IEOR E4602: Quantitative Risk Management Fall 2016
© 2016 by Martin Haugh

Asset Allocation and Risk Management

These lecture notes provide an introduction to asset allocation and risk management. We begin with a brief review of: (i) the mean-variance analysis of Markowitz (1952) and (ii) the Capital Asset Pricing Model (CAPM). We also review the problems associated with implementing mean-variance analysis in practice and discuss some approaches for tackling these problems. We then discuss the Black-Litterman approach for incorporating subjective views on the markets and also describe some simple extensions of this approach. Mean - Value-at-Risk (Mean-VaR) portfolio optimization is also studied and we highlight how this is generally a bad idea as the optimal solution can often inherit the less desirable properties of VaR. Finally we discuss some recent and very useful results for solving mean-CVaR optimization problems. These results lead to problem formulations and techniques that can be useful in many settings including non-Gaussian settings.

We will only briefly discuss issues related to parameter estimation in these notes. This is a very important topic as optimal asset allocations are typically very sensitive to the *estimated* expected returns. Estimated covariance matrices also play an important role and so many estimation techniques including shrinkage, Bayesian and robust estimation techniques are employed in practice. It is also worth mentioning that in practice estimation errors are often significant and this often results in the "optimal portfolios" performing no better than portfolios chosen by simple heuristics. Nonetheless, some of the techniques we describe in these notes can be particularly useful when we have strong views on the market that we wish to incorporate into our portfolio selection. They are also useful in settings where it is desirable to control extreme downside risks such as the CVaR of a portfolio.

---

1 Review of Mean-Variance Analysis and the CAPM

Consider a one-period market with n securities which have identical expected returns and variances, i.e. $\mathrm{E}[R_i] = \mu$ and $\mathrm{Var}(R_i) = \sigma^2$ for $i = 1, \ldots, n$. We also suppose $\mathrm{Cov}(R_i, R_j) = 0$ for all $i \neq j$. Let w_i denote the fraction of wealth invested in the i^{th} security at time $t = 0$. Note that we must have $\sum_{i=1}^{n} w_i = 1$ for any portfolio. Consider now two portfolios:

Portfolio A: 100% invested in security # 1 so that $w_1 = 1$ and $w_i = 0$ for $i = 2, \ldots, n$.

Portfolio B: An equi-weighted portfolio so that $w_i = 1/n$ for $i = 1, \ldots, n$.

Let R_A and R_B denote the random returns of portfolios A and B, respectively. We immediately have

$$
\begin{aligned}
\mathrm{E}[R_A] &= \mathrm{E}[R_B] = \mu \\
\mathrm{Var}(R_A) &= \sigma^2 \\
\mathrm{Var}(R_B) &= \sigma^2/n.
\end{aligned}
$$

The two portfolios therefore have the same expected return but very different return variances. A risk-averse investor should clearly prefer portfolio B because this portfolio benefits from diversification without sacrificing any expected return. This was the cental insight of Markowitz who (in his framework) recognized that investors seek to minimize variance for a given level of expected return or, equivalently, they seek to maximize expected return for a given constraint on variance.

Before formulating and solving the mean variance problem consider Figure 1 below. There were $n = 8$ securities with given mean returns, variances and covariances. We generated $m = 200$ random portfolios from these n securities and computed the expected return and volatility, i.e. standard deviation, for each of them. They are

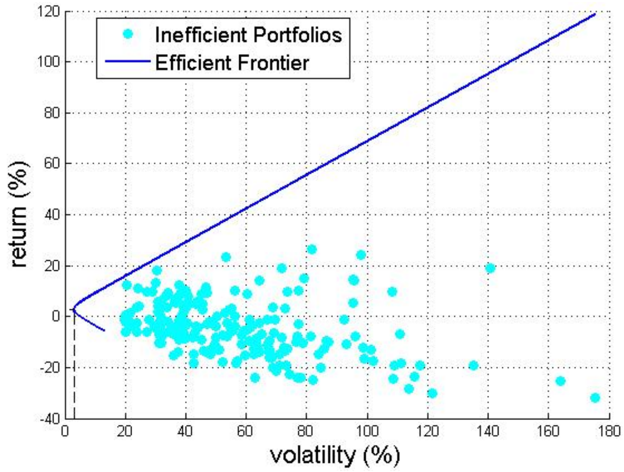

Figure 1: Sample Portfolios and the Efficient Frontier (without a Riskfree Security).

plotted in the figure and are labelled "inefficient". This is because every one of these random portfolios can be improved. In particular, for the same expected return it is possible to find an "efficient" portfolio with a smaller volatility. Alternatively, for the same volatility it is possible to find an efficient portfolio with higher expected return.

