Chapter 8

QP Models: Portfolio Optimization

8.1 Mean-Variance Optimization

Markowitz' theory of mean-variance optimization (MVO) provides a mechanism for the selection of portfolios of securities (or asset classes) in a manner that trades off the expected returns and the risk of potential portfolios. We explore this model in more detail in this chapter.

Consider assets $S_1, S_2, \ldots, S_n$ ($n \geq 2$) with random returns. Let μ_i and σ_i denote the expected return and the standard deviation of the return of asset S_i. For $i \neq j$, ρ_{ij} denotes the correlation coefficient of the returns of assets S_i and S_j. Let $\mu = [\mu_1, \ldots, \mu_n]^T$, and $\Sigma = (\sigma_{ij})$ be the $n \times n$ symmetric covariance matrix with $\sigma_{ii} = \sigma_i^2$ and $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ for $i \neq j$. Denoting by x_i the proportion of the total funds invested in security i, one can represent the expected return and the variance of the resulting portfolio $x = (x_1, \ldots, x_n)$ as follows:

$$\mathrm{E}[x] = x_1\mu_1 + \ldots + x_n\mu_n = \mu^T x,$$

and

$$\mathrm{Var}[x] = \sum_{i,j} \rho_{ij}\sigma_i\sigma_j x_i x_j = x^T \Sigma x,$$

where $\rho_{ii} \equiv 1$.

Since variance is always nonnegative, it follows that $x^T\Sigma x \geq 0$ for any x, i.e., Σ is positive semidefinite. In this section, we will assume that it is in fact positive definite, which is essentially equivalent to assuming that there are no redundant assets in our collection $S_1, S_2, \ldots, S_n$. We further assume that the set of *admissible* portfolios is a nonempty polyhedral set and represent it as $\mathcal{X} := \{x : Ax = b, Cx \geq d\}$, where A is an $m \times n$ matrix, b is an m-dimensional vector, C is a $p \times n$ matrix and d is a p-dimensional vector. In particular, one of the constraints in the set $\mathcal{X}$ is

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

Linear portfolio constraints such as short-sale restrictions or limits on asset/sector allocations are subsumed in our generic notation $\mathcal{X}$ for the polyhedral feasible set.

Recall that a feasible portfolio x is called *efficient* if it has the maximal expected return among all portfolios with the same variance, or alternatively, if it has the minimum variance among all portfolios that have at least a certain expected return. The collection of efficient portfolios form the *efficient frontier* of the portfolio universe. The efficient frontier is often represented as a curve in a two-dimensional graph where the coordinates of a plotted point corresponds to the standard deviation and the expected return of an efficient portfolio.

When we assume that Σ is positive definite, the variance is a strictly convex function of the portfolio variables and there exists a *unique* portfolio in $\mathcal{X}$ that has the minimum variance; see Exercise 7.2. Let us denote this portfolio with $x_{\min}$ and its return $\mu^T x_{\min}$ with $R_{\min}$. Note that $x_{\min}$ is an efficient portfolio. We let $R_{\max}$ denote the maximum return for an admissible portfolio.

Markowitz' *mean-variance optimization* (MVO) problem can be formulated in three different but equivalent ways. We have seen one of these formulations in the first chapter: Find the minimum variance portfolio of the securities 1 to n that yields at least a target value of expected return (say b). Mathematically, this formulation produces a quadratic programming problem:

$$\begin{array}{lrcl} \min_x & \frac{1}{2}x^T\Sigma x & & \\ & \mu^T x & \geq & R \\ & Ax & = & b \\ & Cx & \geq & d. \end{array} \tag{8.1}$$

The first constraint indicates that the expected return is no less than the target value R. Solving this problem for values of R ranging between $R_{\min}$ and $R_{\max}$ one obtains all efficient portfolios. As we discussed above, the objective function corresponds to one half the total variance of the portfolio. The constant $\frac{1}{2}$ is added for convenience in the optimality conditions–it obviously does not affect the optimal solution.

This is a convex quadratic programming problem for which the first order conditions are both necessary and sufficient for optimality. We present these conditions next. x_R is an optimal solution of problem (8.1) if and only if there exists $\lambda_R \in I\!R$, $\gamma_E \in I\!R^m$, and $\gamma_I \in I\!R^p$ satisfying the following conditions:

$$\begin{array}{c} \Sigma x_R - \lambda_R \mu - A^T\gamma_E - C^T\gamma_I = 0, \\ \mu^T x_R \geq R, \quad Ax_R = b, \quad Cx_R \geq d, \\ \lambda_R \geq 0, \quad \lambda_R(\mu^T x_R - R) = 0, \\ \gamma_I \geq 0, \quad \gamma_I^T(Cx_R - d) = 0. \end{array} \tag{8.2}$$

The two other variations of the MVO problem are the following:

$$\begin{array}{lrcl} \max_x & \mu^T x & & \\ & x^T\Sigma x & \leq & \sigma^2 \\ & Ax & = & b \\ & Cx & \geq & d. \end{array} \tag{8.3}$$

$$\begin{array}{rrcl} \max_x & \mu^T x - \frac{\delta}{2} x^T \Sigma x & & \\ & Ax & = & b \\ & Cx & \geq & d. \end{array} \tag{8.4}$$

In (8.3), σ^2 is a given upper limit on the variance of the portfolio. In (8.4), the objective function is a *risk-adjusted return* function where the constant δ serves as a risk-aversion constant. While (8.4) is another quadratic programming problem, (8.3) has a convex quadratic constraint and therefore is not a QP. This problem can be solved using the general nonlinear programming solution techniques discussed in Chapter 5. We will also discuss a reformulation of (8.3) as a second-order cone program in Chapter 10. This opens the possibility of using specialized and efficient second-order cone programming methods for its solution.

Exercise 8.1 *What are the Karush-Kuhn-Tucker optimality conditions for problems (8.3) and (8.4)?*

Exercise 8.2 *Consider the following variant of (8.4):*

$$\begin{array}{rrcl} \max_x & \mu^T x - \eta\sqrt{x^T \Sigma x} & & \\ & Ax & = & b \\ & Cx & \geq & d. \end{array} \tag{8.5}$$

For each η, let $x^*(\eta)$ denote the optimal solution of (8.5). Show that there exists a $\delta > 0$ such that $x^*(\eta)$ solves (8.4) for that δ.

8.1.1 Example

We apply Markowitz's MVO model to the problem of constructing a long-only portfolio of US stocks, bonds and cash. We will use historical return data for these three asset classes to estimate their future expected returns. We note that most models for MVO combine historical data with other indicators such as earnings estimates, analyst ratings, valuation and growth metrics, etc. Here we restrict our attention to price based estimates for expositional simplicity. We use the S&P 500 index for the returns on stocks, the 10-year Treasury bond index for the returns on bonds, and we assume that the cash is invested in a money market account whose return is the 1-day federal fund rate. The annual times series for the "Total Return" are given below for each asset between 1960 and 2003.

Year	Stocks	Bonds	MM
1960	20.2553	262.935	100.00
1961	25.6860	268.730	102.33
1962	23.4297	284.090	105.33
1963	28.7463	289.162	108.89
1964	33.4484	299.894	113.08
1965	37.5813	302.695	117.97
1966	33.7839	318.197	124.34
1967	41.8725	309.103	129.94
1968	46.4795	316.051	137.77
1969	42.5448	298.249	150.12
1970	44.2212	354.671	157.48
1971	50.5451	394.532	164.00
1972	60.1461	403.942	172.74
1973	51.3114	417.252	189.93
1974	37.7306	433.927	206.13
1975	51.7772	457.885	216.85
1976	64.1659	529.141	226.93
1977	59.5739	531.144	241.82
1978	63.4884	524.435	266.07
1979	75.3032	531.040	302.74
1980	99.7795	517.860	359.96
1981	94.8671	538.769	404.48

Year	Stocks	Bonds	MM
1982	115.308	777.332	440.68
1983	141.316	787.357	482.42
1984	150.181	907.712	522.84
1985	197.829	1200.63	566.08
1986	234.755	1469.45	605.20
1987	247.080	1424.91	646.17
1988	288.116	1522.40	702.77
1989	379.409	1804.63	762.16
1990	367.636	1944.25	817.87
1991	479.633	2320.64	854.10
1992	516.178	2490.97	879.04
1993	568.202	2816.40	905.06
1994	575.705	2610.12	954.39
1995	792.042	3287.27	1007.84
1996	973.897	3291.58	1061.15
1997	1298.82	3687.33	1119.51
1998	1670.01	4220.24	1171.91
1999	2021.40	3903.32	1234.02
2000	1837.36	4575.33	1313.00
2001	1618.98	4827.26	1336.89
2002	1261.18	5558.40	1353.47
2003	1622.94	5588.19	1366.73

