neutrality. In an exactly analogous manner to the preceding analysis, we consider in this section the conditions under which an unconstrained portfolio would optimally have a beta of zero. Because we are dealing here with beta sensitivity, it is appropriate to use Sharpe's diagonal model, which gives the expected return of the ith security, r_i, in terms of the alpha of that security, α_i, and beta of that security, β_i, and the expected return of the benchmark security, r_B:

$$r_i = \alpha_i + \beta_i r_B.$$

When this model is used, the beta of the portfolio is

$$\beta_P = \sum_{i=1}^{N} h_i\beta_i = \beta^T h, \tag{15}$$

where $\beta = [\beta_1, \beta_2, \ldots, \beta_N]^T$. The covariance matrix of the security returns is

$$Q = D_\omega + \beta\sigma_B^2\beta^T,$$

where D_ω is a diagonal matrix whose ith diagonal entry is $\omega_i^2 = \text{var}(\alpha_i)$, and $\sigma_B^2 = \text{var}(r_B)$. The diagonal form of this matrix is consistent with the model's assumption that the correlation between any pair of stock return residuals is zero. Using the matrix inversion lemma (Equation 10), the inverse of the covariance matrix is

$$Q^{-1} = D_\omega^{-1} - \frac{D_\omega^{-1}\beta\beta^T D_\omega^{-1}}{\sigma_B^{-2} + \beta^T D_\omega^{-1}\beta}. \tag{16}$$

Using Equation 4 in Equation 15 and setting the portfolio beta equal to zero gives the following general condition for optimality of beta neutrality:

$$\beta^T Q^{-1} r = 0. \tag{17}$$

Then, if Equation 16 is used, the condition shown in Equation 17 becomes

$$\left(\sigma_B^{-2} + \beta^T D_\omega^{-1}\beta\right)\left(\beta^T D_\omega^{-1} r\right) = \left(\beta^T D_\omega^{-1}\beta\right)\left(\beta^T D_\omega^{-1} r\right). \tag{18}$$

The two conditions under which Equation 18 is satisfied are the following: Either

$$\sigma_B^{-2} + \beta^T D_\omega^{-1}\beta = \beta^T D_\omega^{-1}\beta,$$

which would require $\sigma_B^2 = \infty$, and is thus untenable, or

$$\beta^T D_\omega^{-1} r = 0.$$

This second condition, rewritten as a summation, implies that the condition under which an optimal portfolio has zero beta is

$$\sum_{i=1}^{N} \frac{\beta_i r_i}{\omega_i^2} = 0. \tag{19}$$

The left-hand side of Equation 19 can be interpreted as a beta-weighted, risk-adjusted net return. If this quantity is positive, then the optimal portfolio will have a positive beta. Conversely, if this quantity is negative, the optimal portfolio will have a negative beta. Constraining the portfolio beta to be zero when Equation 19 is not satisfied will result in suboptimal portfolio construction.

If one uses the Elton, Gruber, and Padberg approximation for the covariance matrix, one can show that an alternative condition for beta neutrality to be optimal is

$$\beta_p \approx \frac{\tau}{1-\rho} \sum_{i=1}^{N} (\beta_i - \bar{\beta}) \frac{r_i}{\sigma_i} = 0,$$

where beta is a volatility-weighted average beta,

$$\bar{\beta} = \frac{1}{N} \sum_{i=1}^{N} \frac{\beta_i}{\sigma_i}.$$

This expression is analogous to Equation 13 and shows that the portfolio beta is optimally zero when the net risk-adjusted return of all securities weighted by the deviation of their betas from the average is zero.

We have dealt thus far only with absolute return and absolute variance. Most plan sponsors and investment managers, however, are concerned with relative measures rather than absolute measures. In particular, they are interested in maximizing return in excess of a benchmark return while simultaneously minimizing residual risk. In the next section, we extend the previous results to portfolios formed by optimizing such relative measures.