1.1 The Efficient Frontier without a Risk-free Asset

We will consider first the mean-variance problem when a risk-free security is not available. We assume that there are n risky securities with the corresponding return vector, $\mathbf{R}$, satisfying

$$\mathbf{R} \sim \mathsf{MVN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

The mean-variance portfolio optimization problem is formulated as:

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \tag{1}$$

$$\begin{aligned} \text{subject to} \quad \mathbf{w}^\top \boldsymbol{\mu} &= p \\ \text{and} \quad \mathbf{w}^\top \mathbf{1} &= 1. \end{aligned}$$

Note that the specific value of p will depend on the risk aversion of the investor. This is a simple quadratic optimization problem and it can be solved via standard Lagrange multiplier methods.

Exercise 1 *Solve the mean-variance optimization problem (1).*

When we plot the mean portfolio return, p, against the corresponding minimal portfolio volatility / standard deviation we obtain the so-called *portfolio frontier*. We can also identify the portfolio having minimal variance among all risky portfolios: this is called the *minimum variance portfolio*. The points on the portfolio frontier with expected returns greater than the minimum variance portfolio's expected return, $\bar{R}_{mv}$ say, are said to lie on the *efficient frontier*. The efficient frontier is plotted as the upper blue curve in Figure 1 ar alternatively, the blue curve in Figure 2.

Exercise 2 *Let $\mathbf{w}_1$ and $\mathbf{w}_2$ be (mean-variance) efficient portfolios corresponding to expected returns r_1 and r_2, respectively, with $r_1 \neq r_2$. Show that* **all** *efficient portfolios can be obtained as linear combinations of $\mathbf{w}_1$ and $\mathbf{w}_2$.*

The result of the previous exercise is sometimes referred to as a **2-fund theorem**.

1.2 The Efficient Frontier with a Risk-free Asset

We now assume that there is a risk-free security available with risk-free rate equal to r_f. Let $\mathbf{w} := (w_1, \ldots, w_n)'$ be the vector of portfolio weights on the n risky assets so that $1 - \sum_{i=1}^{n} w_i$ is the weight on the risk-free security. An investor's portfolio optimization problem may then be formulated as

$$\min_{\mathbf{w}} \ \frac{1}{2}\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \tag{2}$$

$$\text{subject to} \quad \left(1 - \sum_{i=1}^{n} w_i\right) r_f \ + \ \mathbf{w}^\top \boldsymbol{\mu} \ = \ p.$$

The optimal solution to (2) is given by

$$\mathbf{w} \ = \ \xi \, \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \mathbf{1}) \tag{3}$$

where $\xi := \sigma_{min}^2/(p - r_f)$ and σ_{min}^2 is the minimized variance, i.e., twice the value of the optimal objective function in (2). It satisfies

$$\sigma_{min}^2 \ = \ \frac{(p - r_f)^2}{(\boldsymbol{\mu} - r_f \mathbf{1})^\top \ \boldsymbol{\Sigma}^{-1} \ (\boldsymbol{\mu} - r_f \mathbf{1})} \tag{4}$$

where $\mathbf{1}$ is an $n \times 1$ vector of ones. While ξ (or p) depends on the investor's level of risk aversion it is often inferred from the market portfolio. For example, if we take $p - r_f$ to denote the average excess market return and σ_{min}^2 to denote the variance of the market return, then we can take $\sigma_{min}^2/(p - r_f)$ as the average or market value of ξ.

Suppose now that $r_f < \bar{R}_{mv}$. When we allow our portfolio to include the risk-free security the efficient frontier becomes a straight line that is tangential to the risky efficient frontier and with a y-intercept equal to the risk-free rate. This is plotted as the red line in Figure 2. That the efficient frontier is a straight line when we include the risk-free asset is also clear from (4) where we see that σ_{min} is linear in p. Note that this result is a **1-fund theorem** since every investor will optimally choose to invest in a combination of the risk-free security and a single risky portfolio, i.e. the **tangency portfolio**. The tangency portfolio, $\mathbf{w}^*$, is given by the optimal $\mathbf{w}$ of (3) except that it must be scaled so that its component sum to 1. (This scaled portfolio will not depend on p.)

Exercise 3 *Without using (4) show that the efficient frontier is indeed a straight line as described above. Hint: consider forming a portfolio of the risk-free security with* any *risky security or risky portfolio. Show that the mean and standard deviation of the portfolio varies linearly with α where α is the weight on the risk-free-security. The conclusion should now be clear.*

Exercise 4 *Describe the efficient frontier if no borrowing is allowed.*

The **Sharpe ratio** of a portfolio (or security) is the ratio of the expected excess return of the portfolio to the portfolio's volatility. The **Sharpe optimal portfolio** is the portfolio with maximum Sharpe ratio. It is straightforward to see in our mean-variance framework (with a risk-free security) that the tangency portfolio, $\mathbf{w}^*$, is the Sharpe optimal portfolio.

