FAJ
Financial Analysts Journal
Volume 62 • Number 2
©2006, CFA Institute

Trimability and Fast Optimization of Long–Short Portfolios

Bruce I. Jacobs, Kenneth N. Levy, CFA, and Harry M. Markowitz

Optimization of long–short portfolios through the use of fast algorithms takes advantage of models of covariance to simplify the equations that determine optimality. Fast algorithms exist for widely applied factor and scenario analysis for long-only portfolios. To allow their use in factor and scenario analysis for long–short portfolios, the concept of "trimability" is introduced. The conclusion is that the same fast algorithms that were designed for long-only portfolios can be used, virtually unchanged, for long–short portfolio optimization—provided the portfolio is trimable, which usually holds in practice.

Long–short portfolios can take many forms, including market-neutral equity portfolios that have a zero market exposure and enhanced active equity portfolios that have a full market exposure, such as 120–20 portfolios (with 120 percent of capital long and 20 percent short). We describe a sufficient condition under which a portfolio optimization algorithm designed for long-only portfolios will find the correct long–short portfolio, even if the algorithm's use would violate certain assumptions made in the formulation of the long-only problem.[1] We refer to this condition as the "trimability condition." The trimability condition appears to be widely satisfied in practice.

We also discuss the incorporation of practical and regulatory constraints into the optimization of long–short portfolios. A common assumption of some asset-pricing models is that one can sell a security short without limit and use the proceeds to buy securities long. This assumption is mathematically convenient, but it is unrealistic. In addition, actual constraints on long–short portfolios change over time and, at a given instant, vary from broker to broker and from client to client. The portfolio analyst charged with generating an efficient frontier must take these constraints into account. To our knowledge, all such constraints—whether imposed by regulators, brokers, or the investors themselves—are expressible as linear equalities or weak inequalities. Therefore, they can be incorporated into the general portfolio selection model.

Bruce I. Jacobs is principal at Jacobs Levy Equity Management, Florham Park, New Jersey. Kenneth N. Levy, CFA, is principal at Jacobs Levy Equity Management, Florham Park, New Jersey. Harry M. Markowitz is principal at Harry Markowitz Company, San Diego, California.

In the upcoming sections, we define the general mean–variance problem and outline some of the constraints on portfolio composition in the real world. We then show how the general mean–variance problem can be solved rapidly with a factor, scenario, or historical model by diagonalization of the covariance matrix. We next present the modeling of long–short portfolios and derive a condition under which these fast optimization techniques apply. And we illustrate the results.

General Mean–Variance Problem

Consider a portfolio consisting of n securities with expected returns $\mu_1, \mu_2, \ldots, \mu_n$. The portfolio can include both risky and riskless securities. The portfolio's expected return, E_P, is a weighted sum of the n security returns:

$$E_P = \sum_{i=1}^{n} x_i \mu_i, \tag{1}$$

where $x_1, x_2, \ldots, x_n$ are the security weights in the portfolio. If the covariance between the returns of security i and security j is σ_{ij}, the portfolio's return variance, V_P, is

$$V_P = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j. \tag{2}$$

In addition, security weights may be subject to various constraints. For long-only portfolios, common constraints include the following:

$$\sum_{k=1}^{n} a_{jk} x_k = b_j, \text{ for } j = 1, \ldots, m, \tag{3}$$

and

$$x_i \geq 0, \text{ for } i = 1, \ldots, n, \tag{4}$$

where m is the number of constraints.[2] Equation 3 might include, for example, a budget constraint according to which the sum of the weights must equal a fixed number. Equation 4 is a nonnegativity constraint.

The general single-period mean–variance portfolio selection problem is to find all efficient portfolios (characterized in terms of portfolio expected return E_P and associated portfolio return variance V_P) for all security expected returns μ_i and covariances σ_{ij} in Equations 1 and 2 and all constraint coefficients a_{jk} and b_j in Equations 3 and 4.

Long–Short Constraints in Practice

For a long–short portfolio, the sign of x_i is not constrained. A negative value of x_i is interpreted as a short position.[3] In addition, the capital asset pricing model (CAPM) often assumes that the long–short portfolio is subject only to the full investment constraint,

$$\sum_{i=1}^{n} x_i = 1. \tag{5}$$

Equation 5 is unrealistic as a sole constraint, however, because it permits a portfolio such as the following:

$$x_i = \begin{cases} z & \text{for } i = 1 \\ 1 - z & \text{for } i = 2 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

for all real z. In such a portfolio, an investor could, for example, deposit \$1,000 with a broker, short \$1,000,000 of Stock A, and use the proceeds plus the original deposit to purchase \$1,001,000 of Stock B. Short positions do not, in fact, work this way.

No single constraint set applies to all long–short investors. The portfolio analyst must model the specific set of constraints for the particular investor. To our knowledge, however, all relevant constraints on long–short portfolios can be accommodated if one adopts the convention of representing an n-security long–short portfolio in terms of $2n$ nonnegative variables, $x_1, \ldots, x_{2n}$, in which the first n variables represent the securities in a given set held long, the second n variables represent short sales in the *same set* of securities, and one chooses the long–short portfolio subject to the following constraints:[4]

$$\sum_{k=1}^{2n} a_{jk} x_k = b_j, \text{ for } j = 1, \ldots, m, \tag{7}$$

and

$$x_i \geq 0, \text{ for } i = 1, \ldots, 2n. \tag{8}$$

For the remainder of this article, we assume that long-only portfolios are subject to the constraints of Equations 3 and 4 and long–short portfolios are subject to the constraints of Equations 7 and 8. To illustrate what these assumptions may involve, we outline a few real-world short-sale constraints.

