Gatheral, J. 2006. *The Volatility Surface: A Practitioner's Guide*. New York: John Wiley.
Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*. New York: Springer.
Lehalle, C., and S. Laruelle. 2013. *Market Microstructure in Practice*. Singapore: World Scientific.
Merton, R. 1992. *Continuous-Time Finance*. Oxford: Blackwell.
Protter, P. 2013. A mathematical theory of financial bubbles. In *Paris–Princeton Lectures on Mathematical Finance 2013*, edited by V. Henderson and R. Sircar. New York: Springer.
Swindle, G. 2014. *Valuation and Risk Management in Energy Markets*. Cambridge: Cambridge University Press.

V.10 Portfolio Theory

Thomas J. Brennan, Andrew W. Lo, and Tri-Dung Nguyen

1 Basic Mean–Variance Analysis

Pioneered by the Nobel Prize–winning economist Harry Markowitz over half a century ago, portfolio theory is one of the oldest branches of modern financial economics. It addresses the fundamental question faced by an investor: how should money best be allocated across a number of possible investment choices? That is, what collection or portfolio of financial assets should be chosen? In this article, we describe the fundamentals of portfolio theory and methods for its practical implementation. We focus on a fixed time horizon for investment, which we generally take to be a year, but the period may be as short as days or as long as several years. We summarize many important innovations over the past several decades, including techniques for better understanding how financial prices behave, robust methods for estimating input parameters, Bayesian methods, and resampling techniques.

A *portfolio* is a collection of financial securities, often called *assets*, and the *return* to an asset or a portfolio is the uncertain incremental percentage financial gain or loss that results from holding the asset or portfolio over a particular time horizon. If the price of an asset i at date t is denoted p_{it}, then its return $r_{i,t+1}$ between t and $t+1$ is defined as

$$r_{i,t+1} \equiv \frac{p_{i,t+1} + d_{i,t+1}}{p_{it}} - 1, \tag{1}$$

where $d_{i,t+1}$ denotes any cash payouts made by asset i between t and $t+1$, such as a dividend payment. The return on a portfolio of assets is defined in a similar manner.

Markowitz's seminal idea was to choose an optimal portfolio using two key features of the distribution of a portfolio's return: its mean and variance. For any target level of variance, an allocation yielding the greatest mean (or expected) return should be chosen. Similarly, for any target level of expected return, an allocation yielding the lowest possible level of variance should be chosen. So far, this does not allow a unique portfolio to be selected since either the target variance or the target mean needs to be specified in advance. However, this does already allow us to sketch a curve in mean–variance space corresponding to the characteristics of the portfolios that are *efficient* inasmuch as they satisfy Markowitz's requirements. This curve is known as the *mean–variance efficient frontier* (see figure 1 for an example).

Variance is often replaced by its square root, standard deviation, resulting in an equivalent curve of efficient portfolios in the space defined by mean and standard deviation. We refer to this latter curve as the *efficient frontier*.

Consider a collection of n financial assets that are available as investment choices and let them be indexed by $i = 1, \ldots, n$, with r_i denoting the return of asset i over the applicable investment horizon (we suppress the time subscript t to simplify notation). An initial investment of one dollar in i thus yields $1 + r_i$ dollars at the end of the period of the investment. The expected return of i is $\mu_i = \sum_s r_i(s) p_s$, and the covariance of the returns of i and j is $\sigma_{i,j} = \sum_s (r_i(s) - \mu_i)(r_j(s) - \mu_j) p_s$. Here we assume that the returns have finite distributions and that all the different possible sets of returns $\boldsymbol{r} = (r_1, \ldots, r_n)$ are indexed by the parameter s, with $\boldsymbol{r}_s = (r_1(s), \ldots, r_n(s))$. The probability of state s is denoted by p_s. We write $\boldsymbol{\mu} = [\mu_i]_i$ and $\boldsymbol{\Sigma} = [\sigma_{i,j}]_{i,j}$. All vectors denote column vectors unless stated otherwise.

An allocation of funds among the assets can be thought of as a vector $\boldsymbol{\omega} = [\omega_i]_i$ of weights, with ω_i representing the proportion of available funds invested in asset i. The weights are subject to the constraint $\boldsymbol{e}^t \boldsymbol{\omega} = 1$, where $\boldsymbol{e}$ is a vector of all 1s. This is necessary so that exactly 100% of the available funds are invested. In the general case, weights are permitted to be negative, with the interpretation that these correspond to *short sales* of those assets. A short sale is a specific financial transaction in which an investor can sell a security that he does not own by borrowing it from a third party, such as a broker, with the promise to return it at a later date. Short sales allow investors who expect

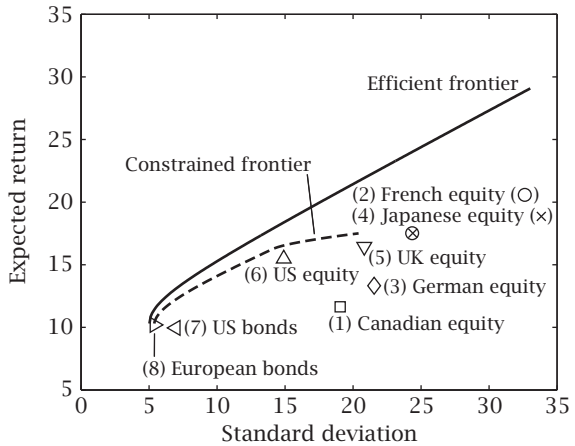

Figure 1 Efficient frontier for the collection of eight assets with characteristics described in table 1. The dashed line is the constrained frontier that does not allow negative asset weights in portfolios. For each individual asset the standard deviation and expected return are indicated on the graph.

an asset's price to decline to benefit from this expectation. When less generality is desired, more constraints can be added, such as the restriction that $\omega_i \geqslant 0$ for all i, implying that no short sales are allowed.

A portfolio formed with the weight vector $\boldsymbol{\omega}$ has expected return $\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\omega}$ and variance $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega}$. Thus, to determine the efficient frontier, we need to solve the minimization problem

$$\min \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega} \quad \text{such that } \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\omega} = \mu_{\mathrm{p}}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\omega} = 1 \qquad (2)$$

for each value of the portfolio expected return μ_{p}. The solution weight vector $\boldsymbol{\omega}^*$ will be the optimal portfolio corresponding to the target expected return μ_{p}.

Figure 1 illustrates the efficient frontier for the set of eight assets with expected returns, standard deviations, and correlations listed in table 1. In the figure, as well as in what follows, we generally restrict our attention to the "upper branch" of the efficient frontier, meaning that we do not include portfolios on the frontier that have returns lower than the return of the global minimum-variance portfolio. For any such portfolios, a higher level of return is possible for the same amount of risk.