1.3 Including Portfolio Constraints

We can easily include linear portfolio constraints in the problem formulation and still easily solve the resulting quadratic program. No-borrowing or no short-sales constraints are examples of linear constraints as are leverage and sector constraints. While analytic solutions are generally no longer available, the resulting problems are still easy to solve numerically. In particular, we can still determine the efficient frontier.

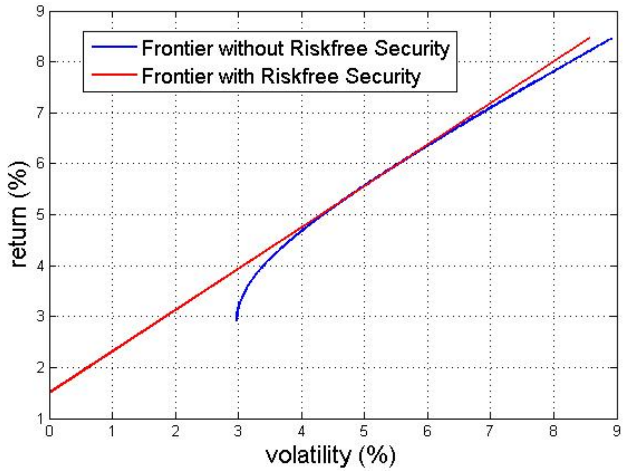

Figure 2: The Efficient Frontier with a Riskfree Security.

1.4 Weaknesses of Traditional Mean-Variance Analysis

The traditional mean-variance analysis of Markowitz has many weaknesses when applied naively in practice. They include:

1. The tendency to produce extreme portfolios combining extreme shorts with extreme longs. As a result, portfolio managers generally do not trust these extreme weights. This problem is typically caused by estimation errors in the mean return vector and covariance matrix.

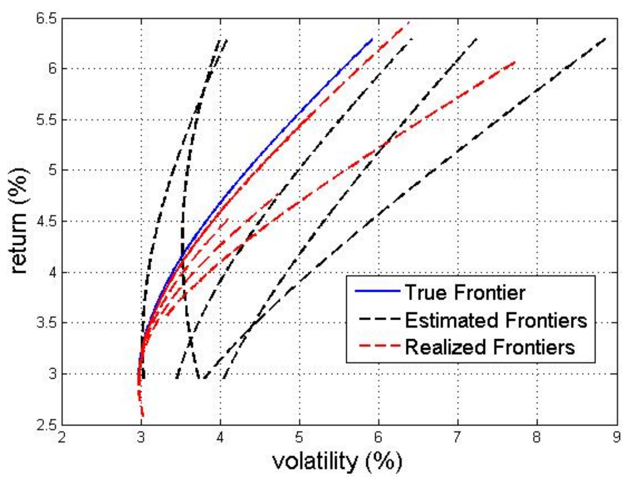

Figure 3: The Efficient Frontier, Estimated Frontiers and Realized Frontiers.

Consider Figure 3, for example, where we have plotted the same efficient frontier (of risky securities) as in Figure 2. In practice, investors can never compute this frontier since they do not know the true mean vector and covariance matrix of returns. The best we can hope to do is to approximate it. But how might we do this? One approach would be to simply estimate the mean vector and covariance matrix using historical data. Each of the black dashed curves in Figure 3 is an *estimated* frontier that we computed by: (i) simulating $m = 24$ sample returns from the true (in this case, multivariate normal) distribution (ii) estimating the mean vector and covariance matrix from this simulated data and (iii) using these estimates to generate the (estimated) frontier. Note that the blue curve in Figure 3 is the true frontier computed using the true mean vector and covariance matrix.

The first observation is that the estimated frontiers are quite random and can differ greatly from the true frontier. They may lie below or above the true frontier or they may cross it and an investor who uses such an estimated frontier to make investment decisions may end up choosing a poor portfolio. How poor? The dashed red curves in Figure 3 are the *realized* frontiers that depict the true portfolio mean - volatility tradeoff that results from making decisions based on the estimated frontiers. In contrast to the estimated frontiers, the realized frontiers must always (why?) lie below the true frontier. In Figure 3 some of the realized frontiers lie very close to the true frontier and so in these cases an investor would do very well. But in other cases the realized frontier is far from the (generally unobtainable) true efficient frontier.