Let I_{it} denote the above "Total Return" for asset $i = 1, 2, 3$ and $t = 0, \ldots T$, where $t = 0$ corresponds to 1960 and $t = T$ to 2003. For each asset i, we can convert the raw data I_{it}, $t = 0, \ldots, T$, into rates of returns r_{it}, $t = 1, \ldots, T$, using the formula

$$r_{it} = \frac{I_{i,t} - I_{i,t-1}}{I_{i,t-1}}.$$

Year	Stocks	Bonds	MM
1961	26.81	2.20	2.33
1962	-8.78	5.72	2.93
1963	22.69	1.79	3.38
1964	16.36	3.71	3.85
1965	12.36	0.93	4.32
1966	-10.10	5.12	5.40
1967	23.94	-2.86	4.51
1968	11.00	2.25	6.02
1969	-8.47	-5.63	8.97
1970	3.94	18.92	4.90
1971	14.30	11.24	4.14
1972	18.99	2.39	5.33
1973	-14.69	3.29	9.95
1974	-26.47	4.00	8.53
1975	37.23	5.52	5.20
1976	23.93	15.56	4.65
1977	-7.16	0.38	6.56
1978	6.57	-1.26	10.03
1979	18.61	-1.26	13.78
1980	32.50	-2.48	18.90
1981	-4.92	4.04	12.37
1982	21.55	44.28	8.95

Year	Stocks	Bonds	MM
1983	22.56	1.29	9.47
1984	6.27	15.29	8.38
1985	31.17	32.27	8.27
1986	18.67	22.39	6.91
1987	5.25	-3.03	6.77
1988	16.61	6.84	8.76
1989	31.69	18.54	8.45
1990	-3.10	7.74	7.31
1991	30.46	19.36	4.43
1992	7.62	7.34	2.92
1993	10.08	13.06	2.96
1994	1.32	-7.32	5.45
1995	37.58	25.94	5.60
1996	22.96	0.13	5.29
1997	33.36	12.02	5.50
1998	28.58	14.45	4.68
1999	21.04	-7.51	5.30
2000	-9.10	17.22	6.40
2001	-11.89	5.51	1.82
2002	-22.10	15.15	1.24
2003	28.68	0.54	0.98

Let R_i denote the random rate of return of asset i. From the above historical data, we can compute the arithmetic mean rate of return for each asset:

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

which gives

	Stocks	Bonds	MM
Arithmetic mean $\bar{r}_i$	12.06 %	7.85 %	6.32 %

Since the rates of return are multiplicative over time, we prefer to use the geometric mean instead of the arithmetic mean. The *geometric mean* is the constant yearly rate of return that needs to be applied in years $t = 0$ through $t = T-1$ in order to get the compounded Total Return I_{iT}, starting from I_{i0}. The formula for the geometric mean is:

$$\mu_i = \left(\prod_{t=1}^{T}(1 + r_{it})\right)^{\frac{1}{T}} - 1.$$

We get the following results.

	Stocks	Bonds	MM
Geometric mean μ_i	10.73 %	7.37 %	6.27 %

We also compute the covariance matrix:

$$\operatorname{cov}(R_i, R_j) = \frac{1}{T}\sum_{t=1}^{T}(r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j).$$

Covariance	Stocks	Bonds	MM
Stocks	0.02778	0.00387	0.00021
Bonds	0.00387	0.01112	-0.00020
MM	0.00021	-0.00020	0.00115

It is interesting to compute the volatility of the rate of return on each asset $\sigma_i = \sqrt{\operatorname{cov}(R_i, R_i)}$:

	Stocks	Bonds	MM
Volatility	16.67 %	10.55 %	3.40 %

and the correlation matrix $\rho_{ij} = \frac{\operatorname{cov}(R_i,R_j)}{\sigma_i \sigma_j}$:

Correlation	Stocks	Bonds	MM
Stocks	1	0.2199	0.0366
Bonds	0.2199	1	-0.0545
MM	0.0366	-0.0545	1

Setting up the QP for portfolio optimization

$$\begin{array}{rcll} \min \quad 0.02778x_S^2 + 2\cdot 0.00387x_Sx_B + 2\cdot 0.00021x_Sx_M & & & \\ +0.01112x_B^2 - 2\cdot 0.00020x_Bx_M + 0.00115x_M^2 & & & \\ 0.1073x_S + 0.0737x_B + 0.0627x_M & \geq & R & \quad (8.6) \\ x_S + x_B + x_M & = & 1 & \\ x_S, x_B, x_M & \geq & 0 & \end{array}$$

and solving it for $R = 6.5\%$ to $R = 10.5\%$ with increments of 0.5% we get the optimal portfolios shown in Table 8.1.1 and the corresponding variance. The optimal allocations on the efficient frontier are also depicted in the right-hand-side graph in Figure 8.1.

Based on the first two columns of Table 8.1.1, the left-hand-side graph of Figure 8.1 plots the maximum expected rate of return R of a portfolio as a function of its volatility (standard deviation). This curve is the *efficient frontier* we discussed earlier. Every possible portfolio of consisting of long positions in stocks, bonds, and money market investments is represented by a point lying on or below the efficient frontier in the standard deviation/expected return plane.

Rate of Return R	Variance	Stocks	Bonds	MM
0.065	0.0010	0.03	0.10	0.87
0.070	0.0014	0.13	0.12	0.75
0.075	0.0026	0.24	0.14	0.62
0.080	0.0044	0.35	0.16	0.49
0.085	0.0070	0.45	0.18	0.37
0.090	0.0102	0.56	0.20	0.24
0.095	0.0142	0.67	0.22	0.11
0.100	0.0189	0.78	0.22	0
0.105	0.0246	0.93	0.07	0

Table 8.1: Efficient Portfolios

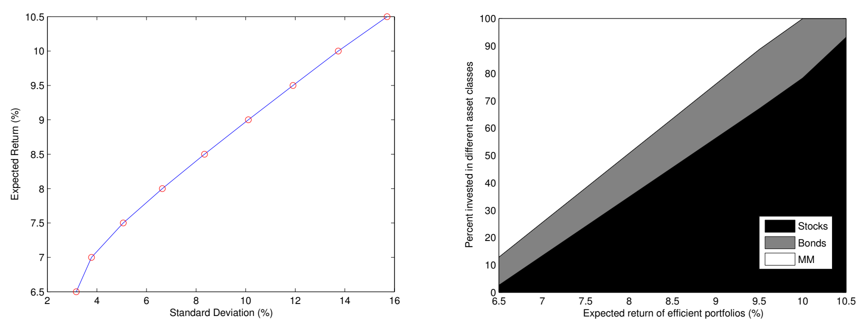

Figure 8.1: Efficient Frontier and the Composition of Efficient Portfolios

Exercise 8.3 Solve Markowitz's MVO model for constructing a portfolio of US stocks, bonds and cash using arithmetic means, instead of geometric means as above. Vary R from 6.5 % to 12 % with increments of 0.5 % . Compare with the results obtained above.

Exercise 8.4 In addition to the three securities given earlier (S&P 500 Index, 10-year Treasury Bond Index and Money Market), consider a 4th security (the NASDAQ Composite Index) with following "Total Return":

Year	NASDAQ
1960	34.461
1961	45.373
1962	38.556
1963	46.439
1964	57.175
1965	66.982
1966	63.934
1967	80.935
1968	101.79
1969	99.389
1970	89.607
1971	114.12
1972	133.73
1973	92.190
1974	59.820

Year	NASDAQ
1975	77.620
1976	97.880
1977	105.05
1978	117.98
1979	151.14
1980	202.34
1981	195.84
1982	232.41
1983	278.60
1984	247.35
1985	324.39
1986	348.81
1987	330.47
1988	381.38
1989	454.82

Year	NASDAQ
1990	373.84
1991	586.34
1992	676.95
1993	776.80
1994	751.96
1995	1052.1
1996	1291.0
1997	1570.3
1998	2192.7
1999	4069.3
2000	2470.5
2001	1950.4
2002	1335.5
2003	2003.4

Construct a portfolio consisting of the S&P 500 index, the NASDAQ index, the 10-year Treasury bond index and cash, using Markowitz's MVO model. Solve the model for different values of R.

Exercise 8.5 Repeat the previous exercise, this time assuming that one can leverage the portfolio up to 50% by borrowing at the money market rate. How do the risk/return profiles of optimal portfolios change with this relaxation? How do your answers change if the borrowing rate for cash is expected to be 1% higher than the lending rate?

8.1.2 Large-Scale Portfolio Optimization

In this section, we consider practical issues that arise when the Mean-Variance model is used to construct a portfolio from a large underlying family of assets. For concreteness, let us consider a portfolio of stocks constructed from a set of n stocks with known expected returns and covariance matrix, where n may be in the hundreds or thousands.