1.1 Analytical Solutions

We can find an exact analytic solution to the problem (2). The method of LAGRANGE MULTIPLIERS [I.3 §10] is applicable, and its use in this context was pioneered by Merton. The appropriate Lagrangian is

$$\mathcal{L} = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega} + \lambda_1(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\omega} - \mu_{\mathrm{p}}) + \lambda_2(\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\omega} - 1). \qquad (3)$$

A calculation setting the gradient of $\mathcal{L}$ equal to zero shows that the solution to the minimization problem is

$$\boldsymbol{\omega}^* = \frac{\mu_{\mathrm{p}}C - B}{D}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{A - \mu_{\mathrm{p}}B}{D}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}, \qquad (4)$$

where $A = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $B = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}$, $C = \boldsymbol{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}$, and $D = AC - B^2$.

The optimal weight of (4) combined with the formula for the variance of the portfolio, namely $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega}$, allows us to calculate the minimum variance possible for a given level of expected return μ_{p}. Specifically, we have

$$(\boldsymbol{\omega}^*)^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega}^* = \frac{C\mu_p^2 - 2B\mu_{\mathrm{p}} + A}{D}. \qquad (5)$$

The minimum variance is thus simply a quadratic function of the level of expected return. Moreover, we can use (5) to see that the global minimum variance is $1/C$ and occurs at the return level $\mu_{\mathrm{p}} = B/C$. The efficient frontier can therefore be generated as the set of points

$$\begin{aligned}\mathcal{F} &= \left\{(\sigma,\mu)\colon \sigma = \frac{\sqrt{C\mu^2 - 2B\mu + A}}{\sqrt{D}}\mu \geqslant \frac{B}{C}\right\} \\ &= \left\{(\sigma,\mu)\colon \mu = \frac{B}{C} + \frac{\sqrt{D(C\sigma^2 - 1)}}{C}\sigma \geqslant \frac{1}{\sqrt{C}}\right\}. \qquad (6)\end{aligned}$$

Table 1 Expected returns, standard deviations, and correlations for a collection of eight assets. The asset numbers correspond to the named assets in figure 1. The values are annualized versions of the statistics appearing in Michaud (1998) and reflect historical data from 1978 through 1995. (ER, expected return; SD, standard deviation.)

	ER (%)	SD (%)	Correlations							
			1	2	3	4	5	6	7	8
1	11.64	19.05	1.00	0.41	0.30	0.25	0.58	0.71	0.26	0.33
2	17.52	24.35		1.00	0.62	0.42	0.54	0.44	0.22	0.26
3	13.32	21.55			1.00	0.35	0.48	0.34	0.27	0.28
4	17.52	24.39				1.00	0.40	0.22	0.14	0.16
5	16.44	20.82					1.00	0.56	0.25	0.29
6	15.48	14.90						1.00	0.36	0.42
7	9.96	6.96							1.00	0.92
8	10.20	5.40								1.00

1.2 Inequality Constraints

It is often desirable to add further constraints to the optimization problem presented in (2). For example, it may be required that the weight of a particular asset be exactly 10%, that all elements of $\boldsymbol{\omega}$ be positive, or

that no one element of $\boldsymbol{\omega}$ be greater than 50%. These constrained portfolio selection problems do not generally have simple closed-form solutions. Instead, numerical methods must be employed to compute optimal portfolios.

To the extent that more constraints are stated in the form of equalities, the Lagrange multiplier formula in (3) can be extended to have additional appropriate terms. If some of the constraints are inequalities, then the method of Lagrange multipliers can be extended using the KARUSH–KUHN–TUCKER CONDITIONS [IV.11 §2].

In practice, computer software packages are used to calculate frontiers with constraints beyond the basic requirement that $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\omega} = 1$. In figure 1, we illustrate a portion of the efficient frontier, as well as an example of a constrained frontier, for the collection of assets with characteristics described in table 1. We computed the points on the efficient frontier analytically, but those on the constrained frontier were computed using quadratic programming software.

1.3 Finding the "Best" Portfolio

We have not yet introduced a way to choose one "best" portfolio for an investor from among all those on the efficient frontier. To do so, we need to know something about the investor's *value function*, i.e., how he ranks different combinations of risk and return. In this section, we describe several commonly used methods for an investor to rank different risk–return profiles. We will refer to these ranking methodologies in subsequent sections.

The Minimum-Variance Portfolio

In an extreme case, for an investor who cares only about risk and not at all about expected return, the portfolio with the lowest level of risk should be chosen. As we saw in section 1.1, this minimum value corresponds to a variance of $1/C$ or a standard deviation of $1/\sqrt{C}$ and is displayed in figure 2. The corresponding set of weights can also be easily derived by solving the Lagrange multiplier problem corresponding to the Lagrangian $\mathcal{L} = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega} + \lambda_1(\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\omega} - 1)$. The resulting portfolio weight is seen to be $\boldsymbol{\omega}_{\min} = (1/C)\boldsymbol{\Sigma}^{-1}\boldsymbol{e}$.

Standard Value Functions

More realistically, an investor may assign higher value to portfolios with higher expected return and lower value to those with higher risk. A commonly used simple value function along these lines is

$$V_\gamma(\sigma_{\mathrm{p}}, \mu_{\mathrm{p}}) = \mu_{\mathrm{p}} - \tfrac{1}{2}\gamma\sigma_{\mathrm{p}}^2. \tag{7}$$

In this case, the value of a portfolio with parameters μ_{p} and σ_{p} is a linear function of expected return (μ_{p}) and variance (σ_{p}^2), and γ is an investor-specific parameter indicating the investor's tolerance for risk. This value function is most appropriate when μ_{p} and σ_{p} are relatively small, as may happen when the investment time horizon is short. For longer time periods, more complex functions are generally needed. These value functions are often derived from more primitive assumptions about an investor's *utility* for wealth, $U(W)$. By setting $W = W_0\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{r}$, where W_0 is initial wealth, the value function V is given by $\mathbb{E}[U(W_0\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{r})]$ according to the axioms of Von Neumann and Morgenstern's *expected utility theory*.

For the special case of the value function defined in (7), we can find the unconstrained optimal portfolio weights $\boldsymbol{\omega}_V$ and determine the corresponding highest achievable value using an appropriate Lagrangian. We find that

$$\boldsymbol{\omega}_V = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{\gamma - B}{\gamma C}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}.$$

The expected return and risk at this point of optimal value are

$$\mu_{\mathrm{p}} = \frac{B}{C} + \frac{D}{\gamma C} \quad \text{and} \quad \sigma_{\mathrm{p}} = \frac{1}{\gamma}\sqrt{\frac{D + \gamma^2}{C}}.$$

Sharpe Ratio

An investor may also look at the reward-to-variability ratio represented by a portfolio, i.e., the expected excess return he receives from the portfolio divided by the risk the portfolio represents. This ratio is referred to as the *Sharpe ratio* and can be written as

$$S(\sigma_{\mathrm{p}}, \mu_{\mathrm{p}}) = \frac{\mu_{\mathrm{p}} - r_{\mathrm{f}}}{\sigma_{\mathrm{p}}},$$

where σ_{p} and μ_{p} are the risk and expected return of the portfolio, respectively, and where r_{f} is the risk-free rate that the investor would receive if he did not invest in the portfolio. By subtracting the risk-free rate, we are measuring the incremental reward in excess of the risk-free rate that the investor receives by investing in the portfolio.