Optimal Long–Short Portfolio with Minimum Residual Risk. The excess return of a portfolio, r_E, is simply $r_A - r_B$, the portfolio's absolute return minus the benchmark return.[9] The residual risk is the variance of the excess return, and can be shown to be

$$\sigma_E^2 = h^T Q h - 2h^T q + \sigma_B^2,$$

where $q = \text{cov}(r, r_B)$ is a column vector of covariances between the individual security returns and the benchmark return. The active portfolio that minimizes the residual risk can be shown to be $h = Q^{-1}q$. Defining this portfolio as the minimum-residual-risk (MRR) portfolio, ψ, will be useful; that is,

$$\psi \equiv Q^{-1} q.$$

This portfolio's absolute return is

$$r_{MRR} = q^T Q^{-1} r,$$

and its residual risk, the minimum attainable with an unequitized portfolio, is

$$\sigma_{MRR}^2 = \sigma_B^2 - q^T Q^{-1} q.$$

Using the same type of analysis as in the previous section, we can state the condition for such a portfolio to be dollar neutral optimally as

$$H \approx \frac{1}{1-\rho} \sum_{i=1}^{N} (\xi_i - \bar{\xi}) \frac{q_i}{\sigma_i} = 0$$

or

$$\sum_{i=1}^{N} \frac{q_i}{\sigma_i^2} = a \sum_{i=1}^{N} \frac{q_i}{\sigma_i}.$$

Thus, the minimum-residual-risk (or minimum-tracking-error) portfolio will optimally be dollar neutral if the net risk-adjusted covariance of the securities' returns with the benchmark return, weighted by the deviations of the returns' stability from the average, is zero.

To find the condition for the optimality of beta neutrality, observe that

$$\begin{aligned} q &= \text{cov}(r, r_B) \\ &= \beta \sigma_B^2, \end{aligned}$$

so

$$\begin{aligned} \psi &= Q^{-1} q \\ &= \sigma_B^2 Q^{-1} \beta, \end{aligned}$$

and the beta of the portfolio is

$$\begin{aligned} \beta_P &= \psi^T \beta \\ &= \sigma_B^2 \beta^T Q^{-1} \beta. \end{aligned}$$

Because Q is positive definite, so too is Q^{-1}. Thus, β_P cannot be zero for any nonzero β.

For the specific case using the Sharpe diagonal model, the preceding expressions can be used to find that the condition for a minimum-excess-variance portfolio to be optimally beta neutral is

$$\sum_{i=1}^{N} \frac{\beta_i^2}{\omega_i^2} = 0,$$

but this equation cannot be satisfied by any portfolio that contains even one security with a nonzero beta. Thus, we reach the conclusion that no practical active portfolio that minimizes residual risk can optimally be beta neutral. This conclusion accords with intuition: A portfolio that minimizes residual risk should have a beta that approaches one, not zero.

Optimal Long–Short Portfolio with Specified Residual Risk. Typically, a plan sponsor gives a manager a mandate to maximize return on a portfolio and simultaneously demands that the standard deviation or variance of that return equal some specified level.[10] For the manager, this task amounts to choosing, at each investment period, a portfolio that optimizes the Lagrangian

$$l = r_E - \lambda\left(\sigma_E^2 - \sigma_D^2\right),$$

where σ_D^2 is the desired excess variance (i.e., residual risk) and λ is a Lagrange multiplier.

Although this approach differs slightly from the more traditional approach of Black (1972), which seeks to minimize variance subject to a constraint on excess return, we believe that the problem posed as return maximization subject to a constrained risk level is a more accurate reflection of the thought processes of plan sponsors and investment managers.