These examples serve to highlight the importance of estimation errors in any asset allocation procedure. Note also that if we had assumed a heavy-tailed distribution for the true distribution of portfolio returns then we might expect to see an even greater variety of sample mean-standard deviation frontiers. In addition, it is worth emphasizing that in practice we may not have as many as 24 *relevant* observations available. For example, if our data observations are weekly returns, then using 24 of them to estimate the joint distribution of returns is hardly a good idea since we are generally more concerned with estimating *conditional* return distributions and so more weight should be given to more recent returns. A more sophisticated estimation approach should therefore be used in practice. More generally, it must be stated that estimating expected returns using historical data is very problematic and is not advisable!

2. The portfolio weights tend to be extremely sensitive to very small changes in the expected returns. For example, even a small increase in the expected return of just one asset can dramatically alter the optimal composition of the entire portfolio. Indeed let $\mathbf{w}$ and $\widehat{\mathbf{w}}$ denote the true optimal and estimated optimal portfolios, respectively, corresponding to the true mean return vector, $\boldsymbol{\mu}$, and the sample mean return vector, $\widehat{\boldsymbol{\mu}}$, respectively. Then Best and Grauer (1991[1]) showed that

$$||\mathbf{w} - \widehat{\mathbf{w}}|| \ \leq \ \xi \, ||\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}|| \, \frac{1}{\gamma_{min}} \left(1 + \frac{\gamma_{max}}{\gamma_{min}}\right)$$

   where γ_{max} and γ_{min} are the largest and smallest eigen values, respectively, of the covariance matrix, $\boldsymbol{\Sigma}$. Therefore the sensitivity of the portfolio weights to errors in the mean return vector grows as the ratio $\gamma_{max}/\gamma_{min}$ grows. But this ratio, when applied to the estimated covariance matrix, $\widehat{\boldsymbol{\Sigma}}$, typically becomes large as the number of asset increases and the number of sample observations is held fixed. As a result, we can expect large errors for large portfolios with relatively few observations.

3. While it is commonly believed that errors in the estimated means are of much greater significance, errors in estimated covariance matrices can also have considerable impact. While it is generally easier to estimate covariances than means, the presence of heavy tails in the return distributions can result in significant errors in covariance estimates as well. These problems can be mitigated to varying extents through the use of more robust estimation techniques.

As a result of these weaknesses, portfolio managers traditionally have had little confidence in mean-variance analysis and therefore applied it very rarely in practice. Efforts to overcome these problems include the use of better estimation techniques such as the use of shrinkage estimators, robust estimators and Bayesian techniques such as the Black-Litterman framework introduced in the early 1990's. (In addition to mitigating the problem of extreme portfolios, the Black-Litterman framework allows users to specify their own subjective views on the market in a consistent and tractable manner.) Many of these techniques are now used routinely in general asset allocation settings. It is worth mentioning that the problem of extreme portfolios can also be mitigated in part by placing no short-sales and / or no-borrowing constraints on the portfolio.
In Figure 4 above we have shown an estimated frontier that was computed using a more robust estimation procedure. We see that it lies much closer to the true frontier which is also the case with it's corresponding realized frontier.

1.5 The Capital Asset Pricing Model (CAPM)

If every investor is a mean-variance optimizer then we can see from Figure 2 and our earlier discussion that each of them will hold the same tangency portfolio of risky securities in conjunction with a position in the risk-free

[1] Sensitivity Analysis for Mean-Variance Portfolio Problems, *Management Science*, 37 (August), 980-989.

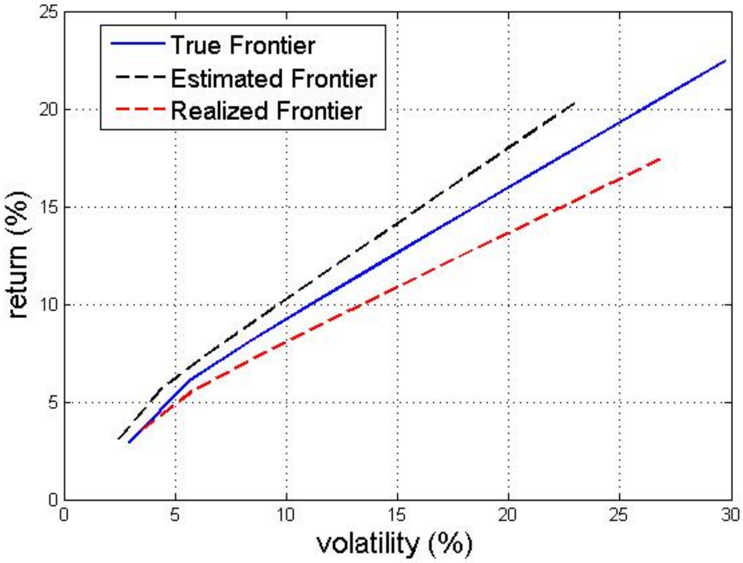

Figure 4: Robust Estimation of the Efficient Frontier.

asset. Because the tangency portfolio is held by all investors and because markets must clear, we can identify this portfolio as the *market* portfolio. The efficient frontier is then termed the *capital market line*.