Using the characterization of points on the efficient frontier given in (6), we can calculate the optimal Sharpe ratio and corresponding portfolio weights. Specifically, we can express σ_{p}, and hence S, as a function of μ_{p},

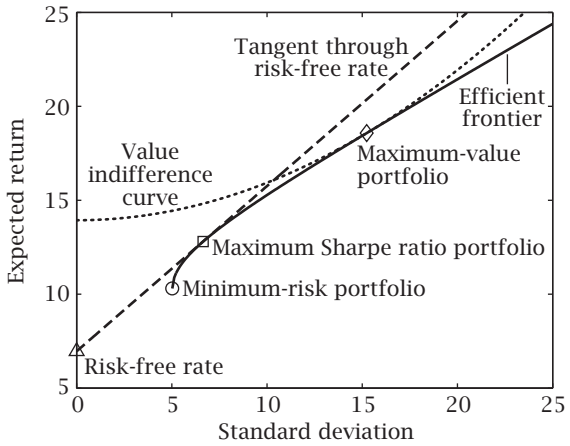

Figure 2 Optimal points on the efficient frontier for the collection of eight assets with characteristics described in table 1. The maximum Sharpe ratio is computed assuming a risk-free rate of $r_f = 6.96\%$, consistent with Michaud (1998), and the maximum value is calculated using a tolerance for risk of $\gamma = 4$.

and we then maximize S as a univariate function of μ_p and find that

$$S_{\max} = \sqrt{Cr_f^2 - 2Br_f + A}$$

for values of r_f less than the return of the minimum-variance portfolio (see figure 2). The portfolio weights are given by (4), with

$$\mu_p = \frac{A - Br_f}{B - Cr_f},$$

namely

$$\boldsymbol{\omega}_{\text{Sharpe}} = \frac{1}{B - Cr_f}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{e}).$$

1.4 Connections to the Capital Asset Pricing Model (CAPM)

The mean–variance framework developed by Markowitz was fundamental to the development of the CAPM. In the 1950s and 1960s, Tobin, Sharpe, and Lintner derived the equilibrium implications under the assumption that all investors held efficient portfolios, and this led to the *capital market line*, the line connecting the risk-free rate on the expected-return axis with the tangency portfolio on the efficient frontier in mean–standard deviation space, which is the same as the optimal Sharpe ratio portfolio derived above. Under the assumptions of the model, all investors hold a combination of positions in this tangency portfolio and the risk-free asset.

In the CAPM, aggregate positions in the risk-free asset net to zero, and the tangency portfolio represents the aggregate position in risky assets across all investors. Accordingly, this portfolio is referred to as the *market portfolio*. In the CAPM world, the expected return, μ_i, for a particular asset i can be expressed in terms of the expected return and risk of the market portfolio, as well as the covariance of the asset with the market portfolio. The precise relationship is

$$\mu_i = r_f + \beta_i(\mu_{\text{mkt}} - r_f),$$

where μ_{mkt} is the expected return of the market portfolio; $\beta_i = \text{cov}(r_i, r_{\text{mkt}})/\sigma^2_{\text{mkt}}$; r_i and r_{mkt} represent the returns on asset i and the market portfolio, respectively; and σ_{mkt} represents the standard deviation of the market portfolio. The return r_i can be written as

$$r_i = r_f + \beta_i(r_{\text{mkt}} - r_f) + \epsilon_i,$$

where ϵ_i is a stochastic variable that is uncorrelated with r_{mkt} and has zero mean. The risk represented by the ϵ_i is known as *idiosyncratic risk*. In matrix notation, the expected return vector and the covariance matrix of asset returns under the CAPM assumptions are thus

$$\left.\begin{aligned}\boldsymbol{\mu}_{\text{CAPM}} &= r_f\mathbf{e} + \boldsymbol{\beta}(\mu_{\text{mkt}} - r_f),\\ \boldsymbol{\Sigma}_{\text{CAPM}} &= \sigma^2_{\text{mkt}}\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}} + \boldsymbol{\Omega}_{\boldsymbol{\epsilon}},\end{aligned}\right\} \tag{8}$$

where $\boldsymbol{\beta}$ is the vector of β values for the assets, and where $\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}$ is the covariance matrix for the idiosyncratic risk.

Efficiency of the Market Portfolio

The importance of the mean–variance efficiency of the market portfolio was recognized early on by many authors and led to a series of debates on the testable implications of the CAPM. Markowitz has argued that empirical deviations from the CAPM are not surprising in light of the counterfactual assumptions on which the CAPM is based. In particular, he observes that: "When one clearly unrealistic assumption of the CAPM is replaced by a real-world version, some of the dramatic CAPM conclusions no longer follow." An example is the fact that unlimited borrowing and lending at identical interest rates is not possible in practice, and this limitation implies that the market portfolio need not be mean–variance efficient in equilibrium.

Impossible Frontiers

If an efficient frontier contains no portfolio with all positive weights, it is incompatible with the CAPM

and is defined to be *impossible* because, by definition, the market portfolio must have a positive weight for each asset, where the weight is proportional to the market capitalization of that asset. Brennan and Lo have demonstrated that, as the number of assets grows large, all efficient frontiers are almost surely impossible for a randomly drawn (with respect to Haar measure) covariance matrix. This result explains the near-universal disdain with which professional portfolio managers regard standard mean–variance optimization techniques: the vast majority of them are constrained to hold long-only portfolios, and hence an impossible frontier is, in fact, literally impossible for them to implement.

2 Techniques for Practical Implementation

In practice, we generally do not know the exact nature of the distributions of returns for the assets we can invest in. In fact, we do not even know the exact values of the inputs required for mean–variance optimization, i.e., $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Instead, we must find ways to estimate these quantities and then use the estimates when carrying out optimizations.

In this section we discuss several methods for the practical determination of optimal portfolios. In section 2.1 we describe the simple approach of using historical data to compute unbiased statistics under the assumption that asset returns follow a stable distribution over time. In section 2.2 we detail several methods for reducing noise, with a focus on estimating the covariance matrix, $\boldsymbol{\Sigma}$, including methods that incorporate theoretical predictions for the structure of $\boldsymbol{\Sigma}$. In section 2.3 we describe Bayesian methods that allow for better estimation of $\boldsymbol{\mu}$, as well as $\boldsymbol{\Sigma}$, and that also allow for incorporation of investor beliefs and theoretical models for the nature of asset returns. In section 2.4 we survey improved and more robust methods for selecting optimal portfolios.