The portfolio that optimizes this Lagrangian can be shown to be

$$h = k\phi + \psi,$$

where ϕ is the unit-risk-tolerance active (URA) portfolio, ψ is the minimum-residual-risk (MRR) portfolio, and

$$k = \sqrt{\frac{\sigma_D^2 - \sigma_{MRR}^2}{\sigma_{URA}^2}}.$$

The optimal portfolio in this case is the sum of the MRR portfolio and a scaled version of the URA portfolio. The scaling factor depends on the desired residual risk, the minimum attainable residual risk, and the variance of the URA portfolio. If the desired residual risk is less than the minimum attainable residual risk, then $\sigma_D^2 - \sigma_{MRR}^2 < 0$ and no portfolio can be constructed. If the desired residual risk is equal to the minimum attainable residual risk, then $\sigma_D^2 - \sigma_{MRR}^2 = 0$ and the optimal portfolio will be simply $h = \psi$, the minimum-residual-risk portfolio. As the desired residual risk increases, the portfolio becomes more like a scaled version of ϕ (the URA portfolio) and k tends asymptotically to the investor's risk tolerance, τ.

The condition under which this portfolio is optimally dollar neutral again has the familiar form

$$H \approx \frac{1}{1-\rho}\sum_{i=1}^{N}(\xi_i - \bar{\xi})\frac{kr_i + q_i}{\sigma_i} = 0, \tag{20}$$

or

$$\sum_{i=1}^{N}\frac{kr_i + q_i}{\sigma_i^2} = a\sum_{i=1}^{N}\frac{kr_i + q_i}{\sigma_i},$$

indicating that this portfolio is optimally dollar neutral if a net risk-adjusted linear combination of the securities' returns and covariances, weighted by deviation of return stability from average, is zero. The interpretation of Equation 20 is similar to that of Equation 13, where the term $kr_i + q_i$ now replaces r_i and the presence of k and q_i reflects the investor's concerns about residual risk.

Similarly, the condition under which this portfolio will optimally be beta neutral is

$$\sum_{i=1}^{N}\frac{\beta_i}{\omega_i^2}(kr_i + q_i) = 0;$$

equivalently, because $r_i = \alpha_i + \beta_i r_B$ and $q_i = \beta_i\sigma_B^2$,

$$k\sum_{i=1}^{N}\frac{\alpha_i\beta_i}{\omega_i^2} + \left(\sigma_B^2 + kr_B\right)\sum_{i=1}^{N}\frac{\beta_i^2}{\omega_i^2} = 0.$$

Optimal Equitized Long–Short Portfolio

We now address the third question posed in the introduction, namely: How should one optimally equitize a long–short portfolio? In this case, in addition to the long–short portfolio, the manager has an explicit benchmark exposure, either through ownership of a physical benchmark security or through a derivative overlay. We determine the optimal portfolio weights and the optimal benchmark exposure in a single integrated step. This approach differs from the approach used by Brush (1997), in which security weights were predetermined for two distinct portfolios—a long–short portfolio and a long-only portfolio—and then capital was allocated between these two existing portfolios. In Brush, the long-only portfolio served to provide both security and benchmark exposure whereas the long–short portfolio provided security but not benchmark exposure.

Treynor and Black showed that, under the assumptions of the diagonal model, an equitized long–short portfolio can be viewed *conceptually* as the outcome of the following separate decisions: selecting an active portfolio to maximize an appraisal ratio, blending the active portfolio with a suitable replica of the market portfolio to maximize the Sharpe ratio, and scaling the positions in the combined portfolio through lending or borrowing while preserving their proportions. These separate decisions are of a different nature from those of Brush. Treynor and Black arrived at the conceptual separability only after performing an explicit integrated optimization in which security positions (long and short) and benchmark exposure were determined jointly.

Treynor and Black showed, among other things, that a security may play two roles simultaneously: (1) a position based entirely on the security's expected independent return (appraisal premium) and (2) a position based solely on the security's role as part of the market portfolio. These two roles must be considered when blending individual security positions with a benchmark exposure. In this section, we derive expressions for the optimal benchmark holding that implicitly account for this dual nature of securities.

