PORTFOLIO OPTIMIZATION

INTRODUCTION:

We revisit the investment problem of trying to maximize gain while minimizing risk that we considered at the beginning of the semester. This time we will be able to do this in a more sophisticated manner. We start with the approach due to Markowitz (for which he received the Nobel Prize in Economics in 1990).

THE MARKOWITZ MODEL:

We assume that we have an amount of money that we wish to invest for some period of time, say a year. Our model is that we invest now, and we are not allowed to change our investment before the time period is up, at which point we sell everything and observe how much we made. We invest in a universe of assets \( A_i \) \( i \in I \). We model the one year return of these assets as random variables, \( \tilde{r}_i \). The ith asset has expected return, \( r_i \). We are allowed to invest in each asset in any amount. This allows us to normalize the amount we have to invest to one. If \( x_i \) represents the amount we invest in asset \( i \), in the simplest case we have \( x_i \geq 0 \), \( i = 1, \ldots, n \), (no short selling), and \( \sum x_i \leq 1 \). The aggregate collection of our investments (non-zero \( x_i \)) is called our portfolio, \( P \). The return (a random variable) is \( \tilde{r} = \sum x_i \tilde{r}_i \), and its expectation is \( r = \sum x_i r_i \). The risk of our portfolio is measured by the variance of the return of the total portfolio. It can be written as \( V = \sum x_i E[(\tilde{r}_i - r_i)(\tilde{r}_j - r_j)] x_j \) where \( E \) represents expectation, and \( E[(\tilde{r}_i - r_i)(\tilde{r}_j - r_j)] \) is the covariance of the random variables \( \tilde{r}_i \) and \( \tilde{r}_j \). Since \( E[(\tilde{r}_i - r_i)(\tilde{r}_j - r_j)] = E[(\tilde{r}_j - r_j)(\tilde{r}_i - r_i)] \), \( V \) is a symmetric quadratic form; i.e., it can be written in the form \( V = x^T Q x \), where \( Q \) is a symmetric matrix. Since variance is non-negative, \( Q \) is positive semi-definite. Assuming that there is no risk free portfolio is essentially equivalent to assuming \( Q \) positive definite. It is convenient to insert the factor 1/2 to the quadratic form. Then, in summary, the simplest Markowitz model is:
Minimize \(-rx + \mu \frac{1}{2} x^T Qx\)
Subject to \(ex = 1\)
\(x \geq 0\)

where \(\mu\) is a non-negative factor that represents a tradeoff between risk and return. If \(\mu\) is large that implies that risk is considered more important, while if \(\mu\) is small, the model assigns more importance to return. An equivalent model of risk-reward tradeoff is:

Minimize \(\frac{1}{2} x^T Qx\)
Subject to \(rx = r_0\)
\(ex = 1\)
\(x \geq 0\)

Where we require a minimum return \(r_0\). The higher we set this, the more risk we must take on. If we allow short selling, then the \(x_i\) are no longer required to be non-negative in the above two models. At this point we need to consider several things:

- how do we get the data - estimate the expected returns, and the covariances of returns,
- how do we solve the optimization problem, and
- how do we make more sophisticated and accurate generalizations of the model?

ESTIMATING DATA:

Achieving good estimates for the expected returns and covariances of the assets in the investment universe is probably the most important aspect of using models such as the Markowitz model for actual investment. For example, you can only imagine what would have happened if one estimated stock data using data from the dot-com period and used that data to guide investment over 2001! Moreover, some errors in using historical data is
unavoidable. See [Leunberger, 1998, Sections 8.5 and 8.6]. We on the other hand are less interested in getting rich, and more interested in learning about optimization; so we will only this brief discussion of this important topic.

QUADRATIC PROGRAMMING [Nocedal & Wright, 1999]:

The Markowitz model is a very simple instance of a convex quadratic program. Besides being the kind of optimization needed to solve the Markowitz model, quadratic programming is worth studying for other reasons. First it is just about the simplest non-linear optimization problem to solve. Second, it is a tool used in many more general optimizers for non-linear programs.

The general quadratic program is:

\[
\begin{align*}
\text{Minimize } & g(x) = px + \frac{1}{2} x^T Q x \\
\text{Subject to } & A^1 x = b^1 \\
& A^2 x \geq b^2
\end{align*}
\]

(QP)

where \( Q \) is a positive semi-definite matrix. (We can also assume without loss of generality that \( Q \) is symmetric because \( X^T Q X = X^T \left( \frac{Q^T + Q}{2} \right) X \).