2.1 Unbiased Estimators Using Historical Data

If we assume that each period's data represent an independent draw from a stable process governing asset returns, we can treat the observed historical returns as a sample from which we can estimate the desired statistics $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. For T observations of historical returns, unbiased estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T} r_{it}, \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}, \tag{9}$$

where r_{it} is the observation of the return on i in period t, and $\boldsymbol{H} = [(r_{it} - \hat{\mu}_i)]_{it}$ is a $T \times n$ matrix. An unbiased estimator of $\boldsymbol{\Sigma}^{-1}$ is $\tilde{\boldsymbol{\Sigma}}^{-1}$, with

$$\tilde{\boldsymbol{\Sigma}} = \frac{T-1}{T-n-2}\hat{\boldsymbol{\Sigma}}.$$

The estimator $\tilde{\boldsymbol{\Sigma}}$ may be more appropriate than $\hat{\boldsymbol{\Sigma}}$ when the object of interest is $\boldsymbol{\Sigma}^{-1}$ instead of $\boldsymbol{\Sigma}$, as in the formulas for the values A, B, and C in (4).

2.2 Covariance Matrix Estimation

Unbiased estimates suffer from estimation error; we describe several techniques designed to provide better estimates, with a particular focus on the covariance matrix, $\boldsymbol{\Sigma}$. These include methods for the reduction of noise as well as for the incorporation of theoretical results regarding the structure of $\boldsymbol{\Sigma}$.

Factor Analysis

The returns of the n available assets will generally not be independent. In fact, the returns may be driven in large part by a small number of common factors. If this is the case, the covariance matrix, $\boldsymbol{\Sigma}$, is less complex and estimators taking this into account may be expected to contain less noise. It may also be assumed that the common factors determine the risk-free vector, $\boldsymbol{\mu}$, but for expositional purposes we allow $\boldsymbol{\mu}$ to remain fully general and focus solely on reducing the complexity of $\boldsymbol{\Sigma}$.

The return vector for the n assets takes the form

$$\boldsymbol{r} = \boldsymbol{\mu} + \boldsymbol{V}^{\mathrm{T}}\boldsymbol{\Lambda} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\mu}$ is a vector of constants; $\boldsymbol{\Lambda}$ is a stochastic vector of n_{f} factors; $\boldsymbol{V}$ is a constant $n_{\mathrm{f}} \times n$ matrix of factor loadings; and $\boldsymbol{\epsilon}$ is a stochastic vector of residual returns for the n assets, each having zero mean. The factors represented by $\boldsymbol{\Lambda}$ are generally thought of as historically observable aggregate market variables, and various economic models give rise to factor structures for asset returns. For example, the CAPM yields a factor model with $n_{\mathrm{f}} = 1$ in which the single factor is the market-value-weighted average of all asset returns.

Because we allow $\boldsymbol{\mu}$ to remain completely general, the factors represented by the elements of $\boldsymbol{\Lambda}$ may each be assumed to have zero mean. This assumption involves no loss of generality because an adjustment to $\boldsymbol{\mu}$ can be made, if necessary, to ensure that the expected returns of the factors are zero. In addition to having zero mean, it is often assumed that $\boldsymbol{\Lambda}$ follows a multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}$. The vector

of residual returns $\boldsymbol{\epsilon}$ is also assumed to follow a multivariate normal distribution, with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$, and to be independent of $\boldsymbol{\Lambda}$. These normality assumptions do result in loss of generality of the distribution of returns, but they greatly simplify the analysis to follow. Indeed, under these assumptions $\boldsymbol{r}$ follows a multivariate normal distribution, with

$$\boldsymbol{r} = \boldsymbol{\mu} + \boldsymbol{V}^{\mathrm{T}}\boldsymbol{\Lambda} + \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{V}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}\boldsymbol{V} + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}).$$

We use the notation $\mathcal{N}(\cdot,\cdot)$ to denote a multivariate normal distribution with mean specified by the first argument and covariance matrix specified by the second argument.

An estimator $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\Lambda}}$ of the covariance matrix of the factors can be found using historical data for factor returns with a methodology similar to that described in section 2.1. Estimators for the matrix of factor loadings $\hat{\boldsymbol{V}}$ and the covariance matrix of the error terms $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}$ can be obtained using historical data and asset-by-asset linear regressions. The regression for the ith asset yields estimators both for the n_{f} elements in the ith column of $\boldsymbol{V}$ and for the variance of the ith element of $\boldsymbol{\epsilon}$, as well as for the ith element of $\boldsymbol{\mu}$. An estimator for the full covariance matrix is, therefore,

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{factor}} = \hat{\boldsymbol{V}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\Lambda}}\hat{\boldsymbol{V}} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}.$$

Covariance Shrinkage

To reduce the estimation error of the covariance matrix, $\hat{\boldsymbol{\Sigma}}$, one can take a weighted average of $\hat{\boldsymbol{\Sigma}}$ and a known covariance matrix $\boldsymbol{F}$, a process known as *shrinking* the covariance matrix $\hat{\boldsymbol{\Sigma}}$ toward the target matrix $\boldsymbol{F}$. The resulting revised estimator for the covariance matrix is

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{LW}} = \alpha\boldsymbol{F} + (1-\alpha)\hat{\boldsymbol{\Sigma}}.$$

This is a special case of *Bayesian shrinkage estimators* in which $\boldsymbol{F}$ plays the role of a *prior* and the *posterior* is given by $\hat{\boldsymbol{\Sigma}}_{\mathrm{LW}}$ (see section 2.3). There are various possibilities for the choice of $\boldsymbol{F}$, but the basic motivation is to select an $\boldsymbol{F}$ that has a known structure that is a plausible alternative to $\hat{\boldsymbol{\Sigma}}$. In this way, the shrinkage process effectively reduces estimation error while still keeping a portion of the characteristics of the unbiased estimator $\hat{\boldsymbol{\Sigma}}$. One possible choice for $\boldsymbol{F}$ is the CAPM covariance matrix, $\boldsymbol{\Sigma}_{\mathrm{CAPM}}$, described in (8), with the assumption that the covariance matrix $\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}$ for the idiosyncratic return components is diagonal.

To determine the appropriate α for the shrinkage procedure, Ledoit and Wolf, who first applied shrinkage estimation to covariance-matrix estimation, derive a consistent estimator for α that minimizes the norm:

$$\|(\alpha\boldsymbol{F} + (1-\alpha)\hat{\boldsymbol{\Sigma}}) - \boldsymbol{\Sigma}\|^2,$$

where $\|\boldsymbol{M}\|^2$ is defined to be the sum of the squares of all the entries of a matrix $\boldsymbol{M}$.

Random-Matrix Theory

Simple estimators for $\boldsymbol{\Sigma}$ based upon historical data become less reliable as the ratio of observation periods to assets ($q = T/n$) decreases. In the extreme case in which $q < 1$, the estimator in (9) is degenerate because the rank of $\boldsymbol{H}$ is less than n. Laloux, Cizeau, Potters, and Bouchaud have addressed the problem of noisy estimation when q is not much larger than 1 by arguing that $\hat{\boldsymbol{\Sigma}}$ should behave like a random matrix in such cases. To the extent that it exhibits behavior other than that of a random matrix, there should be actual information present, and this insight leads to a procedure for purging the matrix of its random noise.