Now let R_m and $\bar{R}_m$ denote the return and expected return, respectively, of the market, i.e. tangency, portfolio. The central insight of the *Capital Asset-Pricing Model* is that in equilibrium the riskiness of an asset is not measured by the standard deviation of its return but by its *beta*. In particular, there is a linear relationship between the expected return, $\bar{R} = \mathrm{E}[R]$ say, of any security (or portfolio) and the expected return of the market portfolio. It is given by

$$\bar{R} \;=\; r_f \;+\; \beta\,(\bar{R}_m - r_f) \tag{5}$$

where $\beta := \mathrm{Cov}(R, R_m)/\mathrm{Var}(R_m)$. In order to prove (5), consider a portfolio with weights α and weight $1-\alpha$ on the risky security and market portfolio, respectively. Let R_α denote the (random) return of this portfolio as a function of α. We then have

$$\begin{aligned} \mathrm{E}\left[R_\alpha\right] &= \alpha\bar{R} + (1-\alpha)\bar{R}_m && (6)\\ \sigma^2_{R_\alpha} &= \alpha^2\sigma^2_R + (1-\alpha)^2\sigma^2_{R_m} + 2\alpha(1-\alpha)\sigma_{R,R_m} && (7) \end{aligned}$$

where $\sigma^2_{R_\alpha}$, σ^2_R and $\sigma^2_{R_m}$ are the returns variances of the portfolio, security and market portfolio, respectively. We use σ_{R,R_m} to denote $\mathrm{Cov}(R, R_m)$. Now note that as α varies, the mean and standard deviation, $(\mathrm{E}\left[R_\alpha\right], \sigma^2_{R_\alpha})$, trace out a curve that cannot (why?) cross the efficient frontier. This curve is depicted as the dashed curve in Figure 5 below. Therefore at $\alpha = 0$ this curve must be tangent to the capital market line. Therefore the slope of the curve at $\alpha = 0$ must equal the slope of the capital market line. Using (6) and (7) we see the former slope is given by

$$\begin{aligned} \left.\frac{d\,\mathrm{E}[R_\alpha]}{d\,\sigma_{R_\alpha}}\right|_{\alpha=0} &= \left.\frac{d\,\mathrm{E}[R_\alpha]}{d\,\alpha}\bigg/\frac{d\,\sigma_{R_\alpha}}{d\,\alpha}\right|_{\alpha=0}\\ &= \left.\frac{\sigma_{R_\alpha}\left(\bar{R}-\bar{R}_m\right)}{\alpha\sigma^2_R - (1-\alpha)\sigma^2_{R_m} + (1-2\alpha)\sigma_{R,R_m}}\right|_{\alpha=0}\\ &= \frac{\sigma_{R_m}\left(\bar{R}-\bar{R}_m\right)}{-\sigma^2_{R_m} + \sigma_{R,R_m}}. \end{aligned}$$

The slope of the capital market line is $\left(\bar{R}_m - r_f\right)/\sigma_{R_m}$ and equating the two therefore yields

$$\frac{\sigma_{R_m}\left(\bar{R}-\bar{R}_m\right)}{-\sigma^2_{R_m} + \sigma_{R,R_m}} = \frac{\bar{R}_m - r_f}{\sigma_{R_m}}$$

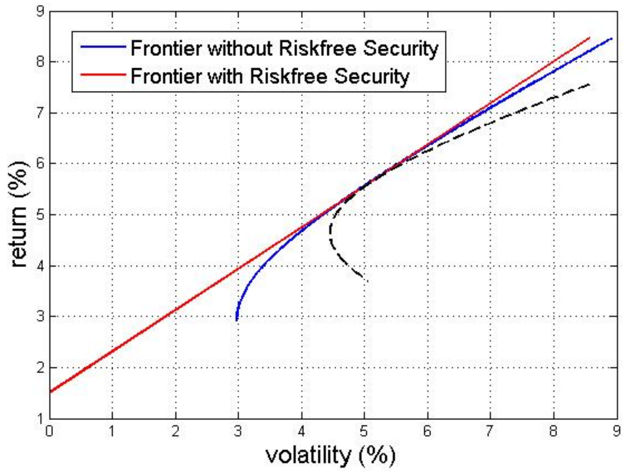

Figure 5: Proving the CAPM relationship.

which upon simplification gives (5).