Note that the gradient of \( g(x) \) is \( \nabla g(x) = p + Q x \), and the Hessian at \( x \) is \( H(x) = Q \). We will assume that \( (QP) \) is feasible, and has an optimal solution. Theorem A.11 from the Appendix guarantees that the objective is convex, and since the constraints are linear the optimization problem as a whole is convex one. Non-convex quadratic programs are much more difficult, and we will not consider them. We can then apply the KKT conditions to this problem. They are sufficient because of convexity, and necessary (we don’t need constraint qualifications for linear constraints). However, at this point it is
convenient to restrict our attention to a simpler special case.

A Simpler Case:

\[
\begin{align*}
\text{(SQP)} & \quad \text{Minimize} & \quad g(x) &= px + \frac{1}{2} x^T Q x \\
& \quad \text{Subject to} & \quad Ax &= b
\end{align*}
\]

Where \( Q \) is now assumed positive definite. The special case \( Q = I \), represents the problem of finding the closest point in the affine space \( \{x \mid Ax = b\} \) to the origin. As before, the KKT conditions are necessary and sufficient. Let \( L(x; \lambda) = px + (1/2)xQx - \lambda(Ax - b) \). The KKT Conditions are then \( \nabla_x L(x; \lambda) = 0 \), and \( \nabla_\lambda L(x; \lambda) = 0 \). The last condition is just the feasibility constraints, the first is \( Qx - A = 0 \). This gives rise to the following \((m+n)x(m+n)\) system of equations:

\[
(*) \quad \begin{pmatrix} Q & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} -p \\ b \end{pmatrix}
\]

The matrix on the left is called the KKT matrix.

**Theorem:** Let \( A \) be of full row rank, and \( Q \) be positive definite, then the KKT matrix is non-singular and there exists a unique pair \( x^*, \lambda^* \) satisfying the system of equations \((*)\).

**Proof:** Suppose \( u \) and \( v \) satisfy:

\[
(**) \quad \begin{pmatrix} Q & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\]

Now form a quadratic form out of the KKT matrix:

\[
(u^T, v^T) \begin{pmatrix} Q & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = u^T Q u = 0.
\]

Then \( u \) must be 0 since \( Q \) is positive definite. Then in \((**)\) we have \( uQ - A^T v = -A^Tv = 0 \). But \( A \) has full row rank, implying \( v \) is 0. So \((**)\) is satisfied if and only if \( u = 0 \), and \( v = 0 \). The condition that \( Q \) be positive definite can be
relaxed to $x^TQx > 0$ for all $x$ such that $Ax = 0$. See [Nocedal & Wright, 1999, pp. 444-445] for details.

Therefore to solve (SQP) we only have to solve the system of equations (*). But this is not trivial for large problems, since the system is $(m+n)x (m+n)$. Several approaches have been developed to do this efficiently. But once this can be done efficiently, it can be used to solve the more general problem (QP) using the "active set method." The general idea is that you use the complementary slackness condition of the KKT conditions for (QP), and the sign of the dual variables to help you determine what subspace of the decision space you operate in.

**Example 1:** The Markowitz Model with Short Selling [Luenberger, 1998, pp. 158-163]:

We consider the Markowitz model in the following form:

\[
\text{Minimize} \quad \frac{1}{2} x^T Q x \\
\text{Subject to} \quad r x = r_0 \\
\quad e x = 1
\]

We form the lagrangian, $L(x; \lambda, \mu) = (1/2)x^TQx - \lambda (rx - r_0) - \mu(ex - 1)$. We then get the following necessary and sufficient optimality conditions for an efficient portfolio (with short selling allowed) with return $r_0$:

\[
Qx - \lambda r - \mu e = 0 \\
rx = r_0 \\
ex = 1
\]

These are uniquely solvable (assuming the constraints are feasible), and the efficient frontier can be generated by parameterizing on $r_0$. But actually, things are even simpler.

Suppose we have two efficient portfolios, that satisfy (***)

and provide
returns \( r_0^1 \) and \( r_0^2 \) respectively. Let their optimal values be \( x_j^j, \lambda_j^j, \) and \( \mu_j^j \) for \( j = 1, 2 \). Then let 
\[
 x^\alpha = (1 - \alpha)x^1 + \alpha x^2, \lambda^\alpha = (1 - \alpha)\lambda^1 + \alpha \lambda^2, \text{ and } \mu^\alpha = (1 - \alpha)\mu^1 + \alpha \mu^2.
\]
Note that \( \alpha \) is not bounded by one, nor even necessarily non-negative. Then by direct substitution, these values also satisfy (***) with return \( r_0^\alpha = (1 - \alpha)r_0^1 + \alpha r_0^2 \). This simple observation has important financial consequences.

The Two-Fund Theorem: Given any two efficient portfolios, any other efficient portfolio can be matched (same mean and variance) as a linear combination of the other two.