Constraint Equations 7 and 8 subsume the budget (or full investment) constraint, with $a_{jk} = 1$ and $b_j = 1$ for all k and for one value of j chosen to be the index of the equation that implements the constraint. Similarly, constraint Equations 7 and 8 may include upper and lower bounds on any particular security. For example, an upper bound of U on the short selling of security i is accomplished by including a new nonnegative slack variable, x_s, setting $b_j = U$, and letting $a_{jk} = 1$ if $k \in \{n + i, s\}$ or $a_{jk} = 0$ otherwise.

Another important constraint, one that is also a special case of Equations 7 and 8, is Regulation T of the U.S. Federal Reserve Board. Reg T margin requirements apply to common stock, convertible bonds, and equity mutual funds. Reg T requires that the sum of the long positions plus the sum of the (absolute value of) short positions not exceed twice the equity in the account. Using the convention of representing a long–short portfolio of n securities in terms of $2n$ nonnegative variables, a generalized form of Reg T requires that

$$\sum_{i=1}^{2n} x_i \leq H. \tag{9}$$

Reg T currently specifies $H = 2$. As a matter of policy, the broker or investor may set H at a lower level. This inequality can be converted to an equality by introduction of a slack variable.[5]

Constraint Equation 9 is a special case of the following more general constraint:

$$\sum_{i=1}^{2n} m_i x_i \leq 1,$$

where m_i represents the net margin requirement of the ith position. This constraint is more general than Equation 9, in that it permits a net short margin requirement that differs from the long margin requirement and it allows inclusion of securities that are exempt from Reg T requirements.[6] The inequality in the more general constraint can be converted into an equality with the use of a nonnegative slack variable.

Yet another important constraint is the requirement that the total value of the long positions minus the total value of the short positions be close to some investor-specified value v. That is,

$$\left|\left(\sum_{i \in L} x_i - \sum_{i \in S} x_i\right) - v\right| \le \tau \tag{10}$$

for some given small nonnegative tolerance level τ, where L is the set of risky securities held long and S is the set of risky securities sold short.

Two strategies encompassed by constraint Equation 10 are market-neutral equity strategies and enhanced active equity strategies. In market-neutral equity strategies, the sum of the long positions equals the sum of the short positions.[7] That is, $v = 0$. In enhanced active equity strategies, such as 120–20 strategies, the portfolio maintains a full market exposure and $v = 1$ (see Jacobs and Levy forthcoming 2006).

There may be additional constraints on the choice of the portfolio. For example, some securities are hard to borrow, so the broker may limit the amount of the short position or not permit short positions in a particular security.

Thus, budget constraints, upper and lower bounds on long and short positions, equality constraints on particular positions, market-neutrality constraints, enhanced active equity constraints, and generalized Reg T types of constraints can all be written in the form of constraint Equations 7 and 8. In addition, this $2n$ formulation can eliminate unrealistic portfolios, such as that in Equation 6. An apparent disadvantage of constraint Equations 7 and 8, insofar as portfolio optimization is concerned, is that they allow long and short positions in the same security. We consider this issue in more detail later.

Diagonalized Models of Covariance

In general, the covariances, σ_{ij}, in Equation 2 are nonzero because the return of any security has at least some relationship to the return of any other security. The covariance matrix will be dense, therefore, with as many nonzero covariances as there are pairs of securities.

Markowitz (1959) showed that the solution of the general mean–variance portfolio selection problem requires the inversion of the covariance matrix.[8] This inversion is one of the major computational burdens in portfolio optimization, and to ease this burden, fast portfolio optimization algorithms have been devised that use rapid methods of inversion. Many rapid methods are derived by constructing a mathematical model of the covariance matrix in such a way that the portfolio selection problem is transformed into a problem requiring only the inversion of a diagonal matrix. A diagonal matrix has zeros everywhere except along the main diagonal; such matrices are particularly easy to invert.

In this section, we show how three types of models—factor models, scenario models, and historical models—can be used to transform the portfolio selection problem into one requiring the inversion of a diagonal (or nearly diagonal) matrix. In problems with a large number of securities, computation time may differ by orders of magnitude between using a dense covariance matrix and using a diagonal or nearly diagonal covariance matrix. We consider long-only portfolios here and extend the results to long–short portfolios in the following section.

Factor Models. A factor model of covariance assumes that the return on a security depends linearly on the movement of one or more factors common to many securities (the general market return, interest rates, etc.) plus the security's independent idiosyncratic term. Specifically, it assumes that the return on the ith security is

$$r_i = \alpha_i + \sum_{k=1}^{K} \beta_{ik} f_k + u_i, \text{ for } i = 1, \ldots, n, \tag{11}$$

where α_i is a constant, f_k is the return on the kth common factor, β_{ik} is the factor loading, K is the number of common factors, and u_i is an idiosyncratic term assumed to be uncorrelated with u_j for all $i \neq j$ and uncorrelated with all f_k for $k = 1, \ldots, K$. For simplicity, we also assume that f_k is uncorrelated with f_j for $j \neq k$.[9]

To perform the diagonalization, one introduces fictitious securities, one for each common factor (see Sharpe 1963; Cohen and Pogue 1967), with the weight of each fictitious security constrained to be a linear combination of the weights of the real securities. Accordingly, one defines a set of K fictitious securities with weights $y_1, \ldots, y_K$ in terms of the real securities as follows:[10]

$$y_k = \sum_{j=1}^{n} x_j \beta_{jk}, \text{ for } k = 1, \ldots, K. \tag{12}$$

With this definition, the portfolio variance can be written (see Jacobs, Levy, and Markowitz 2005) in the form

$$V_P = \sum_{i=1}^{n} x_i^2 V_i + \sum_{k=1}^{K} y_k^2 W_k, \tag{13}$$

where W_k is the variance of f_k. Equation 13 expresses V_P as a positively weighted sum of squares in the n

original securities and K new fictitious securities, which are linearly related to the original securities by Equation 12.