Diversification

In general, there is no reason to expect that solutions to the Markowitz model will be well diversified portfolios. In fact, this model tends to produce portfolios with unreasonably large weights in certain asset classes and, when short positions are allowed, unintuitively large short positions. This issue is well documented in the literature, including the paper by Green and Hollifield [34] and is often attributed to estimation errors. Estimates that may be slightly "off" may lead the optimizer to chase phantom low-risk high-return opportunities by taking large positions. Hence, portfolios chosen by this quadratic program may be subject to idiosyncratic risk. Practitioners often use additional constraints on the x_i's to insure themselves against estimation and model errors and to ensure that the chosen portfolio is well diversified. For example, a limit m may be imposed on the size of each x_i, say

$$x_i \leq m \quad \text{for } i = 1, \ldots, n.$$

One can also reduce sector risk by grouping together investments in securities of a sector and setting a limit on the exposure to this sector. For example, if m_k is the maximum that can be invested in sector k, we add the constraint

$$\sum_{i \text{ in sector } k} x_i \leq m_k.$$

Note however that, the more constraints one adds to a model, the more the objective value deteriorates. So the above approach to producing diversification, at least ex ante, can be quite costly.

Transaction Costs

We can add a portfolio turnover constraint to ensure that the change between the current holdings x^0 and the desired portfolio x is bounded by h. This constraint is essential when solving large mean-variance models since the covariance matrix is almost singular in most practical applications and hence the optimal decision can change significantly with small changes in the problem data. To avoid big changes when reoptimizing the portfolio, turnover constraints are imposed. Let y_i be the amount of asset i bought and z_i the amount sold. We write

$$\begin{aligned}
x_i - x_i^0 \leq y_i, &\quad y_i \geq 0, \\
x_i^0 - x_i \leq z_i, &\quad z_i \geq 0, \\
\sum_{i=1}^n (y_i + z_i) \leq h. &
\end{aligned}$$

Instead of a turnover constraint, we can introduce transaction costs directly into the model. Suppose that there is a transaction cost t_i proportional to the amount of asset i bought, and a transaction cost t_i' proportional to the amount of asset i sold. Suppose that the portfolio is reoptimized once per period. As above, let x^0 denote the current portfolio. Then a reoptimized portfolio is obtained by solving

$$\min \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j$$

subject to

$$\begin{aligned}
&\sum_{i=1}^n (\mu_i x_i - t_i y_i - t_i' z_i) \geq R \\
&\sum_{i=1}^n x_i = 1 \\
&x_i - x_i^0 \leq y_i \quad \text{for } i = 1, \ldots, n \\
&x_i^0 - x_i \leq z_i \quad \text{for } i = 1, \ldots, n
\end{aligned}$$

$$\begin{aligned} &y_i \geq 0 && \text{for } i = 1, \ldots, n \\ &z_i \geq 0 && \text{for } i = 1, \ldots, n \\ &x_i \text{ unrestricted} && \text{for } i = 1, \ldots, n. \end{aligned}$$

Parameter Estimation

The Markowitz model gives us an *optimal* portfolio *assuming* that we have perfect information on the μ_i's and σ_{ij}'s for the assets that we are considering. Therefore, an important practical issue is the estimation of the μ_i's and σ_{ij}'s.

A reasonable approach for estimating these data is to use time series of past returns (r_{it}= return of asset i from time $t-1$ to time t, where $i = 1, \ldots, n$, $t = 1, \ldots, T$). Unfortunately, it has been observed that small changes in the time series r_{it} lead to changes in the μ_i's and σ_{ij}'s that often lead to significant changes in the "optimal" portfolio.

Markowitz recommends using the β's of the securities to calculate the μ_i's and σ_{ij}'s as follows. Let

$$\begin{aligned} r_{it} &= \text{return of asset } i \text{ in period } t, i = 1, \ldots, n, \text{ and } t = 1, \ldots, T, \\ r_{mt} &= \text{market return in period } t, \\ r_{ft} &= \text{return of risk-free asset in period } t. \end{aligned}$$

We estimate β_i by a linear regression based on the capital asset pricing model

$$r_{it} - r_{ft} = \beta_i(r_{mt} - r_{ft}) + \varepsilon_{it}$$

where the vector ε_i represents the idiosyncratic risk of asset i. We assume that $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$. The β's can also be purchased from financial research groups and risk model providers.

Knowing β_i, we compute μ_i by the relation

$$\mu_i - E(r_f) = \beta_i(E(r_m) - E(r_f))$$

and σ_{ij} by the relation

$$\sigma_{ij} = \beta_i\beta_j\sigma_m^2 \quad \text{for } i \neq j$$

$$\sigma_{ii} = \beta_i^2\sigma_m^2 + \sigma_{\varepsilon_i}^2$$

where σ_m^2 denotes the variance of the market return and $\sigma_{\varepsilon_i}^2$ the variance of the idiosyncratic risk.

But the fundamental weakness of the Markowitz model remains, no matter how cleverly the μ_i's and σ_{ij}'s are computed: The solution is extremely sensitive to small changes in the data. Only one small change in one μ_i may produce a totally different portfolio x. What can be done in practice to overcome this problem, or at least reduce it? Michaud [51] recommends to resample returns from historical data to generate alternative μ and σ

estimates, to solve the MVO problem repeatedly with inputs generated this way, and then to combine the optimal portfolios obtained in this manner. Robust optimization approaches provide an alternative strategy to mitigate the input sensitivity in MVO models; we discuss some examples in Chapters 19 and 20. Another interesting approach is considered in the next section.

Exercise 8.6 Express the following restrictions as linear constraints:

(i) The β of the portfolio should be between 0.9 and 1.1 .

(ii) Assume that the stocks are partitioned by capitalization: large, medium and small. We want the portfolio to be divided evenly between large and medium cap stocks, and the investment in small cap stocks to be between two and three times the investment in large cap stocks.

Exercise 8.7 Using historical returns of the stocks in the DJIA, estimate their mean μ_i and covariance matrix. Let R be the median of the μ_is.

(i) Solve Markowitz's MVO model to construct a portfolio of stocks from the DJIA that has expected return at least R.

(ii) Generate a random value uniformly in the interval $[0.95\mu_i, 1.05\mu_i]$, for each stock i. Resolve Markowitz's MVO model with these mean returns, instead of μ_is as in (i). Compare the results obtained in (i) and (ii).

(iii) Repeat three more times and average the five portfolios found in (i), (ii) and (iii). Compare this portfolio with the one found in (i).

8.1.3 The Black-Litterman Model

Black and Litterman [13] recommend to combine the investor's view with the market equilibrium, as follows.

The expected return vector μ is assumed to have a probability distribution that is the product of two multivariate normal distributions. The first distribution represents the returns at market equilibrium, with mean π and covariance matrix $\tau\Sigma$, where τ is a small constant and $\Sigma = (\sigma_{ij})$ denotes the covariance matrix of asset returns (Note that the factor τ should be small since the variance $\tau\sigma_i^2$ of the random variable μ_i is typically much smaller than the variance σ_i^2 of the underlying asset returns). The second distribution represents the investor's view about the μ_i's. These views are expressed as

$$P\mu = q + \varepsilon$$

where P is a $k \times n$ matrix and q is a k-dimensional vector that are provided by the investor and ε is a normally distributed random vector with mean 0 and diagonal covariance matrix Ω (the stronger the investor's view, the smaller the corresponding $\omega_i = \Omega_{ii}$).

The resulting distribution for μ is a multivariate normal distribution with mean

$$\bar{\mu} = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}\pi + P^T\Omega^{-1}q]. \tag{8.7}$$

Black and Litterman use $\bar{\mu}$ as the vector of expected returns in the Markowitz model.

Example 8.1 *Let us illustrate the Black-Litterman approach on the example of Section 8.1.1. The expected returns on Stocks, Bonds and Money Market were computed to be*

	Stocks	*Bonds*	*MM*
Market Rate of Return	*10.73 %*	*7.37 %*	*6.27 %*

This is what we use for the vector π *representing market equilibrium. In practice,* π *is obtained from the vector of shares of global wealth invested in different asset classes via reverse optimization. We need to choose the value of the small constant* τ*. We take* $\tau = 0.1$*. We have two views that we would like to incorporate into the model. First, we hold a strong view that the Money Market rate will be 2% next year. Second, we also hold the view that S&P 500 will outperform 10-year Treasury Bonds by 5% but we are not as confident about this view. These two views can be expressed as follows*

$$\begin{aligned} \mu_M &= 0.02 \quad \textit{strong view: } \omega_1 = 0.00001 \qquad (8.8) \\ \mu_S - \mu_B &= 0.05 \quad \textit{weaker view: } \omega_2 = 0.001 \end{aligned}$$

Thus $P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$, $q = \begin{pmatrix} 0.02 \\ 0.05 \end{pmatrix}$ *and* $\Omega = \begin{pmatrix} 0.00001 & 0 \\ 0 & 0.001 \end{pmatrix}$.
Applying formula (8.7) to compute $\bar{\mu}$*, we get*

	Stocks	*Bonds*	*MM*
Mean Rate of Return $\bar{\mu}$	*11.77 %*	*7.51 %*	*2.34 %*

We solve the same QP as in (8.6) except for the modified expected return constraint:

$$\begin{array}{rrcl} \min & 0.02778x_S^2 + 2 \cdot 0.00387x_Sx_B + 2 \cdot 0.00021x_Sx_M & & \\ & +0.01112x_B^2 - 2 \cdot 0.00020x_Bx_M + 0.00115x_M^2 & & \\ & 0.1177x_S + 0.0751x_B + 0.0234x_M & \geq & R \qquad (8.9) \\ & x_S + x_B + x_M & = & 1 \\ & x_S, x_B, x_M & \geq & 0 \end{array}$$

Solving for $R = 4.0\%$ *to* $R = 11.5\%$ *with increments of 0.5% we now get the optimal portfolios and the efficient frontier depicted in Table 8.1.3 and Figure 8.2.*

Exercise 8.8 Repeat the example above, with the same investor's views, but adding the 4th security of Exercise 8.4 (the NASDAQ Composite Index).