The absolute return on the equitized portfolio now includes a contribution from the return on the benchmark security and is, therefore, given by

$$r_P = h^T r + h_B r_B.$$

The excess return on the equitized portfolio is

$$r_E = h^T r + h_B r_B - r_B = \mathbf{h}^T \mathbf{r},$$

where the augmented holding vector, $\mathbf{h}$, and the augmented return vector, $\mathbf{r}$, for the equitized portfolio are defined as

$$\mathbf{h} = \begin{bmatrix} h \\ \hbar_B \end{bmatrix}; \ \mathbf{r} = \begin{bmatrix} r \\ r_B \end{bmatrix}$$

with $\hbar_B = h_B - 1$. Note that the augmented vectors (which are distinguished from the active portfolio vectors by the use of bold font) incorporate the corresponding active portfolio holding and return vectors.

The variance of the excess return of the equitized portfolio, σ_E^2, is

$$\sigma_E^2 = \mathrm{var}(r_E) = \mathbf{h}^T \mathbf{Q} \mathbf{h},$$

where $\mathbf{Q}$ is the covariance matrix of the augmented return vector $\mathbf{r}$.[11] Noting that $\mathbf{r}$ is a partitioned vector, we can also write $\mathbf{Q}$ in the following partitioned form:

$$\mathbf{Q} = \begin{bmatrix} Q & q \\ q^T & \sigma_B^2 \end{bmatrix}.$$

Optimality of Dollar Neutrality with Equitization. In this section, we consider the active portion of the equitized long–short portfolio and determine the conditions under which that portion is optimally dollar neutral. As before, we consider an unconstrained portfolio designed to maximize the investor's utility. In the presence of equitization, the utility of interest is the portfolio's excess return tempered by the variance of its excess return. Specifically, the objective function to be maximized is

$$J = \mathbf{h}^T \mathbf{r} - \frac{1}{2\tau} \mathbf{h}^T \mathbf{Q} \mathbf{h},$$

where, as before, τ is the risk tolerance of the investor.

By differentiating this objective function with respect to $\mathbf{h}$ and setting the derivative equal to zero, the benchmark and active portfolio weights are found to be

$$\hbar_B = -\tau m \ \text{ or } \ h_B = 1 - \tau m$$

and

$$h = \tau Q^{-1}(r + mq) = (\phi + m\psi)\tau.$$

The scalar m is given by

$$m = \frac{r_{MRR} - r_B}{\sigma_{MRR}^2}.$$

The net holding in the active part of the portfolio is obtained by summing the components of h to give

$$H = \mathbf{1}^T h = \tau \mathbf{1}^T Q^{-1}(r + mq).$$

This quantity will be zero if dollar neutrality is optimal.

Using the constant correlation model discussed previously to provide more specific results for the inverse covariance matrix, we find the net holding to be

$$H \approx \frac{\tau}{1 - \rho} \sum_{i=1}^{N} (\xi_i - \bar{\xi}) \frac{r_i + mq_i}{\sigma_i}.$$

This holding is exactly analogous to the holdings given in Equations 13 and 20. As in those equations, the net holding will be zero when the weighted average of a particular set of risk-adjusted returns is zero. As before, the weighting is the deviation of the stability of each security's return from the average stability. In this case, however, the particular risk-adjusted return includes one part equal to the security's return and a second part equal to a scaled version of the security's correlation with the benchmark security. The scaling, m, depends on the return and variance of the minimum-residual-risk portfolio relative to the return and variance of the benchmark security.

Optimality of Beta Neutrality with Equitization. Following the method discussed in the section on beta neutrality, and using the expressions derived previously, we find that the condition for the active portion of an equitized long–short portfolio to be optimally beta neutral is

$$\sum_{i=1}^{N} \frac{\beta_i}{\omega_i^2}(r_i + mq_i) = 0.$$

Equivalently, because $q_i = \beta_i \sigma_B^2$ and $r_i = \alpha_i + \beta_i r_B$, the condition for the active portion of an equitized long–short portfolio to be optimally beta neutral is

$$\sum_{i=1}^{N} \frac{\beta_i}{\omega_i^2}\left[\alpha_i + \beta_i\left(r_B + m\sigma_B^2\right)\right] = 0.$$