Let $\boldsymbol{H}_{\mathrm{r}}$ be a $T \times n$ matrix with elements independently drawn from a normal random distribution with mean zero and standard deviation σ_{r}. The matrix $\boldsymbol{M}_{\mathrm{r}} = \boldsymbol{H}_{\mathrm{r}}^{\mathrm{T}}\boldsymbol{H}_{\mathrm{r}}$ follows a Wishart distribution, and as $T \to \infty$, $n \to \infty$, and the ratio $q = T/n$ remains constant, the eigenvalues of $\boldsymbol{M}_{\mathrm{r}}$ asymptotically all lie within an interval $[\lambda_-, \lambda_+]$, where

$$\lambda_\pm = \sigma_{\mathrm{r}}^2(1 + 1/q \pm 2\sqrt{1/q}).$$

If a matrix has, with only a few exceptions, eigenvalues that lie in this range, then it may be argued that the outliers correspond to the information content of the matrix while the other eigenvalues correspond to random noise.

Instead of focusing on the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ directly, it is more convenient to consider the corresponding correlation matrix $\hat{\boldsymbol{C}}$ because RANDOM-MATRIX THEORY [IV.24] is most easily applied to cases in which all variances are equal. $\hat{\boldsymbol{C}}$ is defined by its entries $\hat{C}_{i,j} = \hat{\Sigma}_{i,j}/\sqrt{\hat{\Sigma}_{i,i}\hat{\Sigma}_{j,j}}$ and is itself a covariance matrix for the returns of n assets but with each return rescaled so it has unit variance.

Let $\hat{\boldsymbol{C}} = \hat{\boldsymbol{Q}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{Q}}^{\mathrm{T}}$ be an eigendecomposition of $\hat{\boldsymbol{C}}$ with the eigenvalues on the diagonal of $\hat{\boldsymbol{\Lambda}}$ and the corresponding eigenvectors in the columns of $\hat{\boldsymbol{Q}}$. We compare the distribution of the eigenvalues $\hat{\lambda}_i$ to the theoretical distribution predicted for a random matrix $\boldsymbol{M}_{\mathrm{r}} = \boldsymbol{H}_{\mathrm{r}}^{\mathrm{T}}\boldsymbol{H}_{\mathrm{r}}$, where the elements of $\boldsymbol{H}_{\mathrm{r}}$ are drawn independently from a normal distribution with unit variance. The eigenvalues may be separated into two groups by using a suitable method, e.g., by specifying

a threshold, $c_{\mathrm{RM}} > 1$, such that any eigenvalue larger than $c_{\mathrm{RM}}\lambda_+$ is deemed to convey information, with the rest deemed to be random noise.

The correlation matrix can be *cleaned* by replacing eigenvalues corresponding to noise with the average of such eigenvalues and leaving the other eigenvalues as they are. This set of cleaned eigenvalues can be used to create a cleaned covariance matrix in two steps. First, define $\tilde{\boldsymbol{C}}$ to be $\hat{\boldsymbol{Q}}\tilde{\boldsymbol{\Lambda}}\hat{\boldsymbol{Q}}^{\mathrm{T}}$, where $\tilde{\boldsymbol{\Lambda}}$ is the diagonal matrix with diagonal entries equal to the cleaned eigenvalues. Second, define the cleaned covariance matrix by

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{RM}} = [\tilde{\boldsymbol{C}}_{ij}(\hat{\Sigma}_{ii}\hat{\Sigma}_{jj})^{1/2}]_{i,j}.$$

This covariance matrix can now be used along with $\hat{\boldsymbol{\mu}}$ as a basis for portfolio optimization.

It is important to underscore that the approach for cleaning a covariance matrix using the theory of random matrices is most appropriate when the value of $q = T/n$ is not much larger than 1.

Nearest Correlation Matrix

A problem that is related to the estimation of a covariance matrix is that of computing a correlation matrix that is closest in some metric to a given matrix. This problem arises routinely in applications of portfolio optimization in which financial analysts wish to impose their priors by altering various entries of the correlation matrix to conform to their beliefs, e.g., that the correlation between stock A and stock B is 0. Arbitrary changes to elements of a bona fide correlation matrix can easily violate positive-semidefiniteness, implying negative variances for certain portfolios. Nonpositive-semidefiniteness can also arise if elements of the correlation matrix are estimated individually rather than via a matrix estimator.

Under the Frobenius norm, a unique solution to the nearest correlation matrix (NCM) problem exists, and Higham has shown that it can be computed via an alternating projections algorithm that projects onto the space of matrices with unit diagonal and the cone of symmetric positive-semidefinite matrices. Although this algorithm is guaranteed to converge, it does so at a linear rate, which can be slow for large matrices. Using the dual of the NCM problem, it is possible to achieve global quadratic convergence by applying Newton's method, as shown by Qi and Sun. The NCM problem has also been extended to the case where the true correlation matrix is assumed to have a k-factor structure, where k is much less than the dimension of the correlation matrix.

Finding a Possible Frontier

It is possible to modify a covariance matrix so that the corresponding efficient frontier is possible, rather than impossible, in the sense described in section 1.4. The tangency portfolio of the CAPM should satisfy $\boldsymbol{\omega}_{\mathrm{mkt}} = (1/(B - Cr_{\mathrm{f}}))\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_{\mathrm{f}}\boldsymbol{e})$, up to scaling by a positive constant. If this equality does not hold, then an adjustment to $\boldsymbol{\Sigma}$ can be made to restore the equality. This approach was introduced by Brennan and Lo and is related to the Black–Litterman method of asset allocation with prior information, which is described further in section 2.3.

Brennan and Lo's covariance matrix, $\boldsymbol{\Sigma}_{\mathrm{poss}}$, has the desired property that $\boldsymbol{\omega}_{\mathrm{mkt}} = (1/(B - Cr_{\mathrm{f}}))\boldsymbol{\Sigma}_{\mathrm{poss}}^{-1}(\boldsymbol{\mu} - r_{\mathrm{f}}\boldsymbol{e})$, and it is constructed to be the matrix requiring the least amount of change to the original covariance matrix $\boldsymbol{\Sigma}$. Specifically, the nature of their proposed change alters $\boldsymbol{\Sigma}^{-1}$ only with respect to the one-dimensional vector space spanned by $\boldsymbol{\mu} - r_{\mathrm{f}}\boldsymbol{e}$, and the product of $\boldsymbol{\Sigma}^{-1}$ with any vector orthogonal to $\boldsymbol{\mu} - r_{\mathrm{f}}\boldsymbol{e}$ is unaffected. Brennan and Lo argue that this covariance matrix should be used by those whose best estimate of the covariance matrix is otherwise $\boldsymbol{\Sigma}$ but who also have a strong conviction that the CAPM must hold and that $\boldsymbol{\mu}$ and $\boldsymbol{\omega}_{\mathrm{mkt}}$ are, in fact, the correct expected returns and market weights.