Black and Litterman give the following intuition for their approach using the following example. Suppose we know the true structure of the asset returns: For each asset, the return is composed of an equilibrium risk premium plus a common factor and an independent shock.

$$R_i = \pi_i + \gamma_i Z + \nu_i$$

Rate of Return R	Variance	Stocks	Bonds	MM
0.040	0.0012	0.08	0.17	0.75
0.045	0.0015	0.11	0.21	0.68
0.050	0.0020	0.15	0.24	0.61
0.055	0.0025	0.18	0.28	0.54
0.060	0.0032	0.22	0.31	0.47
0.065	0.0039	0.25	0.35	0.40
0.070	0.0048	0.28	0.39	0.33
0.075	0.0059	0.32	0.42	0.26
0.080	0.0070	0.35	0.46	0.19
0.085	0.0083	0.38	0.49	0.13
0.090	0.0096	0.42	0.53	0.05
0.095	0.0111	0.47	0.53	0
0.100	0.0133	0.58	0.42	0
0.105	0.0163	0.70	0.30	0
0.110	0.0202	0.82	0.18	0
0.115	0.0249	0.94	0.06	0

Table 8.2: Black-Litterman Efficient Portfolios

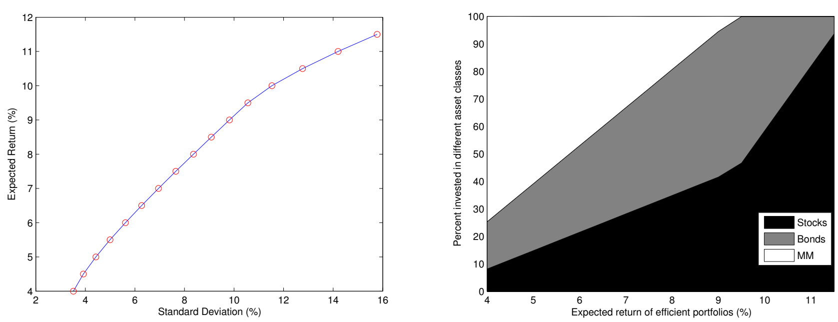

Figure 8.2: Efficient Frontier and the Composition of Efficient Portfolios using the Black-Litterman approach

where

$$
\begin{aligned}
R_i &= \text{the return on the } i\text{th asset,} \\
\pi_i &= \text{the equilibrium risk premium on the } i\text{th asset,} \\
Z &= \text{a common factor,} \\
\gamma_i &= \text{the impact of } Z \text{ on the } i\text{th asset,} \\
\nu_i &= \text{an independent shock to the } i\text{th asset.}
\end{aligned}
$$

The covariance matrix Σ of asset returns is assumed to be known. The expected returns of the assets are given by:

$$\mu_i = \pi_i + \gamma_i E[Z] + E[\nu_i].$$

While a consideration of the equilibrium motivates the Black-Litterman model, they do not assume that $E[Z]$ and $E[\nu_i]$ are equal to 0 which would indicate that the expected excess returns are equal to the equilibrium risk premiums. Instead, they assume that the expected excess returns μ_i are unobservable random variables whose distribution is determined by the distribution of $E[Z]$ and $E[\nu_i]$'s. Their additional assumptions imply that the covariance matrix of expected returns is $\tau\Sigma$ for some small positive scalar τ. All this information is assumed to be known to all investors.

Investors differ in the additional, subjective informative they have about future returns. They express this information as their "views" such as "I expect that asset A will outperform asset B by 2%". Coupled with a measure of confidence, such views can be incorporated into the equilibrium returns to generate conditional distribution of the expected returns. For example, if we assume that the equilibrium distribution of μ is given by the normal distribution $\mathrm{N}(\pi, \tau\Sigma)$ and views are represented using the constraint $P\mu = q$ (with 100% confidence), the mean $\bar{\mu}$ of the normal distribution conditional on this view is obtained as the optimal solution of the following quadratic optimization problem:

$$\begin{array}{ll} \min & (\mu - \pi)^T (\tau\Sigma)^{-1} (\mu - \pi) \\ \text{s.t.} & \mu_A - \mu_B = q. \end{array} \tag{8.10}$$

Using the KKT optimality conditions presented in Section 5.5, the solution to the above minimization problem can be shown to be

$$\bar{\mu} = \pi + (\tau\Sigma)P^T[P(\tau\Sigma)P^T]^{-1}(q - P\pi). \tag{8.11}$$

Exercise 8.9 Prove that $\bar{\mu}$ in (8.11) solves (8.10) using KKT conditions.

Of course, one rarely has 100% confidence in his/her views. In the more general case, the views are expressed as $P\mu = q + \varepsilon$ where P and q are given by the investor as above and ε is an unobservable normally distributed random vector with mean 0 and diagonal covariance matrix Ω. A diagonal Ω corresponds to the assumption that the views are independent. When this is the case, $\bar{\mu}$ is given by the Black-Litterman formula

$$\bar{\mu} = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}\pi + P^T\Omega^{-1}q],$$

as stated earlier. We refer to the Black and Litterman paper for additional details and an example of an international portfolio [13].

Exercise 8.10 Repeat Exercise 8.4, this time using the Black-Litterman methodology outlined above. Use the expected returns you computed in Exercise 8.4 as equilibrium returns and incorporate the view that NASDAQ stocks will outperform the S & P 500 stocks by 4% and that the average of NASDAQ and S & P 500 returns will exceed bond returns by 3%. Both views are relatively strong and are expressed with $\omega_1 = \omega_2 = 0.0001$.

8.1.4 Mean-Absolute Deviation to Estimate Risk

Konno and Yamazaki [43] propose a linear programming model instead of the classical quadratic model. Their approach is based on the observation that different measures of risk, such a volatility and L_1-risk, are closely related, and that alternate measures of risk are also appropriate for portfolio optimization.

The volatility of the portfolio return is

$$\sigma = \sqrt{E[(\sum_{i=1}^{n}(R_i - \mu_i)x_i)^2]}$$

where R_i denotes the random return of asset i, and μ_i denotes its mean.

The L_1-risk of the portfolio return is defined as

$$w = E[|\sum_{i=1}^{n}(R_i - \mu_i)x_i|].$$

Theorem 8.1 *(***Konno and Yamazaki***) If $(R_1, \dots, R_n)$ are multivariate normally distributed random variables, then $w = \sqrt{\frac{2}{\pi}}\sigma$.*

Proof:
Let $(\mu_1, \dots, \mu_n)$ be the mean of $(R_1, \dots, R_n)$. Also let $\Sigma = (\sigma_{ij}) \in I\!R^{n \times n}$ be the covariance matrix of $(R_1, \dots, R_n)$. Then $\sum R_i x_i$ is normally distributed [59] with mean $\sum \mu_i x_i$ and standard deviation

$$\sigma(x) = \sqrt{\sum_i \sum_j \sigma_{ij} x_i x_j}.$$

Therefore $w = E[|U|]$ where $U \sim N(0, \sigma)$.

$$w(x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \int_{-\infty}^{+\infty} |u| e^{-\frac{u^2}{2\sigma^2(x)}} du = \frac{2}{\sqrt{2\pi}\sigma(x)} \int_{0}^{+\infty} u e^{-\frac{u^2}{2\sigma^2(x)}} du = \sqrt{\frac{2}{\pi}}\sigma(x).$$

$\square$

This theorem implies that minimizing σ is equivalent to minimizing w when $(R_1, \dots, R_n)$ is multivariate normally distributed. With this assumption, the Markowitz model can be formulated as

$$\min E[|\sum_{i=1}^{n}(R_i - \mu_i)x_i|]$$

subject to

$$\sum_{i=1}^{n} \mu_i x_i \geq R$$
$$\sum_{i=1}^{n} x_i = 1$$
$$0 \leq x_i \leq m_i \quad \text{for } i = 1, \dots, n.$$

Whether $(R_1, \ldots, R_n)$ has a multivariate normal distribution or not, the above Mean-Absolute Deviation (MAD) model constructs efficient portfolios for the L_1-risk measure. Let r_{it} be the realization of random variable R_i during period t for $t = 1, \ldots, T$, which we assume to be available through the historical data or from future projection. Then

$$\mu_i = \frac{1}{T} \sum_{t=1}^{T} r_{it}$$

Furthermore

$$E[|\sum_{i=1}^{n}(R_i - \mu_i)x_i|] = \frac{1}{T} \sum_{t=1}^{T} |\sum_{i=1}^{n}(r_{it} - \mu_i)x_i|$$

Note that the absolute value in this expression makes it nonlinear. But it can be linearized using additional variables. Indeed, one can replace $|x|$ by $y + z$ where $x = y - z$ and $y, z \geq 0$. When the objective is to minimize $y + z$, at most one of y or z will be positive. Therefore the model can be rewritten as

$$\min \sum_{t=1}^{T} y_t + z_t$$

subject to

$$\begin{aligned}
& y_t - z_t = \sum_{i=1}^{n}(r_{it} - \mu_i)x_i \quad \text{for } t = 1, \ldots, T \\
& \sum_{i=1}^{n} \mu_i x_i \geq R \\
& \sum_{i=1}^{n} x_i = 1 \\
& 0 \leq x_i \leq m_i \quad \text{for } i = 1, \ldots, n \\
& y_t \geq 0,\ z_t \geq 0 \quad \text{for } t = 1, \ldots, T
\end{aligned}$$

This is a linear program! Therefore this approach can be used to solve large scale portfolio optimization problems.

