1. Portfolio Optimization Problems

Portfolio optimization problems or asset allocation problems look at the “best” way for an investor or fund manager to allocate funds between a number of different asset classes. One criterion is to maximize the utility of the portfolio of assets at the end of the investment period. The classical mean variance model, which involves maximizing the portfolio return and minimizing the risk, was proposed by Markowitz [21] in 1952. There has been continual interest in the problem since (see [22, 28, 19, 23, 8, 12] for example), with increasingly sophisticated mathematical models being proposed (see [26, 7, 20, 25] for recent examples). In the most basic setting the planning horizon is just a single period, and transaction costs are ignored. This allows some of the basic ideas to be discussed but limits the applicability of the models.

Let the asset classes be indexed \( \{1, \ldots, n\} \) and let \( x_i \) be the fraction of the portfolio allocated to asset class \( i \). The vector of weights \( x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n \) then satisfies

\[ e^T x = 1 \quad x \geq 0. \]

Here, and subsequently, \( e = (1, \ldots, 1)^T \) is the vector of all 1s of the appropriate dimension. More generally there may be simple lower and upper bounds

\[ \ell_i \leq x_i \leq u_i \quad i = 1, \ldots, n \tag{1.1} \]

on the weights (e.g. the fraction of the portfolio in property must be between 5% and 20%), and general linear constraints

\[ \ell_{n+i} \leq a_i^T x \leq u_{n+i} \quad i = 1, \ldots, m. \]

(e.g. the total fraction of the portfolio allocated to all international assets must not exceed 40%). Writing the constant vectors \( a_i \in \mathbb{R}^n \) as the rows of the matrix \( A \in \mathbb{R}^{m \times n} \) defined by \( A^T = [a_1 \cdots a_m] \) the linear constraints are

\[ \ell \leq \begin{bmatrix} x \\ Ax \end{bmatrix} \leq u, \tag{1.2} \]

where \( \ell \leq u \in \mathbb{R}^{n+m} \). Any equality constraints have \( \ell_i = u_i \). The linear constraints (1.2) define a convex feasible region in \( \mathbb{R}^n \).

2. Mean Variance Models

Given a vector \( r \in \mathbb{R}^n \) of expected returns for the asset classes, the expected return for the portfolio is \( \mu = r^T x \), which is to be maximized. Given a symmetric
positive semi-definite covariance matrix $C \in \mathcal{R}^{n \times n}$, a measure of the risk associated with the portfolio is the portfolio variance $\sigma^2(x) = x^T C x$. The two objectives of maximizing the return and minimizing the risk can be viewed either as a multi-objective optimization problem, or the objectives can be combined using a utility function. A typical problem is thus to maximize the utility function

$$ f(x) = \text{Portfolio return} - \frac{\text{Portfolio variance}}{t} = r^T x - \frac{x^T C x}{t} $$

where $t > 0$ is a risk tolerance parameter. Combining the objective (2.1) and the constraints (1.2) produces the convex quadratic programming problem

$$ \text{Minimize } -r^T x + \frac{x^T C x}{t} \\
x \in \mathcal{R}^n \\
Subject to \ell \leq \begin{bmatrix} x \\ A x \end{bmatrix} \leq u. $$

(2.2)

Alternatively the variance of the portfolio can be minimized for a specified portfolio return $\mu$. The minimal set of constraints then produces the problem

$$ \text{Minimize } x^T C x \\
x \in \mathcal{R}^n \\
Subject to r^T x = \mu \\
e^T x = 1 \\
x \geq 0. $$

(2.3)

Frequently the parametric optimization problem (2.3) is solved for a range of values for $\mu$, giving an efficient frontier for that investor.

Both (2.2) and (2.3) are convex problems, so any local solution is a global solution. Moreover if the covariance matrix $C$ is positive definite then the objective is strictly convex and there is a unique local and global solution. For typical values of $n$ around 10 – 20 problems (2.2) and (2.3) can be solved by a wide variety of software packages on a PC, often in conjunction with a major spreadsheet package. If the covariance matrix is singular then the convex set of optimal solutions may contain more than one element, and the problem may be numerically more difficult to solve.

There are many recognised difficulties with the use of (2.2) or (2.3). These include: getting the problem data $r \in \mathcal{R}^n$ and $C \in \mathcal{R}^{n \times n}$ and what effect errors in this data will have [6]; the use of the symmetric measure $x^T C x$ of risk when downside risk is more relevant [9, 11, 14, 16]; optimization solutions with a large number of active constraints (corner solutions); transaction costs (selling property is much more difficult than selling shares); other utility functions [17]; and the use of a single planning period [7, 25].