2.3 Bayesian Methods

Estimators for the risk-free vector, $\boldsymbol{\mu}$, are often considered particularly problematic because of their large estimation errors. To improve the estimation of $\boldsymbol{\mu}$, as well as $\boldsymbol{\Sigma}$, techniques applying BAYES'S THEOREM [V.11 §1] have been proposed. The basic idea is that an investor specifies certain *prior information* about the nature of asset returns, updates this prior with additional information and observations, and finally obtains a *posterior* distribution for $\boldsymbol{r}$. Improved estimates for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are then recovered from this posterior distribution. In general, there is no restriction on the nature of the possible prior or additional information used in Bayesian methods. For the purposes of our discussion, we will restrict attention to the simple assumption that asset returns follow a normal distribution, so that $\boldsymbol{r} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown a priori, but the additional information and Bayesian updating procedure will allow estimates for these parameters to be made.

The additional information used throughout our discussion consists of a list of m separate pieces of

information, each specifying a probability distribution for the value of a linear transformation of $\boldsymbol{\mu}$. For each $j = 1, \ldots, m$, the applicable transformation is specified by a $k_j \times n$ matrix $\boldsymbol{P}_j$, and the jth piece of information is the assumption that $\boldsymbol{P}_j\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{Q}_j, \boldsymbol{S}_j)$ for some k_j-dimensional vector $\boldsymbol{Q}_j$ and some $k_j \times k_j$ covariance matrix $\boldsymbol{S}_j$. This information specifies that the transformation of $\boldsymbol{\mu}$ given by $\boldsymbol{P}_j\boldsymbol{\mu}$ is normally distributed with mean $\boldsymbol{Q}_j$ and covariance matrix $\boldsymbol{S}_j$. The value of k_j may vary with j.

With the prior and additional information in hand, we can apply Bayes's theorem for continuous distributions to obtain a posterior distribution for asset returns. The procedure is to evaluate the integral with respect to $\boldsymbol{\mu}$ of the probability density for $\boldsymbol{r}$, with unknown values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, multiplied by the product of all the probability densities corresponding to the additional information. The result is a posterior distribution for $\boldsymbol{r}$ that is normally distributed with mean and covariance given by

$$\boldsymbol{\mu}_{\text{post}} = \Big(\sum_{j=1}^{m} \boldsymbol{P}_j^{\mathrm{T}} \boldsymbol{S}_j^{-1} \boldsymbol{P}_j \Big)^{-1} \sum_{j=1}^{m} \boldsymbol{P}_j^{\mathrm{T}} \boldsymbol{S}_j^{-1} \boldsymbol{Q}_j \tag{10}$$

and

$$\boldsymbol{\Sigma}_{\text{post}} = \boldsymbol{\Sigma}\Big(\boldsymbol{I}_n - \Big(\boldsymbol{I}_n + \sum_{j=1}^{m} \boldsymbol{P}_j^{\mathrm{T}} \boldsymbol{S}_j^{-1} \boldsymbol{P}_j \boldsymbol{\Sigma} \Big)^{-1} \Big)^{-1},$$

where $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. Note that the expression for $\boldsymbol{\Sigma}_{\text{post}}$ still involves the unknown covariance matrix $\boldsymbol{\Sigma}$. Additional information can be assumed about $\boldsymbol{\Sigma}$, and a further Bayesian posterior distribution can be calculated to eliminate the dependence on $\boldsymbol{\Sigma}$. However, for our purposes we will simply replace $\boldsymbol{\Sigma}$ with the unbiased estimator $\hat{\boldsymbol{\Sigma}}$ derived from historical returns, as discussed in section 2.1. A shortcut of this sort is typical in many applications of Bayesian analysis because elimination of $\boldsymbol{\Sigma}$ generally involves more complicated integrals.

There is an alternate derivation for the expression of $\boldsymbol{\mu}_{\text{post}}$ from (10): $\boldsymbol{\mu}_{\text{post}}$ is the value of $\boldsymbol{\mu}$ that minimizes the function

$$\mathcal{F}(\boldsymbol{\mu}) = \sum_{j=1}^{m} \|\boldsymbol{P}_j\boldsymbol{\mu} - \boldsymbol{Q}_j\|^2_{\boldsymbol{S}_j^{-1}},$$

where we have used the notation $\|\boldsymbol{v}\|^2_{\boldsymbol{W}} = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{v}$. That is, $\boldsymbol{\mu}_{\text{post}}$ is the point most "compatible" with the constraints $\boldsymbol{P}_j\boldsymbol{\mu} = \boldsymbol{Q}_j$, with the uncertainty around each constraint specified by $\boldsymbol{S}_j$. To see that $\boldsymbol{\mu}_{\text{post}}$ indeed minimizes $\mathcal{F}$, compute the gradient of $\mathcal{F}$ and solve for the value of $\boldsymbol{\mu}$ that makes the gradient equal to $\boldsymbol{0}$. The resulting formula is the same as the formula for $\boldsymbol{\mu}_{\text{post}}$ in (10).

The Grand Mean

Bayesian techniques can be used to determine a posterior mean and covariance matrix based on a combination of historical information and information about a *grand mean*, an n-dimensional vector with all components equal to some real number η. The assumption is that asset returns will ultimately tend toward a common average value and, by incorporating this tendency into a Bayesian analysis, it may be possible to obtain better values for $\boldsymbol{\mu}_{\text{post}}$ and $\boldsymbol{\Sigma}_{\text{post}}$.

Jorion uses a total of $m = T + 1$ pieces of additional information to update the prior that $\boldsymbol{r}$ is normally distributed, where T is the sample size of the available return data. The first T items are based on the historical data observations and are of the form $\boldsymbol{\mu} = \boldsymbol{r}_j$ for $1 \leqslant j \leqslant T$, where $\boldsymbol{r}_j$ is the jth observed return, with uncertainty $\boldsymbol{\Sigma}$, the historical covariance matrix. The final item has the form $\boldsymbol{\mu} = \eta\boldsymbol{e}$, the vector with all elements equal to the common grand mean, with uncertainty specified by $(1/\lambda)\boldsymbol{\Sigma}$ for a suitable $\lambda > 0$.

The appropriate value of λ is estimated as

$$\hat{\lambda} = \frac{n+2}{(\hat{\boldsymbol{\mu}} - \eta_0\boldsymbol{e})^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}} - \eta_0\boldsymbol{e})},$$

where η_0 is an estimate for the expected return of the minimum-risk portfolio. After performing the necessary Bayesian updating, Jorion finds estimators for the expected return and covariance that depend only on historical data, namely

$$\hat{\boldsymbol{\mu}}_J = \frac{T}{T+\lambda}\hat{\boldsymbol{\mu}} + \frac{\lambda}{T+\lambda}\eta_0\boldsymbol{e},$$

$$\hat{\boldsymbol{\Sigma}}_J = \Big(1 + \frac{1}{T+\lambda}\Big)\tilde{\boldsymbol{\Sigma}} + \frac{\lambda}{T(T+1+\lambda)}\Big(\frac{\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}}{\boldsymbol{e}^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{e}}\Big).$$

The values $\hat{\boldsymbol{\mu}}_J$ and $\hat{\boldsymbol{\Sigma}}_J$ tend toward $\hat{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$ as $T \to \infty$, but for smaller values of T there are "correction" terms that take on a greater weight.