Example 8.2 *We illustrate the approach on our 3-asset example, using the historical data on stocks, bonds and cash given in Section 8.1.1. Solving the linear program for $R = 6.5\%$ to $R = 10.5\%$ with increments of 0.5 % we get the optimal portfolios and the efficient frontier depicted in Table 8.2 and Figure 8.3.*

In the above table, we computed the variance of the MAD portfolio for each level R of the rate of return. These variances can be compared with the results obtained in Section 8.1.1 for the MVO portfolio. As expected, the variance of a MAD portfolio is always at least as large as that of the corresponding MVO portfolio. Note however that the difference is small. This indicates that, although the normality assumption of Theorem 8.1 does not hold, minimizing the L_1-risk (instead of volatility) produces comparable portfolios.

Rate of Return R	Variance	Stocks	Bonds	MM
0.065	0.0011	0.05	0.01	0.94
0.070	0.0015	0.15	0.04	0.81
0.075	0.0026	0.25	0.11	0.64
0.080	0.0046	0.32	0.28	0.40
0.085	0.0072	0.42	0.32	0.26
0.090	0.0106	0.52	0.37	0.11
0.095	0.0144	0.63	0.37	0
0.100	0.0189	0.78	0.22	0
0.105	0.0246	0.93	0.07	0

Table 8.3: Konno-Yamazaki Efficient Portfolios

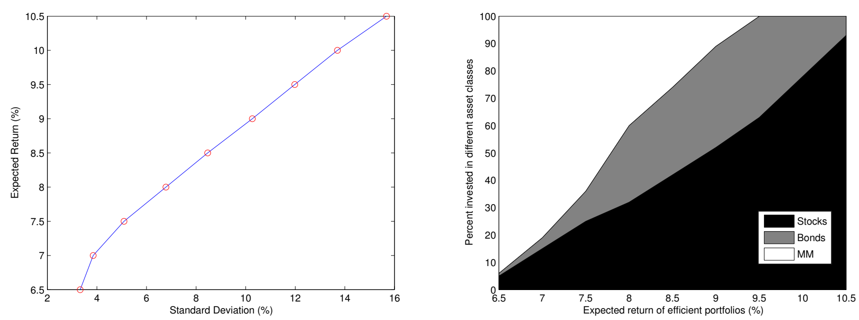

Figure 8.3: Efficient Frontier and the Composition of Efficient Portfolios using the Konno-Yamazaki approach

Exercise 8.11 Add the 4th security of Exercise 8.4 (the NASDAQ Composite Index) to the 3-asset example. Solve the resulting MAD model for varying values of R. Compare with the portfolios obtained in Exercise 8.4.

We note that the portfolios generated using the mean-absolute deviation criteria has the additional property that they are never *stochastically dominated* [64]. This is an important property as a portfolio has second-order stochastic dominance over another one if and only if it is preferred to the other by any concave (risk-averse) utility function. Unfortunately, mean-variance optimization may generate optimal portfolios that are stochastically dominated. This and other criticisms of Markowitz' mean-variance optimization model we mentioned above led to the development of alternative formulations including the Black-Litterman and Konno-Yamazaki models as well as the robust optimization models we consider in Chapter 20. Steinbach provides an excellent review of Markowitz' mean-variance optimization model, its many variations and its extensions to multi-period optimization setting [70].

8.2 Maximizing the Sharpe Ratio

Consider the setting in Section 8.1. Recall that we denote with $R_{\min}$ and $R_{\max}$ the minimum and maximum expected returns for efficient portfolios. Let us define the function

$$\sigma(R) : [R_{\min}, R_{\max}] \to I\!R, \quad \sigma(R) := (x_R^T \Sigma x_R)^{1/2},$$

where x_R denotes the unique solution of problem (8.1). Since we assumed that Σ is positive definite, it is easy to show that the function $\sigma(R)$ is strictly convex in its domain. The efficient frontier is the graph

$$E = \{(R, \sigma(R)) : R \in [R_{\min}, R_{\max}]\}.$$

We now consider a *riskless asset* whose return is $r_f \geq 0$ with probability 1. We will assume that $r_f < R_{\min}$, which is natural since the portfolio $x_{\min}$ has a positive risk associated with it while the riskless asset does not.

Return/risk profiles of different combinations of a risky portfolio with the riskless asset can be represented as a straight line—a *capital allocation line (CAL)*—on the standard deviation vs. mean graph; see Figure 8.4. The optimal CAL is the CAL that lies below all the other CALs for $R > r_f$ since the corresponding portfolios will have the lowest standard deviation for any given value of $R > r_f$. Then, it follows that this optimal CAL goes through a point on the efficient frontier and never goes above a point on the efficient frontier. In other words, the slope of the optimal CAL is a *sub-derivative* of the function $\sigma(R)$ that defines the efficient frontier. The point where the optimal CAL touches the efficient frontier corresponds to the optimal risky portfolio.

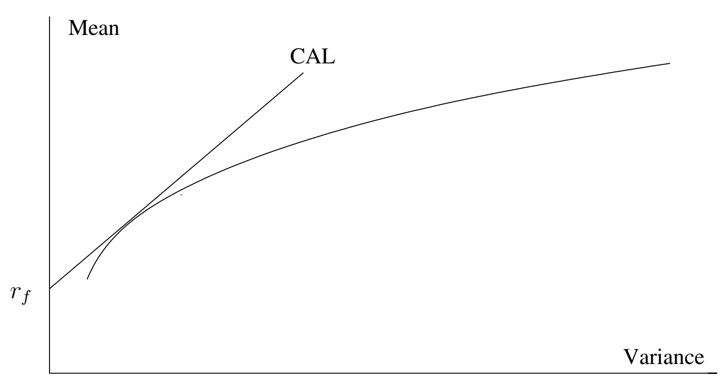

Figure 8.4: Capital Allocation Line

Alternatively, one can think of the optimal CAL as the CAL with the smallest slope. Mathematically, this can be expressed as the portfolio x that maximizes the quantity

$$h(x) = \frac{\mu^T x - r_f}{(x^T \Sigma x)^{1/2}},$$

among all $x \in S$. This quantity is precisely the *reward-to-volatility ratio* introduced by Sharpe to measure the performance of mutual funds [69]. This quantity is now more commonly known as the Sharpe ratio. The portfolio that maximizes the Sharpe ratio is found by solving the following problem:

$$\begin{array}{lrcl} \max_x & \frac{\mu^T x - r_f}{(x^T \Sigma x)^{1/2}} & & \\ & Ax & = & b \\ & Cx & \geq & d. \end{array} \tag{8.12}$$

In this form, this problem is not easy to solve. Although it has a nice polyhedral feasible region, its objective function is somewhat complicated, and worse, it is possibly non-concave. Therefore, (8.12) is not a convex optimization problem. The standard strategy to find the portfolio maximizing the Sharpe ratio, often called the *optimal risky portfolio*, is the following: First, one traces out the efficient frontier on a two dimensional return vs. standard deviation graph. Then, the point on this graph corresponding to the optimal risky portfolio is found as the tangency point of the line going through the point representing the riskless asset and is tangent to the efficient frontier. Once this point is identified, one can recover the composition of this portfolio from the information generated and recorded while constructing the efficient frontier.

Here, we describe a direct method to obtain the optimal risky portfolio by constructing a convex quadratic programming problem equivalent to (8.12). We need two assumptions: First, we assume that $\sum_{i=1}^n x_i = 1$ for any feasible portfolio x. This is a natural assumption since the x_is are the proportions of the portfolio in different asset classes. Second, we assume that there exists a feasible portfolio $\hat{x}$ with $\mu^T \hat{x} > r_f$–if all feasible portfolios have expected return bounded by the risk-free rate, there is no need to optimize, the risk-free investment dominates all others.