The CAPM result is one of the most famous results in all of finance and, even though it arises from a simple one-period model, it provides considerable insight to the problem of asset-pricing. For example, it is well-known that riskier securities should have higher expected returns in order to compensate investors for holding them. But how do we measure risk? Counter to the prevailing wisdom at the time the CAPM was developed, the riskiness of a security is not measured by its return volatility. Instead it is measured by its beta which is proportional to its covariance with the market portfolio. This is a very important insight. Nor, it should be noted, does this contradict the mean-variance formulation of Markowitz where investors do care about return variance. Indeed, we derived the CAPM from mean-variance analysis!

Exercise 5 *Why does the CAPM result not contradict the mean-variance problem formulation where investors do measure a portfolio's risk by its variance?*

The CAPM is an example of a so-called 1-factor model with the market return playing the role of the single factor. Other factor models can have more than one factor. For example, the Fama-French model has three factors, one of which is the market return. Many empirical research papers have been written to test the CAPM. Such papers usually perform regressions of the form

$$R_i - r_f \;=\; \alpha_i + \beta_i\,(R_m - r_f) + \epsilon_i$$

where α_i (not to be confused with the α we used in the proof of (5)) is the intercept and ϵ_i is the *idiosyncratic* or residual risk which is assumed to be independent of R_m and the idiosyncratic risk of other securities. If the CAPM holds then we should be able to reject the hypothesis that $\alpha_i \neq 0$. The evidence in favor of the CAPM is mixed. But the language inspired by the CAPM is now found throughout finance. For example, we use β's to denote factor loadings and α's to denote excess returns even in non-CAPM settings.

2 The Black-Litterman Framework

We now discuss Black and Litterman's[2] framework which was developed in part to ameliorate the problems with mean-variance analysis outlined in Section 1.4 but also to allow investors to impose their own subjective views on the market-place. It has been very influential in the asset-allocation world.

To begin, we now assume the $n \times 1$ vector of *excess*[3] returns, $\mathbf{X}$, is multivariate normal *conditional* on knowing the mean excess return vector, $\boldsymbol{\mu}$. In particular, we assume

$$\mathbf{X} \mid \boldsymbol{\mu} \ \sim \ \mathsf{MVN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where the $n \times n$ covariance matrix, $\boldsymbol{\Sigma}$, is again assumed to be known. In contrast to the mean-variance approach of Section 1, however, we now assume that $\boldsymbol{\mu}$ is also random and satisfies

$$\boldsymbol{\mu} \ \sim \ \mathsf{MVN}_n(\boldsymbol{\pi}, \mathbf{C})$$

where $\boldsymbol{\pi}$ is an $n \times 1$ vector and $\mathbf{C}$ is an $n \times n$ matrix, both of which are assumed to be known. In practice, it is common[4] to take

$$\mathbf{C} = \tau \boldsymbol{\Sigma}$$

where τ is a given subjective constant that we use to quantify the level of certainty we possess regarding the true value of $\boldsymbol{\mu}$. In order to specify $\boldsymbol{\pi}$, Black-Litterman invoked the CAPM and set

$$\boldsymbol{\pi} = \lambda \boldsymbol{\Sigma} \, \mathbf{w}_m \tag{8}$$

where $\mathbf{w}_m$ is some observable *market* portfolio and where λ measures the average level of risk aversion in the market. Note that (8) can be justified by identifying λ with $1/\xi$ in (3). It is worth emphasizing, however, that we are free to choose $\boldsymbol{\pi}$ in any manner we choose and that we are *not* required to select $\boldsymbol{\pi}$ according to the CAPM and (8). For example, one could simply assume that the components of $\boldsymbol{\pi}$ are all equal and represent some average expected level of return in the market.

Remark 1 *It is also worth noting that setting $\boldsymbol{\pi}$ according to (8) can lead to unsatisfactory values for $\boldsymbol{\pi}$. For example, Meucci (2008) considers the simplified case where the available assets are the national stock market indices of Italy, Spain, Switzerland, Canada, the U.S. and Germany. In this case he takes the market portfolio to be $\mathbf{w}_m = (4\%, 4\%, 5\%, 8\%, 71\%, 8\%)$. Note in particular that the US index represents 71% of the "market". Having estimated the covariance matrix, $\boldsymbol{\Sigma}$, he obtains $\boldsymbol{\pi} = (6\%, 7\%, 9\%, 8\%, 17\%, 10\%)$ using (8). But there is no reason to assume that the market expects the U.S. to out-perform the other indices by anywhere from 5% to 11%! Choosing $\boldsymbol{\pi}$ according to (8) is clearly not a good idea in this situation.*