Optimal Equitized Long–Short Portfolio with Specified Residual Risk. For this problem, we define an optimal portfolio to be one that maximizes expected excess return while keeping the variance of the excess return (i.e., the residual risk) equal to some specified or desired level. To find the portfolio, we form the following Lagrangian:

$$\begin{aligned} l &= r_E - \frac{1}{2}\lambda\left(\sigma_E^2 - \sigma_D^2\right) \\ &= h^T r - \frac{1}{2}\lambda\left(\mathbf{h}^T\mathbf{Q}\mathbf{h} - \sigma_D^2\right). \end{aligned}$$

Differentiating the Lagrangian with respect to $\mathbf{h}$ and λ and setting these derivatives equal to zero yields

$$\mathbf{r} = \lambda\mathbf{Q}\mathbf{h} \tag{21}$$

and

$$\mathbf{h}^T\mathbf{Q}\mathbf{h} = \sigma_D^2. \tag{22}$$

Solving Equation 21 for $\mathbf{h}$, substituting this solution into Equation 22, and noting that $\mathbf{Q}$ is Hermitian,[12] we arrive at the following solution for the optimal equitized portfolio:

$$\mathbf{h} = \frac{1}{\lambda}\mathbf{Q}^{-1}\mathbf{r}, \tag{23}$$

where

$$\frac{1}{\lambda} = \frac{\sigma_D}{\sqrt{\mathbf{r}^T\mathbf{Q}^{-1}\mathbf{r}}}. \tag{24}$$

Although Equation 23 enables one to compute the optimal holdings, it does not provide much intuition about the benchmark holding.

We now derive an explicit expression for the optimal benchmark exposure from which we can draw insight. First, use the definitions of $\mathbf{r}$, $\mathbf{h}$, and $\mathbf{Q}$ to rewrite Equation 21 as the following set of equations:

$$Qh + q\hbar_B = \frac{1}{\lambda}r \tag{25}$$

and

$$q^T h + \sigma_B^2 \hbar_B = \frac{1}{\lambda}r_B. \tag{26}$$

Then, solving for h from Equation 25, substituting this solution into Equation 26, and rearranging gives the optimal benchmark holding as

$$\begin{aligned} h_B &= 1 + \hbar_B \\ &= 1 - \frac{\sigma_D}{\sqrt{\mathbf{r}^T\mathbf{Q}^{-1}\mathbf{r}}}\left(\frac{q^T Q^{-1} r - r_B}{\sigma_B^2 - q^T Q^{-1} q}\right). \end{aligned} \tag{27}$$

To attach intuition to Equation 27, it is convenient to state a number of definitions and associations. Define θ to be the unit-risk-tolerance equitized (URE) portfolio that optimizes the unconstrained mean–variance criterion function $J = \mathbf{h}^T\mathbf{r} - \frac{1}{2}\mathbf{h}^T\mathbf{Q}\mathbf{h}$. This portfolio is

$$\theta = \mathbf{Q}^{-1}\mathbf{r}.$$

Its expected excess return and the variance of that return are

$$\begin{aligned} r_{URE} &= \mathbf{r}^T\mathbf{Q}^{-1}\mathbf{r} \\ &= \mathbf{r}^T\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{r} \\ &= \theta^T\mathbf{Q}\theta \\ &= \sigma_{URE}^2. \end{aligned} \tag{28}$$

This variance is the term under the radical in the denominator of Equation 27.

Using the definitions of σ_{URE}^2, σ_{MRR}^2, and r_{MRR} in Equation 27 gives the following equation:

$$h_B = 1 - \frac{\sigma_D}{\sigma_{URE}}\left(\frac{r_{MRR} - r_B}{\sigma_{MRR}^2}\right), \tag{29}$$

from which we can make the following qualitative inferences:

- The quantity in parentheses can be regarded as the risk-adjusted excess return of the minimum-residual-risk portfolio, and the benchmark holding should clearly decrease as this quantity increases. The following specific comments apply:
  1. Generally, $r_{MRR} > r_B$, so the expression in parentheses in Equation 29 is positive.
  2. As the return of the minimum-residual-risk portfolio, r_{MRR}, increases or the return of the benchmark security, r_B, decreases, the holding in the benchmark security should decrease.
  3. As the minimum residual risk, σ_{MRR}^2, increases, the holding of the benchmark should increase.
- The weight in the benchmark security is generally negatively related to the desired residual risk; that is, as the desired residual risk, σ_D, increases, the holding in the benchmark should decrease. If no excess variance can be tolerated,

$\sigma_D = 0$ and $h_B = 1$, so the portfolio should be fully invested in the benchmark. If the investor desires a large residual risk in pursuit of high returns, the benchmark portfolio weight can decrease to less than zero and the investor should sell the benchmark security short.

- The ratio σ_D / σ_{URE} is an important determinant of the relative size of the benchmark holding. It is the ratio of the investor's desired residual risk to the residual risk of a portfolio that a unit-risk-tolerant investor would choose. As the ratio increases, the optimal benchmark holding generally decreases.

Regarding the active portfolio, h, note that the preceding definitions substituted into Equation 25 lead to

$$h = \frac{\sigma_D}{\sigma_{URE}}(\phi + m\psi).$$

As before, the optimal active holding is a function of the unit-risk-tolerance active portfolio and the minimum-residual-risk portfolio. As σ_D / σ_{URE} approaches zero, the optimal holdings in the active portfolio tend to zero. As before, with a requirement for zero excess variance, the optimal holding is a full exposure to the benchmark.

Optimal Equitized Long–Short Portfolio with Constrained Beta. In addition to being required to produce portfolios that maximize return while keeping residual risk at a prescribed level, managers are typically expected to keep the betas of their portfolios very close to one. If a portfolio beta differs significantly from one, the manager may be viewed as taking undue risk or attempting to time the market.

These requirements are captured in the following Lagrangian:

$$l = r_E + \lambda_1\left(\sigma_E^2 - \sigma_D^2\right) + \lambda_2\left(\beta_P - \beta_D\right),$$

where the λs are Lagrange multipliers and β_D is the desired portfolio beta (usually equal to one). This Lagrangian can be optimized with respect to the unknown parameters, but the resulting solution is algebraically untidy and does not provide much insight. Instead, an intuitive result can be achieved by examining the constraint on the portfolio's beta. Specifically, the beta of the portfolio is

$$\beta_P = \sum_{i=1}^{N} h_i \beta_i + h_B,$$

and substituting this expression into the constraint on the portfolio beta gives

$$h_B = \beta_D - \sum_{i=1}^{N} h_i \beta_i = \beta_D - \beta_A, \tag{30}$$

where β_A is the beta of the active portfolio.

An intuitive explanation of Equation 30 is that with a constraint on the portfolio's beta, the benchmark holding is simply the difference between the desired beta and the beta of the active portfolio. One extreme case corresponds to a desired portfolio beta of one and an active portfolio beta of zero; under these conditions, the benchmark holding must be one. That is, the manager should be exposed to the benchmark to the full value of the capital under management.

Conclusion

We derived conditions that a universe of securities must satisfy for an optimal portfolio constructed from that universe to be dollar neutral or beta neutral. Using criterion functions that are most often used in practical investment management, we found conditions under which optimal portfolios become dollar or beta neutral. Only in fairly restrictive cases will optimal portfolios satisfy these conditions. Generally, an optimal long–short portfolio will be dollar neutral if the risk-adjusted returns of its constituent securities, weighted by the deviation of those securities' returns from average, sum to zero. This condition can be used to select a universe of securities that will naturally form a dollar-neutral optimal portfolio. Analogous conditions must hold for a long–short portfolio to be beta neutral.