Proposition 8.1 *Given a set $\mathcal{X}$ of feasible portfolios with the properties that $e^T x = 1,\ \forall x \in \mathcal{X}$ and $\exists \hat{x} \in \mathcal{X}, \mu^T \hat{x} > r_f$, the portfolio x^* with the maximum Sharpe ratio in this set can be found by solving the following problem*

$$\min y^T \Sigma y \ \ s.t.\ \ (y,\kappa) \in \mathcal{X}^+,\ \ (\mu - r_f e)^T y = 1, \tag{8.13}$$

where

$$\mathcal{X}^+ \ \ := \ \ \{x \in I\!R^n, \kappa \in I\!R | \kappa > 0, \frac{x}{\kappa} \in \mathcal{X}\} \cup (0,0). \tag{8.14}$$

If (y,κ) solves (8.13), then $x^* = \frac{y}{\kappa}$. □

Problem (8.13) is a quadratic program and can be solved using the methods discussed in Chapter 7. **Proof:**
By our second assumption, it suffices to consider only those x for which $(\mu - r_f e)^T x > 0$. Let us make the following change of variables in (8.12):

$$\begin{array}{rcl} \kappa & = & \dfrac{1}{(\mu - r_f e)^T x} \\ y & = & \kappa x. \end{array}$$

Then, $\sqrt{x^T \Sigma x} = \frac{1}{\kappa}\sqrt{y^T \Sigma y}$ and the objective function of (8.12) can be written as $1/\sqrt{y^T \Sigma y}$ in terms of the new variables. Note also that

$$(\mu - r_f e)^T x > 0, x \in \mathcal{X} \Leftrightarrow \kappa > 0, \frac{y}{\kappa} \in \mathcal{X},$$

and

$$\kappa = \frac{1}{(\mu - r_f e)^T x} \Leftrightarrow (\mu - r_f e)^T y = 1$$

given $y/\kappa = x$. Thus, (8.12) is equivalent to

$$\min y^T \Sigma y \text{ s.t. } \kappa > 0,\ (y, \kappa) \in \mathcal{X},\ (\mu - r_f e)^T y = 1.$$

Since $(\mu - r_f e)^T y = 1$ rules out (0,0) as a solution, replacing $\kappa > 0,\ (y, \kappa) \in \mathcal{X}$ with $(y, \kappa) \in \mathcal{X}^+$ does not affect the solutions–it just makes the feasible set a closed set. □

Exercise 8.12 Show that $\mathcal{X}^+$ is a cone. If $\mathcal{X} = \{x | Ax \geq b, Cx = d\}$, show that $\mathcal{X}^+ = \{(x, \kappa) | Ax - b\kappa \geq 0, Cx - d\kappa = 0, \kappa \geq 0\}$. What if $\mathcal{X} = \{x : \|x\| \leq 1\}$?

Exercise 8.13 Find the Sharpe ratio maximizing portfolio of the four assets in Exercise 8.4 assuming that the risk-free return rate is 3% by solving the QP (8.13) resulting from its reformulation. Verify that the CAL passing through the point representing the standard deviation and the expected return of this portfolio is tangent to the efficient frontier.

8.3 Returns-Based Style Analysis

In two very influential articles, Sharpe described how constrained optimization techniques can be used to determine the effective asset mix of a fund using only the return time series for the fund and contemporaneous time series for returns of a number of carefully chosen asset classes [67, 68]. Often, passive indices or index funds are used to represent the chosen asset classes and one tries to determine a portfolio of these funds and indices whose returns provide the best match for the returns of the fund being analyzed. The allocations in the portfolio can be interpreted as the fund's style and consequently, this approach has become to known as *returns-based style analysis*, or RBSA.

RBSA provides an inexpensive and timely alternative to *fundamental analysis* of a fund to determine its style/asset mix. Fundamental analysis uses the information on actual holdings of a fund to determine its asset mix. When all the holdings are known, the asset mix of the fund can be inferred easily. However, this information is rarely available, and when it is available, it is often quite expensive and several weeks or months old. Since RBSA relies only on returns data which is immediately available for publicly traded funds, and well-known optimization techniques, it can be employed in circumstances where fundamental analysis cannot be used.

The mathematical model for RBSA is surprisingly simple. It uses the following generic linear factor model: Let R_t denote the return of a security–usually a mutual fund, but can be an index, etc.–in period t for $t = 1, \ldots, T$ where T corresponds to the number of periods in the modeling window. Further, let F_{it} denote the return on factor i in period t, for $i = 1, \ldots, n$, $t = 1, \ldots, T$. Then, R_t can be represented as follows:

$$\begin{aligned} R_t &= w_{1t}F_{1t} + w_{2t}F_{2t} + \ldots + w_{nt}F_{nt} + \varepsilon_t \\ &= F_t w_t + \varepsilon_t,\ t = 1, \ldots, T. \end{aligned} \tag{8.15}$$

In this equation, w_{it} quantities represent the sensitivities of R_t to each one of the n factors, and ε_t represents the non-factor return. We use the notation $w_t = \left[\begin{array}{c} w_{1t}, \ldots, w_{nt} \end{array}\right]^T$ and $F_t = \left[\begin{array}{c} F_{1t}, \ldots, F_{nt} \end{array}\right]$.

The linear factor model (8.15) has the following convenient interpretation when the factor returns F_{it} correspond to the returns of passive investments, such as those in an index fund for an asset class: One can form a benchmark portfolio of the passive investments (with weights w_{it}), and the difference between the fund return R_t and the return of the benchmark portfolio $F_t w_t$ is the non-factor return contributed by the fund manager using stock selection, market timing, etc. In other words, ε_t represents the additional return resulting from active management of the fund. Of course, this additional return can be negative.

The benchmark portfolio return interpretation for the quantity $F_t w_t$ suggests that one should choose the sensitivities (or weights) w_{it} such that they are all nonnegative and sum to one. With these constraints in mind, Sharpe proposes to choose w_{it} to minimize the variance of the non-factor return ε_t. In his model, Sharpe restricts the weights to be constant over the period in consideration so that w_{it} does not depend on t. In this case, we use $w = \left[\begin{array}{c} w_1, \ldots, w_n \end{array}\right]^T$ to denote the time-invariant factor weights and formulate the following quadratic programming problem:

$$\begin{array}{lrcl} \min_{w \in I\!R^n} & \mathrm{var}(\varepsilon_t) = \mathrm{var}(R_t - F_t w) & & \\ \text{s.t.} & \sum_{i=1}^n w_i & = & 1 \\ & w_i & \geq & 0, \forall i. \end{array} \tag{8.16}$$

The objective of minimizing the variance of the non-factor return ε_t deserves some comment. Since we are essentially formulating a tracking problem, and since ε_t represents the "tracking error", one may wonder why we do not minimize the magnitude of this quantity rather than its variance. Since the Sharpe model interprets the quantity ε_t as a consistent management effect, the objective is to determine a benchmark portfolio such that the difference between fund returns and the benchmark returns is as close to constant (i.e., variance 0) as possible. So, we want the fund return and benchmark return graphs to show two almost parallel lines with the distance between these lines corresponding to manager's consistent contribution to the fund return. This objective is almost equivalent to choosing weights in order to maximize the R-square of this regression model. The equivalence is not exact since we are using constrained regression and this may lead to correlation between ε_t and asset class returns.

The objective function of this QP can be easily computed:

$$\begin{aligned}
\operatorname{var}(R_t - w^T F_t) &= \frac{1}{T}\sum_{t=1}^{T}(R_t - w^T F_t)^2 - \left(\frac{\sum_{t=1}^{T}(R_t - w^T F_t)}{T}\right)^2 \\
&= \frac{1}{T}\|R - Fw\|^2 - \left(\frac{e^T(R - Fw)}{T}\right)^2 \\
&= \left(\frac{\|R\|^2}{T} - \frac{(e^T R)^2}{T^2}\right) - 2\left(\frac{R^T F}{T} - \frac{e^T R}{T^2}e^T F\right) w \\
&\quad + w^T\left(\frac{1}{T}F^T F - \frac{1}{T^2}F^T e e^T F\right) w.
\end{aligned}$$

Above, we introduced and used the notation

$$R = \begin{bmatrix} R_1 \\ \vdots \\ R_T \end{bmatrix}, \text{ and } F = \begin{bmatrix} F_1 \\ \cdots \\ F_T \end{bmatrix} = \begin{bmatrix} F_{11} & \ldots & F_{n1} \\ \vdots & \ddots & \vdots \\ F_{1T} & \cdots & F_{nT} \end{bmatrix}$$

and e denotes a vector of ones of appropriate size. Convexity of this quadratic function of w can be easily verified. Indeed,

$$\frac{1}{T}F^T F - \frac{1}{T^2}F^T e e^T F = \frac{1}{T}F^T\left(I - \frac{ee^T}{T}\right)F, \tag{8.17}$$

and the symmetric matrix $M = I - \frac{ee^T}{T}$ in the middle of the right-hand-side expression above is a positive semidefinite matrix with only two eigenvalues: 0 (multiplicity 1) and 1 (multiplicity $T-1$). Since M is positive semidefinite, so is $F^T M F$ and therefore the variance of ε_t is a convex quadratic function of w. Therefore, the problem (8.16) is convex quadratic programming problem and is easily solvable using well-known optimization techniques such as interior-point methods we discussed in Chapter 7.