Exercise 6 *Can you provide some reasons for why (8) leads to a poor choice of $\boldsymbol{\pi}$ in this case?*

Subjective Views

The Black-Litterman framework allows the user to express subjective views on the vector of mean excess returns, $\boldsymbol{\mu}$. This is achieved by specifying a $k \times n$ matrix $\mathbf{P}$, a $k \times k$ covariance matrix $\boldsymbol{\Omega}$ and then defining

$$\mathbf{V} := \mathbf{P}\boldsymbol{\mu} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim \mathsf{MVN}_k(\mathbf{0}, \boldsymbol{\Omega})$ independently of $\boldsymbol{\mu}$. In practice it is common to set $\boldsymbol{\Omega} = \mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^\top / c$ for some scalar, $c > 0$, that represents the level of confidence we have in our views. Specific views are then expressed by

[2] Asset Allocation: Combining Investor Views with Market Equilibrium, (1990) *Goldman Sachs Fixed Income Research*.

[3] $\mathbf{X} = \mathbf{R} - r_f$ in the notation of the previous section. We follow the notation of Meucci's 2008 paper The Black-Litterman Approach: Original Model and Extensions, in this section.

[4] This is what Black-Litterman did in their original paper. They also set $\lambda = 2.4$ in (8) in their original paper but obviously other values could also be justified.

conditioning on $\mathbf{V} = \mathbf{v}$ where $\mathbf{v}$ is a given $k \times 1$ vector. Note also that we can only express *linear* views on the expected excess returns. These views can be *absolute* or *relative*, however. For example, in the international portfolio setting of Remark 1, stating that the Italian index is expected to rise by 10% would represent an absolute view. Stating that the U.S. will outperform Germany by 12% would represent a relative view. Both of these views could be expressed by setting

$$P \;=\; \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

and $\mathbf{v} = (10\%, 12\%)'$. If the i^{th} view, v_i, is more *qualitative* in nature, then one possibility would be to express this view as

$$v_i \;=\; \mathbf{P}_{i.}\boldsymbol{\pi} \;+\; \eta\,\sqrt{(\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^\top)_{i,i}} \tag{9}$$

for some $\eta \in \{-\beta, -\alpha, \alpha, \beta\}$ representing "very bearish", "bearish", "bullish" and "very bullish", respectively. Meucci (2008) recommends taking $\alpha = 1$ and $\beta = 2$ though of course the user is free to set these values as she sees fit. Note that $\mathbf{P}_{i.}$ appearing in (9) denotes the i^{th} row of $\mathbf{P}$.

2.1 The Posterior Distribution

The goal now is to compute the conditional distribution of $\boldsymbol{\mu}$ given the views, $\mathbf{v}$. Conceptually, we assume that $\mathbf{V} =$ has been observed equal to $\mathbf{v}$ and we now want to determine the conditional or posterior distribution of $\boldsymbol{\mu}$. This posterior distribution is given by Proposition 1 which follows from a straightforward application of Baye's rule.

Proposition 1 *The conditional distribution of $\boldsymbol{\mu}$ given $\mathbf{v}$ is multivariate normal. In particular, we have*

$$\boldsymbol{\mu} \mid \mathbf{v} \;\sim\; \mathit{MVN}_n(\boldsymbol{\mu}_{bl}, \boldsymbol{\Sigma}_{bl}) \tag{10}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{bl} \;&:=\; \boldsymbol{\pi} \;+\; \mathbf{C}\,\mathbf{P}^\top\,(\mathbf{P}\mathbf{C}\,\mathbf{P}^\top \;+\; \boldsymbol{\Omega})^{-1}\,(\mathbf{v} - \mathbf{P}\boldsymbol{\pi}) \\ \boldsymbol{\Sigma}_{bl} \;&:=\; \mathbf{C} \;-\; \mathbf{C}\,\mathbf{P}^\top(\mathbf{P}\mathbf{C}\,\mathbf{P}^\top \;+\; \boldsymbol{\Omega})^{-1}\,\mathbf{P}\mathbf{C}. \end{aligned}$$