Market Equilibrium and Investor Beliefs

One can approach portfolio optimization by starting from a neutral market-implied set of expected returns and allowing investors to overlay their views to inform and modify these values. This method avoids the well-known difficulty of correctly predicting future returns from historical data and also provides a mechanism through which investors can express views different from those implied by history or the market.

A Bayesian technique along these lines was developed by Black and Litterman, who started with the usual prior that the distribution of $\boldsymbol{r}$ is normally distributed but assumed that the mean and covariance matrix are unknown. This prior is then updated with two types of additional information. The first is that the value of $\boldsymbol{\mu}$ has a distribution centered at $\boldsymbol{\mu}_{\text{mkt}}$, the expected return vector implied by the market, discussed further below. The second is information provided by the investor regarding his beliefs about the distribution of $\boldsymbol{\mu}$. Using all this additional information, a posterior normal distribution for $\boldsymbol{r}$ is computed.

To determine the expected return vector implied by the market, assume that the CAPM holds and that the tangency portfolio is given by $\boldsymbol{\omega}_{\text{mkt}} = (1/\gamma)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\boldsymbol{e})$, where γ is a constant that can be interpreted as the level of investor risk tolerance. Because the portfolio of market weights is easily observed, it is useful to rewrite this formula as an expression for $\boldsymbol{\mu}$ and to define this to be the value of $\boldsymbol{\mu}$ implied by the market, namely $\boldsymbol{\mu}_{\text{mkt}} = r_f\boldsymbol{e} + \gamma\boldsymbol{\Sigma}\boldsymbol{\omega}_{\text{mkt}}$. The value γ can be estimated using historical data as the level of risk tolerance that is compatible with the risk in the market portfolio and the excess return to the market portfolio over the risk-free rate. Black and Litterman assume that the uncertainty around $\boldsymbol{\mu}_{\text{mkt}}$ is $\tau\boldsymbol{\Sigma}$, where τ is a small positive number. The uncertainty is therefore assumed to be proportional to the covariance matrix for returns, but small. The exact value of τ should be chosen in such a way as to reflect the uncertainty in the mean estimator $\boldsymbol{\mu}_{\text{mkt}}$. The value of τ should generally be close to zero to reflect the idea that the uncertainty in the market-implied mean return vector should not be very large.

The second step of Black and Litterman's method is the incorporation of specified investor beliefs. An investor is allowed to express a number of views of the form

$$p_{k,1}\mu_1 + \cdots + p_{k,n}\mu_n = q_k + \epsilon_k,$$

where $k = 1,\ldots,K$ runs through the list of views expressed. The quantity ϵ_k represents a normally distributed random variable with zero mean, reflecting the uncertainty in the kth investor belief. The $p_{k,i}$ and q_k are real numbers, and μ_i represents the expected return for the ith asset. This set of beliefs can be written compactly in matrix form as

$$\boldsymbol{P}_{\text{inv}}\boldsymbol{\mu} = \boldsymbol{Q}_{\text{inv}} + \boldsymbol{\epsilon}_{\text{inv}}, \tag{11}$$

where $\boldsymbol{P}_{\text{inv}} = [p_{k,i}]_{k,i}$ is a $K \times n$ matrix, $\boldsymbol{Q}_{\text{inv}} = [q_k]_k$ is a K-dimensional vector, and $\boldsymbol{\epsilon}_{\text{inv}}$ is a multivariate normal random variable with zero mean and diagonal covariance matrix $\boldsymbol{\Omega}_{\text{inv}}$.

To combine the investor beliefs of (11) with the market-implied returns and determine a posterior distribution for $\boldsymbol{r}$, one can follow the methodology described at the start of section 2.3 with $m = 2$ pieces of information. To incorporate market-implied returns, they set $\boldsymbol{P}_1 = \boldsymbol{I}_n$, $\boldsymbol{Q}_1 = \boldsymbol{\mu}_{\text{mkt}}$, and $\boldsymbol{S}_1 = \tau\boldsymbol{\Sigma}$. To incorporate investor beliefs, they set $\boldsymbol{P}_2 = \boldsymbol{P}_{\text{inv}}$, $\boldsymbol{Q}_2 = \boldsymbol{Q}_{\text{inv}}$, and $\boldsymbol{S}_2 = \boldsymbol{\Omega}_{\text{inv}}$. They then find that the posterior mean is

$$\boldsymbol{\mu}_{BL} = ((\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}_{\text{inv}}^{\text{T}}\boldsymbol{\Omega}_{\text{inv}}^{-1}\boldsymbol{P}_{\text{inv}})^{-1} \times ((\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\mu}_{\text{mkt}} + \boldsymbol{P}_{\text{inv}}^{\text{T}}\boldsymbol{\Omega}_{\text{inv}}^{-1}\boldsymbol{Q}_{\text{inv}}).$$

This is the expected value of $\boldsymbol{r}$ that is most consistent (subject to the specified levels of uncertainty) with both the expected returns implied by the market and investor beliefs.

2.4 Other Approaches and Metrics

It is possible to extend the basic framework of mean–variance optimization by using measures of risk and reward other than the mean vector, $\boldsymbol{\mu}$, and the covariance matrix, $\boldsymbol{\Sigma}$. It is also possible to find additional ways to implement the basic mean–variance framework beyond the methods already described. In the remainder of this section we describe a few of these approaches.

One-Sided Risk Measures

Some of the most common alternative measures of portfolio risk are "one sided," in that they focus on downside risk rather than symmetric risk around an expected return. Examples include *value at risk* (VaR) and *shortfall risk*. To measure VaR, an investor must specify a threshold $0 < \theta < 1$. The VaR is an amount such that the probability that portfolio losses will equal or exceed such an amount is exactly equal to θ. Thus, an investor knows that with probability $1 - \theta$ losses will not exceed the VaR level. To measure shortfall risk, an investor must specify a benchmark, b, relative to which performance can be measured. The shortfall risk is the probability that the portfolio return will fall below b multiplied by the average amount by which the portfolio return falls below b conditional on being below b. The shortfall risk thus provides an investor with an expected value of downside exposure.

By replacing variance with a one-sided measure in choosing optimal portfolios, an investor is able to control his worst-case scenarios to within a certain

confidence threshold (for VaR) or his average loss below a benchmark (for shortfall risk). A difficulty with using these measures, however, is that they do not yield closed-form expressions and they are less intuitive. Also, from a computational perspective, nonlinear optimization will generally be necessary to find optimal portfolios with respect to these measures, which is far less efficient than the linear and quadratic programming algorithms applicable to standard mean–variance optimization problems. Nonetheless, despite the additional computational complexity, it is possible to optimize relative to these alternative measures of risk, and doing so may be desirable for investors with preferences or asset-return dynamics that are especially asymmetric.