Exercise 8.14 Implement the returns-based style analysis approach to determine the effective asset mix of your favorite mutual fund. Use the following asset classes as your "factors": Large growth stocks, large value stocks, small growth stocks, small value stocks, international stocks, and fixed income investments. You should obtain time series of returns representing these asset classes from on-line resources. You should also obtain a corresponding time series of returns for the mutual fund you picked for this exercise. Solve the problem using 30 periods of data (i.e., $T = 30$).

8.4 Recovering Risk-Neural Probabilities from Options Prices

Recall our discussion on risk-neutral probability measures in Section 4.1.2. There, we considered a one-period economy with n securities. Current prices of these securities are denoted by S_0^i for $i = 1, \ldots, n$. At the end of the current period, the economy will be in one of the states from the state space Ω. If the economy reaches state $\omega \in \Omega$ at the end of the current period, security i will have the payoff $S_1^i(\omega)$. We assume that we know all S_0^i's and $S_1^i(\omega)$'s but do not know the particular terminal state ω, which will be determined randomly.

Let r denote the one-period (riskless) interest rate and let $R = 1 + r$. A risk neutral probability measure (RNPM) is defined as the probability measure under which the present value of the expected value of future payoffs of a security equals its current price. More specifically,

- (**discrete case:**) on the state space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$, an RNPM is a vector of positive numbers $p_1, p_2, \ldots, p_m$ such that
  1. $\sum_{j=1}^m p_j = 1$,
  2. $S_0^i = \frac{1}{R} \sum_{j=1}^m p_j S_1^i(\omega_j), \ \forall i$.
- (**continuous case:**) on the state space $\Omega = (a, b)$ an RNPM is a density function $p : \Omega \to I\!R_+$ such that
  1. $\int_a^b p(\omega) d\omega = 1$,
  2. $S_0^i = \frac{1}{R} \int_a^b p(\omega) S_1^i(\omega) d\omega, \ \forall i$.

Also recall the following result from Section 4.1.2 that is often called the *First Fundamental Theorem of Asset Pricing*:

Theorem 8.2 *A risk-neutral probability measure exists if and only if there are no arbitrage opportunities.*

If we can identify a risk-neutral probability measure associated with a given state space and a set of observed prices we can price any security for which we can determine the payoffs for each state in the state space. Therefore, a fundamental problem in asset pricing is the identification of a RNPM consistent with a given set of prices. Of course, if the number of states in the state space is much larger than the number of observed prices, this problem becomes under-determined and we cannot obtain a sensible solution without introducing some additional structure into the RNPM we seek. In this section, we outline a strategy that guarantees the smoothness of the RNPM by constructing it through cubic splines. We first describe spline functions briefly:

Consider a function $f : [a, b] \to I\!R$ to be estimated using its values $f_i = f(x_i)$ given on a set of points $\{x_i\}$, $i = 1, \ldots, m+1$. It is assumed that $x_1 = a$ and $x_{m+1} = b$.

A **spline function**, or **spline**, is a *piecewise* polynomial approximation $S(x)$ to the function f such that the approximation agrees with f on each node x_i, i.e., $S(x_i) = f(x_i), \forall i$.

The graph of a spline function S contains the data points (x_i, f_i) (called *knots*) and is continuous on $[a, b]$.

A spline on $[a, b]$ is of order n if (i) its first $n-1$ derivatives exist on each interior knot, (ii) the highest degree for the polynomials defining the spline function is n.

A cubic (third order) spline uses cubic polynomials of the form $f_i(x) = \alpha_i x^3 + \beta_i x^2 + \gamma_i x + \delta_i$ to estimate the function in each interval $[x_i, x_{i+1}]$ for $i = 1, \ldots, m$. A cubic spline can be constructed in such a way that it has second derivatives at each node. For $m + 1$ knots ($x_1 = a, \ldots x_{m+1} = b$) in $[a, b]$ there are m intervals and, therefore $4m$ unknown constants to evaluate. To determine these $4m$ constants we use the following $4m$ equations:

$$f_i(x_i) = f(x_i),\ i = 1, \ldots, m, \ \text{ and } \ f_m(x_{m+1}) = f(x_{m+1}), \tag{8.18}$$
$$f_{i-1}(x_i) = f_i(x_i),\ i = 2, \ldots, m, \tag{8.19}$$
$$f'_{i-1}(x_i) = f'_i(x_i),\ i = 2, \ldots, m, \tag{8.20}$$
$$f''_{i-1}(x_i) = f''_i(x_i),\ i = 2, \ldots, m, \tag{8.21}$$
$$f''_1(x_1) = 0 \text{ and } f''_m(x_{m+1}) = 0. \tag{8.22}$$

The last condition leads to a so-called *natural* spline.

We now formulate a quadratic programming problem with the objective of finding a risk-neutral probability density function (described by cubic splines) for future values of an underlying security that best fits the observed option prices on this security.

We choose a security for consideration, say a stock or an index. We then fix an exercise date–this is future the date for which we will obtain a probability density function of the price of our security. Finally, we fix a range $[a, b]$ for possible terminal values of the price of the underlying security at the exercise date of the options and an interest rate r for the period between now and the exercise date. The inputs to our optimization problem are current market prices C_K of call options and P_K for put options on the chosen underlying security with strike price K and the chosen expiration date. This data is easily available from newspapers and online sources. Let $\mathcal{C}$ and $\mathcal{P}$, respectively, denote the set of strike prices K for which reliable market prices C_K and P_K are available. For example, $\mathcal{C}$ may denote the strike prices of call options that were traded on the day the problem is formulated.

Next, we fix a super-structure for the spline approximation to the risk-neutral density, meaning that we choose how many knots to use, where to place the knots and what kind of polynomial (quadratic, cubic, etc.) functions to use. For example, we may decide to use cubic splines and $m + 1$ equally spaced knots. The parameters of the polynomial functions that comprise the spline function will be the variables of the optimization problem we are formulating. For cubic splines with $m + 1$ knots, we will have $4m$ variables $(\alpha_i, \beta_i, \gamma_i, \delta_i)$ for $i = 1, \ldots, m$. Collectively, we will represent these variables with y. For all y chosen so that the corresponding polynomial functions f_i satisfy the equations (8.19)–(8.22) above, we will have a particular choice of a natural spline function defined on the interval $[a, b]$[1]. Let

[1]Note that we do not impose the conditions (8.18), because the values of the probability

$p_y(\cdot)$ denote this function. Imposing the following additional restrictions we make sure that p_y is a probability density function:

$$\begin{aligned} p_y(x) &\geq 0, \forall x \in [a, b] & (8.23) \\ \int_a^b p_y(\omega) d\omega &= 1. & (8.24) \end{aligned}$$

The constraint (8.24) is a linear constraint on the variables $(\alpha_i, \beta_i, \gamma_i, \delta_i)$ of the problem and can be enforced as follows:

$$\sum_{s=1}^{n_s} \int_{x_s}^{x_{s+1}} f_s(\omega) d\omega = 1. \tag{8.25}$$

On the other hand, enforcing condition (8.23) is not straightforward as it requires the function to be nonnegative for *all* values of x in $[a, b]$. Here, we relax condition (8.23), and require the cubic spline approximation to be nonnegative only at the knots:

$$p_y(x_i) \geq 0,\ i = 1, \ldots, m. \tag{8.26}$$

While this relaxation simplifies the problem greatly, we cannot guarantee that the spline approximation we generate will be nonnegative in its domain. We will discuss in Chapter 10.3 a more sophisticated technique that rigorously enforces condition (8.23).

Next, we define the discounted expected value of the terminal value of each option using p_y as the risk-neutral density function:

$$\begin{aligned} C_K(y) &:= \frac{1}{1+r} \int_a^b (\omega - K)^+ p_y(\omega) d\omega, & (8.27) \\ P_K(y) &:= \frac{1}{1+r} \int_a^b (K - \omega)^+ p_y(\omega) d\omega. & (8.28) \end{aligned}$$

Then, $C_K(y)$ is the theoretical option price if p_y is the true risk-neutral probability measure and

$$(C_K - C_K(y))^2$$

is the squared difference between the actual option price and this theoretical value. Now consider the aggregated error function for a given y:

$$E(y) := \sum_{K \in \mathcal{C}} (C_K - C_K(y))^2 + \sum_{K \in \mathcal{P}} (P_K - P_K(y))^2$$

The objective now is to choose y such that conditions (8.19)–(8.22) of spline function description as well as (8.26) and (8.24) are satisfied and $E(y)$ is minimized. This is essentially a constrained least squares problem.

We choose the number of knots and their locations so that the knots form a superset of $\mathcal{C} \cup \mathcal{P}$. Let $x_0 = a, x_1, \ldots, x_m = b$ denote the locations of the knots. Now, consider a call option with strike K and assume that

density function we are approximating are unknown and will be determined as a solution of an optimization problem.