Note that the variance expression in Equation 13 contains only two single sums (whereas the variance expression in Equation 2 contained a nested double sum). Therefore, Equation 13 can be written in terms of a diagonal covariance matrix.

Scenario Models. As in the scenario models analyzed by Markowitz and Perold (1981a, 1981b), we assume that one of S mutually exclusive scenarios will occur with probability P_s, where $s = 1, \ldots, S$. If scenario s occurs, the return of the ith security is

$$r_{is} = \mu_{is} + u_{is}, \tag{14}$$

where μ_{is} is a constant for scenario s and u_{is} is a random variable with mean zero and variance V_{is}. We assume that u_{is} is uncorrelated with u_{js} for $i \neq j$.

The expected return of the portfolio is

$$E_P = \sum_{i=1}^{n} x_i \mu_i$$

with

$$\mu_i = \sum_{s=1}^{S} \mu_{is} P_s.$$

To perform the diagonalization, we define a set of S fictitious securities with weights $y_1, \ldots, y_S$ as follows:

$$y_s = \sum_{i=1}^{n} \left(\mu_{is} - \mu_i \right) x_i, \text{ for } s = 1, \ldots, S. \tag{15}$$

With this definition, the variance of the portfolio's return can be written (see Jacobs, Levy, and Markowitz 2005) as

$$V_P = \sum_{i=1}^{n} x_i^2 V_i + \sum_{s=1}^{S} y_s^2 P_s, \tag{16}$$

where

$$V_i = \sum_{s=1}^{S} P_s V_{is}.$$

Thus, portfolio variance can be expressed as a positively weighted sum of squares in the n original securities and S new fictitious securities, which are linearly related to the original securities by Equation 15. Again, therefore, portfolio variance can be written in terms of a diagonal covariance matrix.

Apart from notation (e.g., using S for K and Equation 15 for Equation 12), the scenario model is formally the same as the factor model. That is, the meanings of the coefficients are different, but with a change of notation, the portfolio selection problem with a scenario model of covariance is the same as that for a factor model of covariance.[11]

Historical Covariance Models. Consider the case in which T historical periods (e.g., months or days) are used to estimate covariances between n securities (see Markowitz, Todd, Xu, and Yamane 1992). Define a fictitious security

$$y_t = \sum_{i=1}^{n} \left(r_{it} - \bar{r}_i \right) x_i, \text{ for } t = 1, \ldots, T, \tag{17}$$

where r_{it} is the return on the ith security during period t and $\bar{r}_i$ is the ith security's historical average return:

$$\bar{r}_i = \frac{1}{T} \sum_{t=1}^{T} r_{it}.$$

Thus, y_t is the difference between portfolio return in the tth period and the portfolio's average return.

The historical variance of the portfolio is a constant times

$$V_P = \sum_{t=1}^{T} y_t^2. \tag{18}$$

It is a sum of squares in new, fictitious securities that are linearly related to the old securities by Equation 17.

We call the factor models, scenario models, and historical models described in this section "diagonalizable models." In each case, transformation into the diagonalized form allows one to write portfolio variance in terms of simple sums (i.e., Equations 13, 16, and 18) rather than in terms of nested double sums, such as in Equation 2. Diagonalization transforms the variance expressions from ones couched in terms of dense covariance matrices (i.e., matrices containing mostly nonzero entries) to expressions containing matrices that are slightly larger but have nonzero entries only along their diagonals. Inversion of such matrices is trivial.

Modeling Long–Short Portfolios

For this discussion, we adopt the convention described earlier of representing an n-security long–short portfolio in terms of $2n$ nonnegative variables $x_1, \ldots, x_{2n}$. Let r_i be the return on security i for $i = 1, \ldots, n$ and r_c be the return on cash or collateral. The portfolio's return, R_P, is then

$$R_P = \sum_{i=1}^{n} r_i x_i + \sum_{i=n+1}^{2n} \left(-r_{i-n} \right) x_i + r_c \sum_{i=n+1}^{2n} h_{i-n} x_i. \tag{19}$$

The first term on the right of Equation 19 represents the return contribution of the securities held long.

The second term represents the contribution of the securities sold short. The third term represents the short rebate, where

$$h_i \le 1, \text{ for } i = 1, \ldots, n$$

is the investor's portion of the interest received on the proceeds from the short sale of security i. Usually, $h_i \ge 0$, but this condition is sometimes violated for stocks that are hard to borrow and is not required for our results. Also, usually, $h_i < 1$, but the case of $h_i = 1$ is conceivable (and is also covered by our results).[12]

If no riskless security is ever sold short, we may modify the second term of Equation 19 to sum only over the risky securities or we may leave it as is and impose constraints to force short sales of riskless securities to be zero.[13]