This result depends strongly on the allowing of short selling. Also, it turns out that portfolios with short selling pick many assets for their efficient portfolios, but when short selling is not allowed, fewer are chosen. Also note that the model allowing short selling is a relaxation of the one where short selling is prohibited, so that short selling yields greater (or at least no lesser) returns.

Example 2:

\[
\text{Minimize } (x - a)^2 + (y-b)^2 + (z-c)^2
\]
Subject to \( x \geq 0 \)
\[
-x \geq -1
\]
\[
y \geq 0
\]
\[
-y \geq -1
\]
\[
z \geq 0
\]
\[
-z \geq -1
\]

That is, we want to minimize the distance from \((a, b, c)\) to the unit cube. Depending on the values of \( a, b, \) and \( c, \) the nearest point can be inside the cube if, \((a, b, c)\) is inside the cube, or if \((a, b, c)\) is outside the cube or on its surface the closest point can be on a planar face, a one-dimensional edge of the cube, or at an extreme point of the cube.
ALTERNATE MEASURES OF RISK:

How do we measure risk? There are many ways. Markowitz measured the risk of a portfolio by treating portfolios as random variables and using the variance of the return as a measure of risk. Variance measures the fluctuation of the value about the mean, both above and below the mean. However, for our investments we are usually much more concerned with the fluctuations below the mean (the “downside risk”); we are only too pleased if the fluctuations are above the mean. If the random fluctuations of the return are symmetric about the expected return; for example, if we assume they are normal then this is no issue because the fluctuations above the mean are statistically the same as the fluctuations below so the two measures are related by a constant factor.

We can also measure risk, based on scenarios, much like we did in our first example earlier in the semester. These techniques can give “one sided” measures of risk in a straightforward way. One way we can measure the variation is by the sum of the absolute values of the differences between the actual return and its expected value [Konno et al, 1991]. We look for the portfolio with the given expected return that minimizes the risk as measured by the sum over the periods of the absolute deviations of the total portfolios value from the mean. Let This leads to the following optimization program for determining optimal portfolios using the $L_1$ measure of risk. Let $a_{ij} = r_{ij} - r_j$ be the difference between the observed return (the ratio of the share price at time $i$ divided by its price at time $i-1$) at time $i$ for security $j$ and its expected value calculated as described above.
Minimize \( \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \)

Subject to \( \sum_{j=1}^{n} P_j x_j = M_0 \)

\( \sum_{j=1}^{n} R_j P_j x_j \geq R_0 M_0 \)

where \( x_j \geq 0, \ j = 1, \ldots, n \).

which is equivalent to:

Minimize \( \sum_{j=1}^{n} y_j \)

Subject to \( y_i + \sum_{j=1}^{n} a_{ij} x_j \geq 0, \ i = 1, \ldots, m \)

\( y_i - \sum_{j=1}^{n} a_{ij} x_j \geq 0, \ i = 1, \ldots, m \)

\( \sum_{j=1}^{n} P_j x_j = M_0 \)

\( \sum_{j=1}^{n} R_j P_j x_j \geq R_0 M_0 \)

where \( x_j \geq 0, \ j = 1, \ldots, n \)

If you want to consider only the downside risk you can drop the second set of inequalities involving the \( y_j \). So by using this alternative approach we can replace a quadratic program with a linear program, and use one-sided risk in the bargain.

You can perform a similar kind of analysis where you measure risk as the minimum value of your portfolio over the data collecting period. This also gives rise to a linear programming model [Young, 1998].
IMPROVEMENTS AND GENERALIZATIONS:

Many refinements and extensions of the basic Markowitz models are relatively easily modeled using standard optimization techniques. For example, in the models we discussed one can invest in any number of shares the investor wants, even small fractions of one share. Using integer programming you can limit invests to even shares, or to even lots (100 shares). This helps take account of transaction costs. For practical and theoretical reasons, it is often useful to have a "riskless" asset; U.S. treasury notes, bills, and bonds are often used for this. If we have a riskless asset the Hessian matrix is not positive definite (it is still positive semi-definite), but there are straightforward ways to handle this case. More complicated optimization techniques can model reinvestments.

Many theoretically trivial but practically important features of portfolios can easily be modeled by adding linear constraints. Some examples might be: suppose you didn’t want to have any asset represent more than 5% of the portfolio - this could easily by modeled by upper bounds on the $x_i$; or maybe you were willing to allow short selling but only for some percentage, say 25%, of the portfolio value - this could be modeled by a range constraint on the sum of the $x_i$; you might want to limit your investments in certain categories: technology stocks, real estate investment trusts, bond, .... - you simply put a sum constraint on each of the relevant categories.

References:


Richard Van Slyke, Revised November 16, 2007