Addressing these issues often produces more sophisticated optimization problems. Although many of these can readily be solved using modern numerical optimization methods, some are intrinsically more difficult. This paper considers the computational solution of two optimization problems arising from models used to address the
downside nature of risk. Firstly the use of scenarios to model the stochastic nature of the problem is discussed.

3. Scenarios

Given a set of assets \{1, \ldots, n\}, the return on these assets during the planning period must be estimated. One approach is to represent the stochastic nature of the problem using scenarios. The investor tries to predict the performance of each asset in each of a set of representative scenarios. Let the number of scenarios be \(s\) and let the return on asset \(i\) under scenario \(k\) be \(R_{ki}\) for \(k = 1, \ldots, s\) and \(i = 1, \ldots, n\). This gives a scenario return matrix \(R \in \mathbb{R}^{s \times n}\). Additionally the probability of each scenario occurring is required. Let \(p \in \mathbb{R}^s\) be the scenario probability vector. Thus \(p \geq 0\) and \(e^T p = 1\), where \(p_k\) is the probability of scenario \(k\) for \(k = 1, \ldots, s\).

The expected asset returns are \(\mu = R^T p \in \mathbb{R}^n\), the scenario returns for a portfolio with weights \(x\) are \(y = Rx \in \mathbb{R}^s\), and the expected portfolio return is \(\mu = p^T y = r^T x = p^T Rx\). Define \(P, D \in \mathbb{R}^{s \times s}\) and \(S \in \mathbb{R}^{s \times n}\) by

\[
P = pe^T, \quad D = \text{diag}(p) \quad \text{and} \quad S = (I - P^T)R,
\]

(3.1)

where \(e = (1, \ldots, 1)^T \in \mathbb{R}^s\). For any vector \(y\) let \(y^j\) denote the vector with each component raised to the power \(j\). The difference between the scenario returns and the expected portfolio return is

\[
z = y - \mu e = Rx - (p^T Rx)e = (I - P^T)Rx = Sx.
\]

Hence the variance \(\sigma^2(x)\) satisfies

\[
\sigma^2(x) = \mathbb{E}[(y - \mu e)^2] = p^T S^2 x = x^T S^T DSx,
\]

(3.2)

\[
\nabla \sigma^2(x) = 2S^T DSx,
\]

(3.3)

\[
\nabla^2 \sigma^2(x) = 2S^T DS,
\]

(3.4)

so the covariance matrix is \(C = S^T DS\).

4. A Semi-variance Model

The variance model measures risk by the amount the scenario returns differ from the expected portfolio return. An investor is not concerned if their return is higher than expected, only if their return is lower than expected. The downside-risk approach to investment decisions [11] concentrates on the variability of returns below a specified target or benchmark. One measure of this is the semi-variance defined by

\[
\varphi(x) = \mathbb{E}[\min(0, y - \mu)] = p^T (\min(0, Sx)^2).
\]

(4.1)

The semi-variance is once continuously differentiable, but not necessarily twice continuously differentiable. It is a convex piecewise quadratic function, where each quadratic piece has a positive semi-definite Hessian \(G(x)\). The semi-variance is an example of a function of order 1 smoothness, or an \(LC^1\) function. Rapidly convergent methods for \(LC^1\) functions have been proposed recently in [27].
Let $\hat{D}(x) \in \mathbb{R}^{s \times s}$ be the diagonal matrix defined by

$$
\hat{D}_{kk}(x) = \begin{cases} 
0 & \text{if } e^T_k S x > 0 \\
0 \text{ or } p_k & \text{if } e^T_k S x = 0 \\
p_k & \text{if } e^T_k S x < 0 
\end{cases} \quad k = 1, \ldots, s.
$$

(4.2)

Then

$$
\varphi(x) = x^T S^T \hat{D}(x) S x,
$$

(4.3)

$$
\nabla \varphi(x) = 2 S^T \hat{D}(x) S x,
$$

(4.4)

$$
G(x) = 2 S^T \hat{D}(x) S.
$$

(4.5)

$G(x)$ is an element of the generalized Hessian of $\phi(x)$ for any $\hat{D}(x)$ defined by (4.2). If $e^T_k S x \neq 0$ for all $k$ then $x$ is in the interior of a smooth quadratic piece of the semi-variance.

The semi-variance $\varphi(x)$ can be used instead of the variance $\sigma^2(x)$ in (2.3). As the semi-variance is a convex function, the resulting problem is a convex programming problem and so any local minimizer is a global minimizer. As in the mean – variance model, an efficient frontier for the semi-variance model can be calculated by varying the expected return $\mu$.