We next considered optimal equitized portfolios and derived conditions under which the active portion of such portfolios will be dollar neutral or beta neutral. We derived an expression for the holding of a benchmark security that sets the residual risk of an equitized long–short portfolio equal to a desired value while simultaneously maximizing the portfolio's return. We showed that the optimal holding of the benchmark security depends on such parameters as the ratio of the desired residual risk level to the residual risk level of a portfolio that a unit-risk-tolerant investor would choose and the risk-adjusted excess return of the minimum-variance active portfolio over the benchmark return. The benchmark holding should decrease in the following circumstances: when the investor's appetite for residual risk increases, when the expected return of the minimum-variance active portfolio increases, when the variance of the minimum-variance active portfolio decreases, or when the expected return of the benchmark portfolio decreases. The portfolio should be fully equitized when the investor has no appetite for residual risk or when the active portfolio has a zero beta and the equitized portfolio is to be constrained to have a beta of one.

Optimal portfolios demand the use of integrated optimization. In the case of active long–short portfolios, the optimization must consider all individual securities (both long and short) simultaneously, and in the case of equitized long–short portfolios, this consideration must also encompass the benchmark security.

Notes

1. Recent tax rulings have made long–short investing more attractive to certain classes of investors than in the past. For example, borrowing *cash* to purchase stock (i.e., debt financing through margin purchases) can give rise to a tax liability for tax-exempt investors. However, according to a January 1995 Internal Revenue Service ruling (IRS Ruling 95-8), borrowing *stocks* to initiate short sales does not constitute debt financing, so profits realized when short sales are closed out are not considered unrelated business taxable income (UBTI). Furthermore, the August 1997 rescission of the "short-short" rule has enabled mutual funds to implement long–short investing. Under IRS Code sec. 851(b)(3), the short-short rule had required that in order to qualify for tax pass-throughs, a mutual fund must have derived less than 30 percent of its gross income from positions held less than three months. This rule severely restricted funds' ability to sell short, because profits from closing short positions were considered to be short-term gains and thus included in this provision.
2. The practice of blending separate long and short portfolios may have arisen from investors with traditional long-only managers adding a dedicated short seller either to neutralize market risk or to enhance overall portfolio return.
3. Portfolios can be constrained to be neutral with respect to any particular factor, such as interest rates. Furthermore, portfolios can be constrained to be insensitive to several factors simultaneously. We focus on dollar neutrality and beta neutrality because they appear to be of greatest interest to investors. Application of our results to other cases is straightforward.
4. As discussed in Note 1, from a taxation perspective, interest indebtedness generates UBTI for tax-exempt investors. For instance, a 200 percent long position would give rise to margin debt in the amount of 100 percent of capital, which would generate UBTI. But investing capital both 100 percent long and 100 percent short incurs no interest indebtedness while providing the maximum amount of leverage under U.S. Federal Reserve Board Regulation T. From an accounting perspective, balanced long and short positions can easily be monitored. Because true parameter values are unknown and can be estimated only with uncertainty, market neutrality is problematic. Thus, investors may be more comfortable with the accounting certainty of dollar balance. From a behavioral and "mental accounting" perspective, investors can easily categorize all beta-neutral long–short portfolios as market neutral and may prefer knowing that certain "pockets" of assets are neutralized from market movements—especially when the investor wants to separate the security selection decision and the derivative overlay decision.
5. As described by Sharpe (1991), "A 'short position' is achieved by borrowing an asset such as a share of stock, with a promise to repay in kind, typically on demand. The borrowed asset is then sold, generating a cash receipt. If the proceeds of the sale may be used for other types of investment, the overall effect is equivalent to a negative holding of [the borrowed asset]" (p. 500).
6. In general, we use lower-case subscripts to refer to a generic security and upper-case subscripts to refer to particular entities. Thus, for example, the subscript i indicates that the variable under consideration is an unspecified security i. The subscript B refers to a particular chosen benchmark, and P refers to the particular portfolio.
7. Regulation T represents an institutional friction. In this analysis, it conveniently drops out of the specification of the problem, and the analysis continues to be consistent with the assumption in Note 5. For a review of the institutional aspects of the market, see Jacobs and Levy (1997).
8. It can be shown that the proportional change in utility when the portfolio is constrained to be dollar neutral is $\Delta U/U = -(\mathbf{1}^T Q^{-1} r)^2/[(\mathbf{1}^T Q^{-1}\mathbf{1})(r^T Q^{-1} r)]$. This change has a maximum value of zero (which occurs when the condition for dollar neutrality is satisfied) and is otherwise always negative.
9. Strictly, the excess return is $r_E = [(1 + r_A)/(1 + r_B)] - 1$, but the two measures of excess return are similar for small constituent returns and the expression used in the text is more convenient arithmetically.
10. Sponsors are often content with a specification of residual risk and are concerned with risk taking that exceeds the specified level or with closet indexing, where risk is below the intended level. Jacobs and Levy (1996a) showed that enhanced passive searches that consider exclusively managers having risk of a certain level or less are suboptimal.
11. Our approach is valid for the usual case in which the benchmark return cannot be expressed as a linear combination of the returns of the individual securities in the portfolio. If the benchmark return can be expressed in such a way (for example, if the portfolio consists of every single one of the securities used to construct the benchmark), then the augmented covariance matrix is singular and an analogous but slightly different approach must be taken to find the optimal portfolio.
12. A Hermitian matrix is one that is equal to its transpose (or conjugate transpose if it is complex). Because $\mathbf{Q}$ is Hermitian, $(\mathbf{Q}^{-1})^T\mathbf{Q}$ is equal to the identity matrix and cancels out during derivation of Equation 24.