For the short positions, let the return be

$$r_i = h_{i-n} r_c - r_{i-n}, \text{ for } i = n+1, \ldots, 2n. \tag{20}$$

Let μ_i be the expected value of r_i for $i = 1, \ldots, 2n$. Then, the expected return of the long–short portfolio is

$$E_P = \sum_{i=1}^{2n} x_i \mu_i. \tag{21}$$

To diagonalize, we assume a multifactor model with returns given by Equation 11 and we define K new fictitious securities, $y_1, \ldots, y_K$, in terms of the real securities, as follows:

$$y_k = \sum_{j=1}^{n} x_j \beta_{jk} - \sum_{j=1}^{n} x_{n+j} \beta_{jk}, \text{ for } k = 1, \ldots, K.$$

From this definition, it follows (see Jacobs, Levy, and Markowitz 2005) that the variance of the portfolio's return is

$$V_P = \sum_{i=1}^{2n} x_i^2 V_i + \sum_{k=1}^{K} y_k^2 W_k - 2\sum_{i=1}^{n} x_i x_{n+i} V_i. \tag{22}$$

Equation 22 is the expression for the variance of the return of a long–short portfolio when a multifactor covariance model is assumed. Note that, with the exception of the cross-product terms, Equation 22 has exactly the same form as Equation 13, which applied exclusively to long-only portfolios. The next section demonstrates how the similarity between these two expressions can be exploited to derive fast algorithms for long–short portfolios.

Applying Fast Techniques to the Long–Short Model

In this section, we consider applying existing fast (long-only) portfolio optimizers to the long–short portfolio selection problem. In particular, we consider the conditions under which a portfolio optimizer that ignores the cross-product terms in Equation 22 will still produce the correct efficient frontier. If one obtains the correct long–short efficient frontier even when ignoring those terms, then existing fast long-only portfolio optimizers can be used for long–short portfolios; the only change necessary will be the addition of n new variables to represent the short positions.

For this analysis, it is useful to define a "trim" portfolio as a long–short portfolio that has no simultaneous long and short positions in the same security. That is, a trim portfolio has

$$x_i x_{n+i} = 0, \text{ for } i = 1, \ldots, n,$$

because either x_i or x_{n+i} is zero, or both are zero. Trim portfolios have the useful property that, for them, Equation 22 has precisely the same form as Equation 13.

It is also useful to define the following modified variance:

$$V_P' = \sum_{i=1}^{2n} x_i^2 V_i + \sum_{k=1}^{K} y_k^2 W_k. \tag{23}$$

For trim portfolios, the modified variance, V_P', in Equation 23 is precisely equal to the original variance, V_P, of Equation 22. We will refer to the portfolio selection model with appropriate constraints and (E_P, V_P) given by Equations 21 and 22 as the *original* model. We will refer to the portfolio selection model that is the same except that V_P from Equation 22 is replaced with V_P' from Equation 23 as the *modified* model.

An important case in which an efficient set of portfolios for the modified model is always an efficient set for the original model is the diagonalized historical model. Equation 18 for the historical model is analogous to Equations 13 and 16 for the factor and scenario models. However, the equation for the historical model, unlike those for the other models, contains only a term involving the fictitious securities, no term involving V_i (see Jacobs, Levy, and Markowitz 2005). Therefore, $V_P' \equiv V_P$ Here, we are making no assumption about the constraint set or expected returns other than the background assumptions that the model is feasible (i.e., meets the specified constraints) and has efficient portfolios.

Trimability. For the factor and scenario models described earlier, an efficient set of portfolios for the modified model is not always an efficient set for the original model. For the factor and scenario models, a further assumption is needed for such an identity to hold. In particular, what is needed is the ability to transform a feasible portfolio that is not trim into a feasible portfolio that is trim in such a way that the transformation does not reduce the portfolio's expected return. We call this ability the "trimability condition."[14] A portfolio selection model that satisfies it is called "trimable."

In other words, for a guarantee that an efficient set for the modified model is an efficient set for the original model, we must be able to do the following:

- remove the overlap from simultaneous long and short positions in each security in such a way that the smaller of the two positions diminishes to zero,
- add the overlap to a risk-free security holding,
- leave all other risky security holdings unchanged,
- maintain feasibility, and
- not reduce the expected return of the portfolio.

If we can remove all simultaneous long and short positions in the same securities in this way, the resulting portfolio is trim.[15]

Although models with arbitrary constraint sets may not satisfy the trimability condition, a wide variety of constraints met in practice do satisfy it. Suppose, for example, that the choice of a long–short portfolio is subject to the following:

- the nonnegativity requirement (Equation 8);
- the budget constraint,

  $$\sum_{i=1}^{n} x_i + x_c - x_b \le 1, \tag{24}$$

  where x_c is a cash balance and x_b is an amount borrowed;[16] and
- any or all of the following constraints:
  A. a Reg T type of constraint as in Equation 9, perhaps with $H > 2$ for an investor not subject to Reg T,
  B. upper bounds on individual long or short positions, and/or
  C. the requirement that the total value of the long positions be related to the total value of the short positions, as in constraint Equation 10.

If a portfolio holding simultaneous long and short positions in the same security meets any or all of the above constraints, then a trimmed version of the portfolio also meets the constraints. Also, the expected return of the trimmed portfolio is greater than or equal to E_P.

Thus, on the one hand, a constraint set consisting of the nonnegativity constraint, budget constraint, and any or all of A, B, and C does satisfy the trimability condition. Note that the trimability condition requires only that the trimmed portfolio be *feasible*, not that it necessarily be efficient; thus, the investor need not be concerned, in checking the trimability condition, that, say, the trimmed portfolio might be improved by reducing the amount borrowed rather than increasing the cash balance in case $x_b > 0$.