5. A Skewness Model

Other attempts to model the asymmetric nature of risk are based on maximizing the skewness

$$
\phi(x) = \frac{\kappa(x)}{\sigma^3(x)},
$$

(5.1)

where

$$
\kappa(x) = \mathcal{E}[(y - \mu e)^3] = p^T (Sx)^3.
$$

(5.2)

The skewness function $\phi(x)$ is the third moment $\kappa(x)$ of the returns scaled by the the standard deviation $\sigma(x)$ cubed. A positively skewed portfolio is more likely to have returns exceeding the expected portfolio return. Alternatively a symmetric return distribution, with equal probability of getting returns lower or higher than expected, will have zero skewness. Unlike semi-variance, the skewness function is $C^\infty$, as long as the variance is non-zero. When $C$ is positive definite any feasible allocation $x$ is non-zero as $e^T x = 1$, and hence the portfolio variance $\sigma^2(x)$ is non-zero. However when the covariance matrix is singular it is possible that the variance may be zero for a feasible asset allocation.

Remembering that $z = Sx$ is the difference between the scenario returns and the expected portfolio return, let $E(x) = \text{diag}(DSx) = \text{diag}(Dz) \in \mathbb{R}^{s \times s}$. The third moment $\kappa(x)$ has gradient $\nabla \kappa(x) = 3 S^T D(Sx)^2 = 3 S^T E(x) z$ and Hessian $\nabla^2 \kappa(x) = 6 S^T E(x) S$. The gradient and the Hessian of $\phi(x)$ are

$$
\nabla \phi(x) = \frac{3}{\sigma^3(x)} S^T E(x) z - \frac{3 \kappa(x)}{\sigma^5(x)} S^T D z
$$

(5.3)
\[ \nabla^2 \phi(x) = \frac{6}{\sigma^6(x)} S^T E(x) S + \frac{15\kappa(x)}{\sigma^4(x)} S^T Dz z^T DS \]
\[- \frac{3}{\sigma^4(x)} \left( \kappa(x) S^T DS + 3S^T D(z(z^T)^2 + z^2 z^T)DS \right). \tag{5.4} \]

The skewness model can be formulated as maximizing \( \phi(x) \) subject to the same set of constraints as in the mean variance model (2.3). The variance is not restricted as a constraint specifying the value of variance will lead to a quadratic constraint which is much harder to solve than a linearly constrained problem. As \( \phi(x) \) is the third moment \( \kappa(x) \) divided by the variance \( \sigma^2(x) \) raised to the power \( \frac{3}{2} \), maximizing the skewness tends to maximize the third moment while minimizing the variance.

An alternative skewness model has been proposed \([15, 16, 9]\) in which the absolute deviation of the return from the mean is used as a surrogate for the variance. Instead of maximizing the skewness their model maximizes the third moment, which simplifies the objective function. A quadratic constraint on the variance is replaced by a piecewise linear constraint controlling the absolute deviation of the return. Additional variables are then used to convert the absolute deviation into a number of linear constraints, which increases the size of the problem substantially.

Incorporating the skewness recognises the asymmetric nature of return and provides a better approximation than the mean variance model. However, as the skewness function is not convex, the problem may have several different local minima. If the global minimum is not found then the efficient frontier may have discontinuities (corresponding to jumps between different local minima), making it useless.

### 6. Numerical Results

These problems were solved for a data set with \( n = 9 \) and \( s = 15 \). One feature of this data set was an asset class, corresponding to a short term bank deposit, which had the same predicted return in every scenario. This meant that the covariance matrix \( C = S^T DS \) was singular and not positive definite, so the variance was not strictly convex. The results are given in Table 1 for \( \mu = 6.5\% \), Table 2 for \( \mu = 8\% \) and Table 3 for \( \mu = 8.75\% \). Problem (2.3) was solved using the variance \( \sigma^2(x) \) (MV column), the semi-variance \( \varphi(x) \) (SV column) and the skewness \( \phi(x) \) (SK column) as the objective function.

The mean-variance model (2.3) produces a convex quadratic programming problem which can be readily solved. The singularity of the covariance matrix did not cause any difficulties.