Resampling the Efficient Frontier

A technique known as *resampling* (which is closely related to the *bootstrap technique* in statistics) can be used to determine portfolio weights. The idea is to smooth out errors arising from uncertainty in estimators for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by generating a large number of alternative possibilities for these values from a single data set, constructing a resampled efficient frontier in each alternative case, and then averaging all of the alternatives to find an average resampled frontier. Optimal portfolios are then selected from among the points on the average resampled frontier.

To construct a resampled frontier, Michaud generates a hypothetical alternative history of realized asset returns. These are chosen from a multivariate normal distribution with mean and covariance equal to the unbiased estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, determined based on the true history. The alternative history is then used to calculate portfolios on a resampled efficient frontier. This process is repeated a large number of times, with the alternative history being chosen independently each time. The number of portfolios computed on the frontier is fixed across all resamplings. In addition, an upper bound for possible returns on the resampled frontiers can be specified in order to limit the returns on all portfolios considered to a finite range. This limit may be taken to be the largest element in $\hat{\boldsymbol{\mu}}$, for example.

The individual portfolios on the resampled frontiers are averaged together to form an average resampled frontier. All these frontiers are discrete, rather than continuous, but if the number of portfolios computed is large, a close approximation to a continuous frontier is obtained. We calculate expected returns and standard deviations for each portfolio on the average resampled frontier using estimates for mean and return such as $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$. These values allow us to determine which portfolio is optimal under a specified metric, such as minimum risk, maximum utility, or maximum Sharpe ratio. Alternatively, we could compute the optimal portfolio along each resampled frontier and average the results. The answer in this situation is generally not the same as the optimal portfolio on the average resampled frontier, but it is generally computationally much easier because it avoids the need to compute all points on the frontier for each resampling.

Ordering of Returns

Mean–variance portfolio optimization can be extended by allowing an investor to specify less information about asset returns than is encompassed by complete knowledge of the vector $\boldsymbol{\mu}$ of expected returns. For example, an investor may specify a list of inequalities and interrelationships that will hold for elements of the return vector $\boldsymbol{r}$. This type of information leads to a much larger set of "efficient" portfolios than mean–variance optimization, and it is thus more complicated to select a single optimal portfolio from among all the efficient ones. Almgren and Chriss resolve this difficulty by introducing a methodology for ranking portfolios in light of the information specified by the investor. They also cast their methodology in a manner that makes it computationally feasible to determine an optimal portfolio.

Acknowledgments. We thank Nick Higham for many helpful comments and suggestions and Jayna Cummings for excellent editorial assistance. Research support from the MIT Laboratory for Financial Engineering and the Northwestern University School of Law Faculty Research Program is gratefully acknowledged.

Further Reading

Almgren, R., and N. Chriss. 2006. Optimal portfolios from ordering information. *Journal of Risk* 9:1–47.

Black, F., and R. Litterman. 1992. Global portfolio optimization. *Financial Analysts Journal* 48(5):28–43.

Brennan, T., and A. Lo. 2010. Impossible frontiers. *Management Science* 56:905–23.

Higham, N. J. 2002. Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis* 22:329–43.

Jorion, P. 1986. Bayes–Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21: 279–92.

Laloux, L., P. Cizeau, M. Potters, and J.-P. Bouchaud. 2000. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance* 3(3): 391–97.

Markowitz, H. 1987. *Mean–Variance Analysis in Portfolio Choice and Capital Markets*. Oxford: Basil Blackwell.

Merton, R. 1972. An analytical derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis* 7(4):1851–72.

Michaud, R. 1998. *Efficient Asset Management*. Boston, MA: Harvard Business School Press.

V.11 Bayesian Inference in Applied Mathematics

Desmond J. Higham

1 Use of Bayes's Theorem

Deterministic models can be used to represent a huge variety of physical systems. Applied mathematicians are traditionally schooled in a framework where a given mathematical model is presented, or developed, with known values for all input data, such as problem domain, initial conditions, boundary conditions, and model parameters. Our task is then to analyze properties of the solution by whatever means we have at our disposal, for example, finding the exact solution, developing approximations under asymptotic limits, or applying computational methods. However, even when a deterministic model is appropriate, there are many realistic scenarios where uncertainty arises and it becomes beneficial to employ tools from statistics. Here are four illustrative examples.

(1) A partial differential equation describes the spread of a pollutant after an environmental disaster. However, the initial location and quantity of the pollutant can only be estimated. What is our uncertainty in the pollutant level after one week, given the uncertainty in the initial data? This is a *problem-sensitivity* or *conditioning* question.

(2) A pair of dice are rolled on a table. It is not practical to measure the initial location and velocity of the dice and then solve their equations of motion until they come to a halt. Instead, we can model each die independently as a random variable that is equally likely to take the value $1, 2, \ldots, 6$. Here, we are introducing randomness as a convenient *modeling approximation*.

(3) A large chemical reaction network is represented by a system of ordinary differential equations. However, the value of one reaction rate constant is not currently known and is too tricky to measure directly from a laboratory experiment. In this case, we would like to *infer* a value for the unknown rate constant using whatever data is available about the full system. This is a *parameter-fitting*, or *model-calibration*, problem, where the key question is what parameter value causes the model to best fit the data or, from a statistical inference perspective: for each possible choice of the parameter value, what is my degree of belief that this choice is correct?

(4) Biologists have two competing, and incompatible, theories for the mechanism by which signals are passed through a transduction network. These lead to two different deterministic mathematical models, each having one or more unknown modeling parameters. Given some experimental data, representing outputs from the network, which of the two models is most likely to be correct? This is a *model-comparison* problem.

Questions of this type, where models based on mechanistic laws of motion meet real data, lie at the intersection between applied mathematics and applied statistics. In recent years the term UNCERTAINTY QUANTIFICATION [II.34] has been coined to describe this field, although the phrase also has many other connotations.

A powerful tool for statistical inference, and hence for uncertainty quantification, is Bayes's theorem (also commonly referred to as Bayes' theorem), which allows us to update our beliefs according to data. The theorem is named after the Reverend Thomas Bayes (1701–61), a British mathematician and Presbyterian minister.

Considering item (3) from the list above, let γ denote the unknown problem parameter in our model and let Y denote observational data consisting of a time series of chemical species levels. Bayes's theorem may then be written

$$P(\gamma \mid Y) = \frac{P(Y \mid \gamma)P(\gamma)}{P(Y)}.$$

Here, $P(\gamma \mid Y)$, known as the *posterior distribution*, answers our question. It quantifies the probability of the model parameter γ given the data Y. On the right-hand side we have $P(Y \mid \gamma)$, known as the *likelihood*, quantifying the probability of the data Y arising given the model parameter γ. This value is available to us, since we have access to the mathematical model. Also