On the other hand, for an example of a constraint set that does *not* satisfy the trimability condition, consider an upper bound on the holding of cash:

$$x_c \le U_c. \tag{25}$$

If there are no upper bounds on the other x_i, then the portfolio composed of the maximum amount of cash (i.e., $x_c = U_c$) plus overlapping long and short positions in any one security (say, $x_1 = x_{n+1} = 1 - U_c$) and no holding of any other security is feasible. However, x_1 and x_{n+1} cannot be reduced because the cash variable cannot be adjusted in the manner required by the trimability condition without violating constraint Equation 25.

Consequences of Trimability. If the trimability condition holds in the original model, then each efficient (E_P,V_P) combination has one and only one trim portfolio with the same (E_P,V_P), although there may be efficient portfolios with this (E_P,V_P) that are not trim (see Jacobs, Levy, and Markowitz 2005). Also, if the trimability condition holds in the original model, then the modified model has the same set of efficient (E_P,V_P) combinations as does the original model, and it has a unique set of efficient portfolios [one for each efficient (E_P,V_P) combination] that is the same as the unique set of trim efficient portfolios in the original model.

These facts assure us that for factor and scenario models, when the trimability condition holds, we can naively use a portfolio optimizer that assumes variance is given by Equation 23 (thereby knowingly ignoring the negative correlation between u_i and u_{n+i}) and still get the correct long–short efficient frontier.

When the trimability condition does not hold, we may not get the correct efficient frontier if we ignore the cross-product terms. For example, consider a diagonalized model with a Reg T constraint (with $H = 2$), a budget constraint as in Equation 24, and an upper bound on cash ($U_c < 1.0$). In the original model, the portfolio with

$$x_i = \begin{cases} 1 & \text{if } i = 1 \\ 1 & \text{if } i = n+1 \\ 0 & \text{otherwise} \end{cases}$$

is feasible and has zero variance. Therefore, zero variance is feasible and some portfolio (not necessarily the above portfolio) has zero variance and is efficient. But the modified version of this model has no feasible zero-variance portfolios: The upper bound on cash implies that the portfolio must hold a positive amount of some risky security, which implies that the portfolio variance is greater than zero. Thus, in the absence of some assumption such as the trimability condition, an efficient set for the modified model may not be an efficient set for the original model.

Example

In this section, we provide an example of a three-security, one-factor, long–short model subject only to Reg T, the budget constraint, and nonnegativity constraints. In this case, Equation 11 may be written as

$$r_i = \alpha_i + \beta_i f + u_i, \text{ for } i = 1, 2, 3.$$

The tables illustrate the example. In all tables, the long positions in the three securities are labeled 1L, 2L, and 3L and short positions are labeled 1S, 2S, and 3S.

Table 1 presents inputs for the long positions in the three securities. It also shows the lending rate and the borrowing rate. The variance of underlying factor *f*, *V(f)*, is 0.04.

Table 1. Illustrative Three-Security One-Factor Model

Security, i	Expected Return, μ_i	Beta, β_i	Idiosyncratic Variance, V_i	Rebate Fraction, h_i
1L	0.10	0.80	0.0768	0.5
2L	0.12	1.00	0.1200	0.5
3L	0.16	1.25	0.1875	0.5
Lending rate	0.03	0.00	0.0000	na
Borrowing rate	0.05	0.00	0.0000	na

na = not applicable.

The betas of the securities, their idiosyncratic variances, and the variance of the underlying factor can be used to compute the covariances between the long positions according to the formulas

$$\text{cov}\left(r_i, r_j\right) = \beta_i \beta_j V(f), \text{ for } i \neq j$$

and

$$V\left(r_i\right) = \beta_i^2 V(f) + V_i, \text{ for } i = 1, 2, 3.$$

The result of this calculation for the present example is shown in **Table 2**.

Table 2. Covariances between Long Positions

	1L	2L	3L
1L	0.1024	0.0320	0.0400
2L	0.0320	0.1600	0.0500
3L	0.0400	0.0500	0.2500

For a long-only portfolio analysis, the covariance matrix for a one-factor model can be transformed into a sum of squares by introducing a new variable, the portfolio beta (PB), as in Equation 12. **Table 3** shows the covariance matrix for this four-security version of the three-security single-factor model. The covariance matrix is now diagonal, with nonzero entries on the diagonal, rather than the dense covariance matrix such as that shown in Table 2.

Table 3. Covariances When Dummy Security Is Included

	1L	2L	3L	PB
1L	0.0768	0.0000	0.0000	0.0000
2L	0.0000	0.1200	0.0000	0.0000
3L	0.0000	0.0000	0.1875	0.0000
PB	0.0000	0.0000	0.0000	0.0400

The advantage of diagonalizing the covariance matrix increases with the number of securities in the portfolio analysis. To illustrate, the second column of **Table 4** shows the number of input coefficients required by the diagonal model of covariance—namely, n betas, n idiosyncratic variances, and one factor f variance. The third column of Table 4 shows the number of unique covariances needed by a computation using an arbitrary covariance matrix—namely, $n(n + 1)/2$. Specifically, with three securities, there are actually more coefficients in the diagonal model than in the nondiagonalized version. With 5,000 securities, the diagonal model works with about 10,000 coefficients whereas the 5,000 × 5,000 covariance matrix

Table 4. Number of Unique Coefficients Required by Model of Covariance

Number of Securities	With Dummy Security	Without Dummy Security
3	7	6
20	41	210
100	201	5,050
500	1,001	125,250
1,000	2,001	500,500
3,000	6,001	4,501,500
5,000	10,001	12,502,500

of the general model has more than 12 million unique covariances (counting $\sigma_{ij} = \sigma_{ji}$ as one covariance). Both versions of the model also need n expected returns.