If the semi-variance \( \varphi(x) \) replaces the variance in (2.3) the the problem is a convex programming problem with a piecewise quadratic objective. A sequential quadratic programming (SQP) method with trust region algorithm can be used to efficiently solve the problem. At each iteration an element of the generalized Hessian is used in a quadratic model of the objective. This quadratic piece is minimized, subject to any linear constraints on the problem and the restriction that the new point lies in a trust region around the current iterate. Using the infinity norm to define the trust region just changes the simple bounds on the variables in the quadratic programming...
Table 1. Results for $\mu = 6.5\%$

<table>
<thead>
<tr>
<th></th>
<th>MV</th>
<th>SV</th>
<th>SK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^*$</td>
<td>0.0000</td>
<td>0.0041</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\sigma^2(x^*)$</td>
<td>0.0000</td>
<td>0.0064</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\varphi(x^*)$</td>
<td>0.4846</td>
<td>0.4610</td>
<td>0.6364</td>
</tr>
<tr>
<td>$\phi(x^*)$</td>
<td>0.1910</td>
<td>0.0511</td>
<td>0.0082</td>
</tr>
<tr>
<td></td>
<td>0.0465</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.0022</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.0095</td>
<td>0.0300</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.0130</td>
<td>0.4610</td>
<td>0.6364</td>
</tr>
<tr>
<td></td>
<td>0.0190</td>
<td>0.2090</td>
<td>0.0931</td>
</tr>
<tr>
<td></td>
<td>0.0266</td>
<td>0.2654</td>
<td>0.2623</td>
</tr>
<tr>
<td></td>
<td>0.0103</td>
<td>0.0112</td>
<td>0.0451</td>
</tr>
<tr>
<td></td>
<td>0.0024</td>
<td>0.0021</td>
<td>0.0052</td>
</tr>
<tr>
<td></td>
<td>1.6414</td>
<td>1.8280</td>
<td>3.3767</td>
</tr>
</tbody>
</table>

Table 2. Results for $\mu = 8.0\%$

<table>
<thead>
<tr>
<th></th>
<th>MV</th>
<th>SV</th>
<th>SK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^*$</td>
<td>0.0270</td>
<td>0.0490</td>
<td>0.1424</td>
</tr>
<tr>
<td>$\sigma^2(x^*)$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\varphi(x^*)$</td>
<td>0.0390</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\phi(x^*)$</td>
<td>0.0000</td>
<td>0.1080</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.1080</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.3186</td>
<td>0.2768</td>
<td>0.3725</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.1368</td>
<td>0.1516</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>0.5137</td>
<td>0.4145</td>
<td>0.4851</td>
</tr>
<tr>
<td></td>
<td>0.1560</td>
<td>0.1481</td>
<td>0.2344</td>
</tr>
<tr>
<td></td>
<td>0.0601</td>
<td>0.0542</td>
<td>0.1271</td>
</tr>
<tr>
<td></td>
<td>0.0203</td>
<td>0.9111</td>
<td>0.1927</td>
</tr>
</tbody>
</table>

The skewness function is not convex so standard methods converge to a local minimum. The three SK columns in Tables 1, 2 and 3 are local maximizers of $\phi(x)$ obtained starting from the mean variance solution, from the semi-variance solution and from another starting point. Taking the best solution from a number of different starting points provides a crude estimate of the global maximum. The different solutions found correspond to different active sets. For each value of $\mu$ in Tables 1, 2 and 3 the semi-variance at the local maximizers of the skewness model is significantly larger than the minimum value of the semi-variance.

The skewness was maximized by minimizing $-\phi(x)$ using an SQP method with a trust region algorithm. The Hessian $\nabla^2\phi(x)$ was typically indefinite with large
positive and large negative eigenvalues. For example at the first SK solution in Table 2 the eigenvalues of $\nabla^2 \phi(x^*)$ ranged between 2,750 and $-86$ (remember $x^*$ is a local maximizer of $\phi(x)$). Some quadratic programming subproblems produced a local minimizer of the quadratic model whose value was larger than the value at the current point. The resulting direction was often not a descent direction. Thus when the Hessian is indefinite either the method for solving the quadratic programming subproblem must find a global solution of the quadratic subproblem or at least a local solution with value less than the starting point (which typically corresponds to a zero step). In such cases an eigenvector corresponding to a negative eigenvalue may be used to try to find a descent direction for use in a line search. An alternative is to add a multiple of the identity matrix to the Hessian or the reduced Hessian to ensure the search direction is a descent direction.

7. Conclusions

Minimizing the semi-variance produces a convex piecewise quadratic optimization problem with linear constraints. This can be solved efficiently using a sequential quadratic programming algorithm. However models based on maximizing the skewness are both nonlinear and non-convex. This makes them much harder to solve, and methods typically only give a local rather than a global solution. Moreover it is not clear that maximizing the skewness produces a asset allocation with less risk.

As in [6, 2, 3, 4] the sensitivity of the optimal asset allocations $x^*$ with respect to perturbations in the problem data $p$ and $R$ is of interest.

References