References

Arnott, Robert D., and David J. Leinweber. 1994. "Long–Short Strategies Reassessed." Letter to the editor. *Financial Analysts Journal,* vol. 50, no. 5 (September/October):76–80.

Black, Fischer. 1972. "Capital Market Equilibrium with Restricted Borrowing." *Journal of Business,* vol. 45, no. 3 (July):444–55.

Brush, John S. 1997. "Comparisons and Combinations of Long and Long/Short Strategies." *Financial Analysts Journal*, vol. 53, no. 3 (May/June):81–89.

Elton, Edwin J., Martin J. Gruber, and Manfred W. Padberg. 1976. "Simple Criteria for Optimal Portfolio Selection." *Journal of Finance,* vol. 31, no. 5 (December):1341–57.

Jacobs, Bruce I., and Kenneth N. Levy. 1995. "More on Long–Short Strategies." Letter to the editor. *Financial Analysts Journal,* vol. 51, no. 2 (March/April):88–90.

———. 1996a. "Residual Risk: How Much Is Too Much?" *Journal of Portfolio Management,* vol. 22, no. 3 (Spring):10–15.

———. 1996b. "20 Myths about Long–Short." *Financial Analysts Journal,* vol. 52, no. 5 (September/October):81–85.

———. 1997. "The Long and Short on Long–Short." *Journal of Investing,* vol. 6, no. 1 (Spring):73–86.

Kailath, Thomas. 1980. *Linear Systems.* Englewood Cliffs, NJ: Prentice-Hall.

Levy, H., and Harry Markowitz. 1979. "Approximating Expected Utility by a Function of Mean and Variance." *American Economic Review,* vol. 69, no. 3 (June):308–17.

Magnus, Jan R., and Heinz Neudecker. 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester, U.K.: John Wiley & Sons.

Markowitz, Harry. 1952. "Portfolio Selection." *Journal of Finance,* vol. 7, no. 1 (March):77–91.

Michaud, Richard O. 1993. "Are Long–Short Equity Strategies Superior?" *Financial Analysts Journal,* vol. 49, no. 6 (November/December):44–49.

Sharpe, William F. 1991. "Capital Asset Prices with and without Negative Holdings." *Journal of Finance,* vol. 46, no. 2 (June):489–509.

———. 1994. "The Sharpe Ratio." *Journal of Portfolio Management,* vol. 21, no. 1 (Fall):49–58.

Treynor, Jack L., and Fischer Black. 1973. "How to Use Security Analysis to Improve Portfolio Selection." *Journal of Business,* vol. 46, no. 1 (January):66–86.

von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.