The two versions will perform the same number of iterations and arrive at the same efficient frontier. The work per iteration depends on how many securities are in the portfolio and on the total number of securities considered for inclusion. For moderate to large portfolios, much less work is required by the diagonal model per iteration. If the portfolio contains n_I securities, the Sharpe algorithm requires a few more than $3n + 7n_I$ multiplications and divisions plus $3n + 5n_I$ additions, whereas the general algorithm requires $2n_I n + 5n + 2n_I^2 - n_I$ multiplications and divisions plus $2n_I n + 3n + 2n_I^2 - 2n_I$ additions. Thus, if $n = 1{,}000$ and $n_I = 10$ (as at the high-return/high-variance end of the frontier) or $n_I = 100$ (as might occur at the low-return/low-variance end of the frontier), the diagonal model requires, respectively, 3,070 or 3,700 multiplications and divisions for the iteration whereas the general algorithm requires 25,190 or 269,900.

Table 5 provides the expected returns, betas, and idiosyncratic variances for the long and short securities corresponding to the long securities in Table 1. The expected returns for the short positions were computed according to Equation 20. The betas of the short positions are the negatives of those for the corresponding long positions, whereas the idiosyncratic variances are the same for the short positions as for the corresponding long positions.

Table 5. Illustrative Three-Security One-Factor Model with Long and Short Positions

Security, i	Expected Return, μ_i	Beta, β_i	Idiosyncratic Variance, V_i
1L	0.100	0.80	0.0768
2L	0.120	1.00	0.1200
3L	0.160	1.25	0.1875
1S	–0.085	–0.80	0.0768
2S	–0.105	–1.00	0.1200
3S	–0.145	–1.25	0.1875
Lending rate	0.030	0.00	0.0000
Borrowing rate	0.050	0.00	0.0000

The covariances between long and short positions, presented in **Table 6**, were derived by replicating Table 2 in the manner necessitated by the extension of the portfolio to include both long and short positions. We could compute an efficient frontier for the long–short model from the expected returns in Table 5 and the covariance matrix in Table 6 by using a general portfolio analysis program that permits an arbitrary covariance matrix, but using Sharpe's technique is much more efficient.

Performing the Sharpe technique of expressing return as a linear function of the amount invested in the factor plus the amounts invested in the idiosyncratic terms produces the covariance for the long–short model presented in **Table 7**. Note that the covariance matrix is no longer diagonalized because, for example, the 1L idiosyncratic term has a –1.0 correlation with 1S.

Table 6. Covariances between Long and Short Positions

	1L	2L	3L	1S	2S	3S
1L	0.1024	0.0320	0.0400	–0.1024	–0.0320	–0.0400
2L	0.0320	0.1600	0.0500	–0.0320	–0.1600	–0.0500
3L	0.0400	0.0500	0.2500	–0.0400	–0.0500	–0.2500
1S	–0.1024	–0.0320	–0.0400	0.1024	0.0320	0.0400
2S	–0.0320	–0.1600	–0.0500	0.0320	0.1600	0.0500
3S	–0.0400	–0.0500	–0.2500	0.0400	0.0500	0.2500

Table 7. Covariances When Dummy Security Included

	1L	2L	3L	1S	2S	3S	PB
1L	0.0768	0.0000	0.0000	–0.0768	0.0000	0.0000	0.0000
2L	0.0000	0.1200	0.0000	0.0000	–0.1200	0.0000	0.0000
3L	0.0000	0.0000	0.1875	0.0000	0.0000	–0.1875	0.0000
1S	–0.0768	0.0000	0.0000	0.0768	0.0000	0.0000	0.0000
2S	0.0000	–0.1200	0.0000	0.0000	0.1200	0.0000	0.0000
3S	0.0000	0.0000	–0.1875	0.0000	0.0000	0.1875	0.0000
PB	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0400

If the data in Table 5 are presented to an algorithm that implements Sharpe's technique, the algorithm will operate as though the covariance matrix were, in fact, diagonal—as shown in **Table 8**. Satisfaction of the trimability condition assures us that the efficient frontier computed by using the diagonal covariance matrix in Table 8 is the same as the efficient frontier computed by using the correct covariance matrix in Table 7 and that for any number of securities, we will get the correct result, even if we ignore the correlations between the idiosyncratic terms for a many-factor model, a scenario model, or a mixed factor-and-scenario model. It further assures us that the efficient frontier is correctly computed, even if additional constraints are imposed on the choice of portfolio—provided that the constraint set satisfies the trimability condition.

In the case of the n-security one-factor model with long and short positions, the advantage of using the diagonal model rather than a general model is again given by Table 4. In this case, however, an n-security long–short model has $2n$ securities. If there are 500 securities in the universe, the diagonal model will be told that there are 1,001 securities whose covariance structure is described by 2,001 coefficients. In contrast, the general model would require 500,500 unique covariances.

Summary

Realistic models of long–short portfolio restrictions can be written as systems of linear equality or inequality constraints. Examples of such constraints include budget constraints, the Reg T margin requirement, upper bounds on long or short positions in individual or groups of assets, and the requirement that the difference between the sum of long positions and the sum of short positions be close to an investor-chosen value. Market-neutral equity strategies correspond to a chosen value of 0, and enhanced active equity strategies correspond to a chosen value of 1.