K coincides with the location of the jth knot, i.e., $x_j = K$. Recall that y denotes collection of variables $(\alpha_i, \beta_i, \gamma_i, \delta_i)$ for $i = 1, \ldots, m$. Now, we can derive a formula for $C_K(y)$:

$$
\begin{aligned}
(1+r)C_K(y) &= \int_a^b S_y(\omega)(\omega - K)^+ d\omega \\
&= \sum_{i=1}^m \int_{x_{i-1}}^{x_i} S_y(\omega)(\omega - K)^+ d\omega \\
&= \sum_{i=j+1}^m \int_{x_{i-1}}^{x_i} S_y(\omega)(\omega - K) d\omega \\
&= \sum_{i=j+1}^m \int_{x_{i-1}}^{x_i} \left(\alpha_i\omega^3 + \beta_i\omega^2 + \gamma_i\omega + \delta_i\right)(\omega - K) d\omega.
\end{aligned}
$$

It is easily seen that this expression for $C_K(y)$ is a linear function of the components $(\alpha_i, \beta_i, \gamma_i, \delta_i)$ of the y variable. A similar formula can be derived for $P_K(y)$. The reason for choosing the knots at the strike prices is the third equation in the sequence above—we can immediately ignore some of the terms in the summation and the $(\cdot)^+$ function is linear (and not piecewise linear) in each integral.

Now, it is clear that the problem of minimizing $E(y)$ subject to spline function conditions, (8.26) and (8.24) is a quadratic optimization problem and can be solved using the techniques of the previous chapter.

8.5 Additional Exercises

Exercise 8.15 Recall the mean-variance optimization problem we considered in Section 8.1:

$$
\begin{array}{lrcl}
\min_x & x^T\Sigma x & & \\
& \mu^T x & \geq & R \\
& Ax & = & b \\
& Cx & \geq & d.
\end{array}
\tag{8.29}
$$

Now, consider the problem of finding the feasible portfolio with smallest overall variance, without imposing any expected return constraint:

$$
\begin{array}{lrcl}
\min_x & x^T\Sigma x & & \\
& Ax & = & b \\
& Cx & \geq & d.
\end{array}
\tag{8.30}
$$

(i) Does the optimal solution to (8.30) give an efficient portfolio? Why?

(ii) Let x_R, $\lambda_R \in I\!R$, $\gamma_E \in I\!R^m$, and $\gamma_I \in I\!R^p$ satisfy the optimality conditions of (8.29) (see system (8.2)). If $\lambda_R = 0$, show that x_R is an optimal solution to (8.30). (Hint: What are the optimality conditions for (8.30)? How are they related to (8.2)?)

Exercise 8.16 Classification problems are among the important classes of problems in financial mathematics that can be solved using optimization models and techniques. In a classification problem we have a vector of "features" describing an entity and the objective is to analyze the features to determine which one of the two (or more) "classes" each entity belongs to. For example, the classes might be "growth stocks" and "value stocks", and the entities (stocks) may be described by a feature vector that may contain elements such as stock price, price-earnings ratio, growth rate for the previous periods, growth estimates, etc.

Mathematical approaches to classification often start with a "training" exercise. One is supplied with a list of entities, their feature vectors and the classes they belong to. From this information, one tries to extract a mathematical structure for the entity classes so that additional entities can be classified using this mathematical structure and their feature vectors. For two-class classification, a hyperplane is probably the simplest mathematical structure that can be used to "separate" the feature vectors of these two different classes. Of course, a hyperplane is often not sufficient to separate two sets of vectors, but there are certain situations it may be sufficient.

Consider feature vectors $a_i \in I\!R^n$ for $i = 1, \ldots, k_1$ corresponding to class 1, and vectors $b_i \in I\!R^n$ for $i = 1, \ldots, k_2$ corresponding to class 2. If these two vector sets can be linearly separated, there exists a hyperplane $w^T x = \gamma$ with $w \in I\!R^n, \gamma \in I\!R$ such that

$$\begin{array}{rcl} w^T a_i & \geq & \gamma, \text{ for } i = 1, \ldots, k_1 \\ w^T b_i & \leq & \gamma, \text{ for } i = 1, \ldots, k_2. \end{array}$$

To have a "strict" separation, we often prefer to obtain w and γ such that

$$\begin{array}{rcl} w^T a_i & \geq & \gamma + 1, \text{ for } i = 1, \ldots, k_1 \\ w^T b_i & \leq & \gamma - 1, \text{ for } i = 1, \ldots, k_2. \end{array}$$

In this manner, we find two parallel lines ($w^T x = \gamma + 1$ line and $w^T x = \gamma - 1$) that form the boundary of the class 1 and class 2 portion of the vector space. This type of separation is shown in Figure 8.5.

There may be several such parallel lines that separate the two classes. Which one should one choose? A good criterion is to choose the lines that have the largest margin (distance between the lines).

a) Consider the following quadratic problem:

$$\begin{array}{lrcl} \min_{w,\gamma} & \|w\|_2^2 & & \\ & a_i^T w & \geq & \gamma + 1, \text{ for } i = 1, \ldots, k_1 \\ & b_i^T w & \leq & \gamma - 1, \text{ for } i = 1, \ldots, k_2. \end{array} \tag{8.31}$$

Show that the objective function of this problem is equivalent to maximizing the margin between the lines $w^T x = \gamma + 1$ and $w^T x = \gamma - 1$.

b) The linear separation idea we presented above can be used even when the two vector sets $\{a_i\}$ and $\{b_i\}$ are not linearly separable. (Note

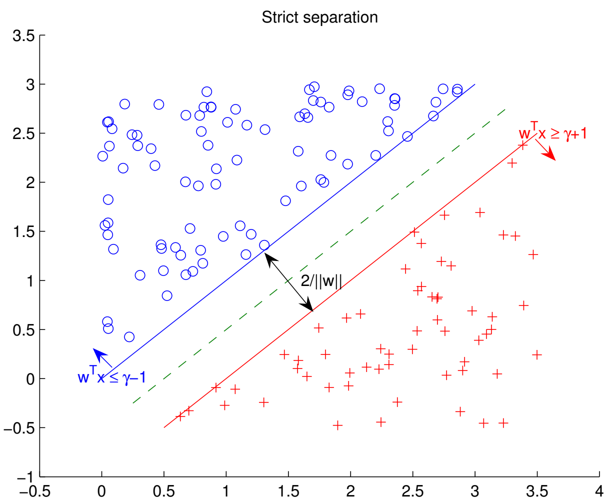

Figure 8.5: Linear separation of two classes of data points

that linearly inseparable sets will result in an infeasible problem in formulation (8.31).) This is achieved by introducing a nonnegative "violation" variable for each constraint of (8.31). Then, one has two objectives: to minimize the total of the violations of the constraints of (1) and to maximize the margin. Develop a quadratic programming model that combines these two objectives using an adjustable parameter that can be chosen in a way to put more weight on violations or margin, depending on one's preference.

Exercise 8.17 The classification problems we discussed in the previous exercise can also be formulated as linear programming problems, if one agrees to use 1-norm rather than 2-norm of w in the objective function. Recall that $\|w\|_1 = \sum_i |w_i|$. Show that if we replace $\|w\|_2^2$ with $\|w\|_1$ in the objective function of (1), we can write the resulting problem as an LP. Show also that, this new objective function is equivalent to maximizing the distance between $w^T x = \gamma + 1$ and $w^T x = \gamma - 1$ if one measures the distance using ∞-norm ($\|g\|_\infty = \max_i |g_i|$).

8.6 Case Study

Investigate the performance of one of the variations on the classical Markowitz model proposed by Michaud, or Black-Litterman or Konno-Yamazaki.

Possible suggestions:

- Choose 30 stocks and retrieve their historical returns over a meaningful horizon.

- Use the historical information to compute expected returns and the variance-covariance matrix for these stock returns.
- Set up the model and solve it with MATLAB or Excel's Solver for different levels R of expected return. Allow for short sales and include no diversification constraints.
- Recompute these portfolios with no short sales and various diversification constraints.
- Compare portfolios constructed in period t (based on historical data up to period t) by observing their performance in period $t+1$, using the actual returns from period $t+1$.
- Investigate how sensitive the optimal portfolios that you obtained are to small changes in the data. For example, how sensitive are they to a small change in the expected return of the assets?
- You currently own the following portfolio: $x_i^0 = 0.20$ for $i = 1, \ldots, 5$ and $x_i^0 = 0$ for $i = 6, \ldots, 30$. Include turnover constraints to reoptimize the portfolio for a fixed level R of expected return and observe the dependency on h, the total turnover allowed for reoptimization.
- You currently own the following portfolio: $x_i^0 = 0.20$ for $i = 1, \ldots, 5$ and $x_i^0 = 0$ for $i = 6, \ldots, 30$. Reoptimize the portfolio considering transaction costs for buying and selling. Solve for a fixed level R of expected return and observe the dependency on transaction costs.