The speed of portfolio optimization can be increased significantly by taking advantage of models (including factor and scenario models) that define new fictitious securities that are linearly related to the real securities in such a way that the covariance matrix of the securities' returns becomes diagonal or almost so. Existing fast algorithms take advantage of the resultant sparse, well-structured sets of equations to increase dramatically the speed of portfolio optimization.

We discussed the conditions under which such fast algorithms, designed for long-only portfolios, will produce correct long–short portfolios. In general, even if long-only positions in n securities satisfy the assumptions of a factor or scenario model, a $2n$-variable long–short model does not satisfy these same assumptions. In particular, the idiosyncratic terms of the long–short model are not uncorrelated.

Despite this violation of the assumption of no correlation of the idiosyncratic terms, a portfolio optimization program that uses factor or scenario models will compute the correct long–short portfolio as long as the trimability condition holds. The trimability condition requires that if a portfolio with short and long positions in the same stock is feasible, then it is also feasible to reduce both positions while keeping the holdings of all other risky securities the same and not reducing the expected return of the portfolio.

The acceleration in computation that results from the use of diagonalized versions of factor, scenario, or historical models is approximately equal to the ratio of nonzero coefficients in the equations of the two models. For large problems, this time saving can be considerable.

This article qualifies for 1 PD credit.

Table 8. Covariances Based on Consequences of the Trimability Condition

	1L	2L	3L	1S	2S	3S	PB
1L	0.0768	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2L	0.0000	0.1200	0.0000	0.0000	0.0000	0.0000	0.0000
3L	0.0000	0.0000	0.1875	0.0000	0.0000	0.0000	0.0000
1S	0.0000	0.0000	0.0000	0.0768	0.0000	0.0000	0.0000
2S	0.0000	0.0000	0.0000	0.0000	0.1200	0.0000	0.0000
3S	0.0000	0.0000	0.0000	0.0000	0.0000	0.1875	0.0000
PB	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0400

Notes

1. For detailed derivations and mathematical proofs of the results presented in this article, see Jacobs, Levy, and Markowitz (2005).
2. If the constraint set includes inequalities, they can be converted into equalities by using nonnegative "slack variables." Slack variables can be interpreted as zero-variance (riskless) securities.
3. See Jacobs and Levy (2000, 2005). To sell short for any customer, a broker must borrow the stock to be sold. The brokerage firm may borrow the stock from itself, typically from customer stock held in a street name in margin accounts, or the broker may borrow the stock from another investor, typically a large institutional investor. Some intermediary may facilitate the process of bringing together demand and supply of "stock to lend." Sometimes, a lender cannot be found for the desired stock. In this case, the stock cannot be sold short. Furthermore, the lender retains the right to call back the stock; if the lender does and another lender is not readily available, the investor must cover the short position and deliver the stock to the lender. The proceeds of a short sale are posted as collateral to the lender of the stock. In fact, if the stock is borrowed from another investor, the broker must typically put up about 105 percent of the proceeds. The broker and the stock lender retain a portion of the interest earned on these proceeds. A large institutional investor that shorts stock usually receives a portion of the interest (referred to as a "short rebate"). A small retail customer who sells short typically receives no part of the interest.
4. Because the second n variables represent short sales in the same set of securities, if security i is held long, there will be a positive entry in x_i and if security i is sold short, there will be a positive entry in x_{n+i}.
5. Constraint Equation 9 with $H = 2$ is referred to as a 50 percent margin requirement on both short and long positions. In practice, the nature of this margin requirement is different for short and long positions. In the case of a long position, the customer may borrow as much as 50 percent of the value of the position from the broker. In the case of a short position, the customer does not borrow money from the broker; the margin requirement is a collateral requirement. Furthermore, the Reg T requirements are for initial margin—the equity required in the account to establish initial positions. It does not constrain the value of the positions maintained after they have been established. The security exchanges and brokers, however, do impose maintenance margin requirements. Consequently, one motive of the investor in setting her or his own H in constraint Equation 9 is to reduce the probability of needing additional cash for maintenance margin [see Regulation T, "Credit by Brokers and Dealers" (12 CFR 220), available online at www.federalreserve.gov/regulations]. Jacobs and Levy (1993) discussed margin requirements and cash needed for liquidity. The Reg T initial short margin requirement is stated as 150 percent—of which 100 percent is to be supplied by the proceeds of the sale of the borrowed stock.
6. Reg T can be circumvented in several ways. For example, hedge funds often set up off-shore accounts, which are not subject to Reg T. Alternatively, a large hedge fund can register as a broker/dealer, with a real broker/dealer acting as the back office. As a broker/dealer, the hedge fund is subject to broker/dealer capital requirements rather than Reg T requirements. Broker/dealer capital requirements allow much more leverage than Reg T. In the extreme, the only constraint is what the broker imposes on the hedge fund's portfolio to assure that, in the case of unfavorable market movements, the broker is secure. A hedge fund can also circumvent Reg T by having a broker set up a proprietary trading account of its own, which is managed by the fund. Gains and losses in the proprietary trading account are transferred to the hedge fund via prearranged swap contracts. The only constraint imposed by this arrangement is the broker's own capital requirements plus whatever constraints the broker imposes. Rule 15c3-1 of the Securities Exchange Act of 1934 governing capital requirements for broker/dealers includes the provision that indebtedness cannot exceed 1500 percent of net capital (800 percent for 12 months after commencing business as a broker or dealer). Also lying outside Reg T are certain arrangements that allow the investor to use noncash collateral, including existing long positions, to collateralize the shares borrowed to sell short, which frees up the proceeds from short sales to be used for further purchases and short sales. (In these cases, however, as in the exceptions to Reg T noted previously, the broker/dealer imposes its own constraints on leverage to assure its own security.) Noncash collateral may consist of securities or letters of credit and usually amounts to 100–105 percent of the amount borrowed. Noncash collateral is marked to market, together with the shares borrowed, and the borrower must make good on any shortfall between the value of the noncash collateral and the value of the shares borrowed. Gains and losses on the collateral accrue to the borrower; the lender is generally paid a fee for the use of the securities.
7. Jacobs, Levy, and Starer (1998, 1999) addressed the conditions under which optimal portfolios that are constrained to hold roughly equal amounts in long and short positions are equivalent to optimal portfolios without this constraint.
8. The problem requires the inversion of a bordered covariance matrix (i.e., a covariance matrix onto which is added bordering coefficients that serve to implement the constraints on the portfolio). See, for example, Markowitz (1987), Markowitz and Todd (2000), and Perold (1984).
9. The mathematical details of the more general case, in which the factors are not necessarily mutually uncorrelated, are discussed in Jacobs, Levy, and Markowitz (2005).
10. Sharpe's diagonalized version of the n-security one-factor model is frequently considered to be *the* diagonal model.
11. For models that combine both scenarios and factors, see Markowitz and Perold (1981a, 1981b).
12. Large institutional investors often perform mean–variance analysis at an asset-class level and then implement the asset-class allocations by using index funds or internal or external fund managers. For instance, if an internal market-neutral fund borrows shares from, say, an internal large-capitalization or small-capitalization fund, the allocation of interest on the proceeds between the borrowing fund and the lending fund is arbitrary. The institution's policy might allocate all the interest to the borrowing fund because the institution's policy might prohibit external stock lending, so the particular interest income would not exist except for the internal market-neutral fund's activities.
13. Equation 19 does not include tax considerations and, therefore, is applicable to tax-exempt organizations, such as university endowments and corporate pension plans.
14. This condition is called "Property P" in Jacobs, Levy, and Markowitz (2005).
15. Another effect of moving the overlap to a riskless security is that it increases the amount of slack available to the investor in complying with Reg T.
16. The sum is taken only over long securities. Recall that, unlike CAPM investors who have Equation 5 as their only constraint, investors are typically constrained by Reg T margin requirements and do not get to spend the proceeds from selling short, although they may share the interest collected on these proceeds.

References

Cohen, K.J., and J.A. Pogue. 1967. "An Empirical Evaluation of Alternative Portfolio Selection Models." *Journal of Business*, vol. 40, no. 2 (April):166–193.

Jacobs, Bruce I., and Kenneth N. Levy. 1993. "The Generality of Long–Short Equitized Strategies: A Correction" (letter in response to C.B. Garcia and F.G. Gould, "The Generality of Long–Short Equitized Strategies," *Financial Analysts Journal*, September/October 1992). *Financial Analysts Journal*, vol. 49, no. 2 (March/April):22.

———. 2000. *Equity Management: Quantitative Analysis for Stock Selection*. New York: McGraw-Hill.

———. 2005. *Market-Neutral Strategies*. Hoboken, NJ: John Wiley.

———. Forthcoming 2006. "Enhanced Active Equity Strategies: Relaxing the Long-Only Constraint in the Pursuit of Active Return." *Journal of Portfolio Management*.

Jacobs, Bruce I., Kenneth N. Levy, and Harry M. Markowitz. 2005. "Portfolio Optimization with Factors, Scenarios, and Realistic Short Positions." *Operations Research*, vol. 53, no. 4 (July/August):586–599.

Jacobs, Bruce I., Kenneth N. Levy, and David Starer. 1998. "On the Optimality of Long–Short Strategies." *Financial Analysts Journal*, vol. 54, no. 2 (March/April):40–51.

———. 1999. "Long–Short Portfolio Management: An Integrated Approach." *Journal of Portfolio Management*, vol. 25, no. 2 (Winter):23–32.

Markowitz, Harry M. 1959. *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons; 2nd ed., 1991. Cambridge, MA: Basil Blackwell.

———. 1987. *Mean–Variance Analysis in Portfolio Choice and Capital Markets.* Cambridge, MA: Basil Blackwell.

Markowitz, Harry M., and André F. Perold. 1981a. "Portfolio Analysis with Factors and Scenarios." *Journal of Finance*, vol. 36, no. 4 (September):871–877.

———. 1981b. "Sparsity and Piecewise Linearity in Large Portfolio Optimization Problems." In *Sparse Matrices and Their Uses.* Edited by I.S. Duff. London: Academic Press.

Markowitz, Harry M., and Peter Todd. 2000. *Mean–Variance Analysis in Portfolio Choice and Capital Markets* (revised reissue of Markowitz 1987 with chapter by Peter Todd). New Hope, PA: Frank J. Fabozzi Associates.

Markowitz, Harry M., Peter Todd, Gan Lin Xu, and Yuji Yamane. 1992. "Fast Computation of Mean–Variance Efficient Sets Using Historical Covariances." *Journal of Financial Engineering*, vol. 1, no. 2 (September):117–132.

Perold, André F. 1984. "Large-Scale Portfolio Optimization." *Management Science*, vol. 30, no. 10 (October):1143–60.

Sharpe, William F. 1963. "A Simplified Model for Portfolio Analysis." *Management Science*, vol. 9, no. 2 (January):277–293.