Sketch of Proof: Since $\mathbf{V}|\boldsymbol{\mu} \sim \mathsf{MVN}(\mathbf{P}\boldsymbol{\mu}, \boldsymbol{\Omega})$, the posterior density of $\boldsymbol{\mu}|\mathbf{V} = \mathbf{v}$ satisfies

$$\begin{aligned} f(\boldsymbol{\mu}|\mathbf{v}) \;&\propto\; f(\mathbf{v}|\boldsymbol{\mu})\, f(\boldsymbol{\mu}) \\ &\propto\; \exp\left(-\frac{1}{2}(\mathbf{v} - \mathbf{P}\boldsymbol{\mu})^\top\boldsymbol{\Omega}^{-1}(\mathbf{v} - \mathbf{P}\boldsymbol{\mu})\right)\, \exp\left(-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\pi})^\top\mathbf{C}^{-1}(\boldsymbol{\mu} - \boldsymbol{\pi})\right). \end{aligned}$$

After some tedious algebraic manipulations, we obtain

$$f(\boldsymbol{\mu}|\mathbf{v}) \;\propto\; \exp\left(-\frac{1}{2}\,\boldsymbol{\mu}^\top\mathbf{A}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^\top\;\left(\mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{v} + \mathbf{C}^{-1}\boldsymbol{\pi}\right)\right) \tag{11}$$

where $\mathbf{A}^{-1} := \mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1}$. By completing the square in the exponent in (11) we find that

$$\boldsymbol{\mu} \mid \mathbf{v} \;\sim\; \mathsf{MVN}_n\left(\mathbf{A}\left(\mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{v} + \mathbf{C}^{-1}\boldsymbol{\pi}\right),\; \mathbf{A}\right). \tag{12}$$

A result from linear algebra, which can be confirmed directly, states that

$$\left(\mathbf{P}^\top\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1}\right)^{-1} = \mathbf{C} - \mathbf{C}\mathbf{P}^\top\left(\mathbf{P}\mathbf{C}\mathbf{P}^\top + \boldsymbol{\Omega}\right)^{-1}\mathbf{P}\mathbf{C}$$

and this result, in conjunction with (12), can now be used to verify the statements in the proposition. □

Of course what we really care about is the conditional distribution of $\mathbf{X}$ given $\mathbf{V} = \mathbf{v}$. Indeed using the results of Proposition 1 it is easy to show that

$$\mathbf{X} \mid \mathbf{V} = \mathbf{v} \sim \mathsf{MVN}(\boldsymbol{\mu}_{bl}, \boldsymbol{\Sigma}^x_{bl}) \tag{13}$$

where $\boldsymbol{\Sigma}^x_{bl} := \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_{bl}$.

Exercise 7 *Show (13).*

Once we have expressed our views the posterior distribution can be computed and then used in any asset allocation setting including, for example, the mean-variance setting of Section 1 or the mean-CVaR setting of Section 4.

Exercise 8 *Suppose we choose $\boldsymbol{\pi}$ in accordance with (8) and then solve a mean-variance portfolio optimization problem using the posterior distribution of $\mathbf{X}$. Explain why we are much less likely in general to obtain the extreme corner solutions that are often obtained in the traditional implementation of the mean-variance framework where historical returns are used to estimate expected returns and covariances.*

Exercise 9 *What happens to the posterior distribution of $\mathbf{X}$ as[5] $\boldsymbol{\Omega} \to \infty$? Does this make sense?*

2.2 An Extension of Black-Litterman: Views on Risk Factors

In the Black-Litterman framework, views are expressed on $\boldsymbol{\mu}$, the vector of mean excess returns. It may be more meaningful, however, to express views directly on the market outcome, $\mathbf{X}$. (It would certainly be more natural for a portfolio manager to express views regarding $\mathbf{X}$ rather than $\boldsymbol{\mu}$.) It is straightforward to model this situation and indeed we can view $\mathbf{X}$ more generally as denoting a vector of changes in *risk factors*. As we shall see, we can allow these risk factors to influence the portfolio return in either a linear or non-linear manner. We no longer model $\boldsymbol{\mu}$ as a random vector but instead assume $\boldsymbol{\mu} \equiv \boldsymbol{\pi}$ so that the reference model becomes

$$\mathbf{X} \ \sim \ \mathsf{MVN}_n(\boldsymbol{\pi}, \boldsymbol{\Sigma}) \tag{14}$$

We express our (linear) views on the market via a matrix, $\mathbf{P}$, and assume that

$$\mathbf{V}|\mathbf{X} \ \sim \ \mathsf{MVN}_k(\mathbf{P}\mathbf{X}, \boldsymbol{\Omega}) \tag{15}$$

where again $\boldsymbol{\Omega}$ represents our uncertainty in the views. Letting $\mathbf{v}$ denote our realization of $\mathbf{V}$, i.e. our view, we can compute the posterior distribution of $\mathbf{X}$. An application of Baye's Theorem leads to

$$\mathbf{X}|\mathbf{V} = \mathbf{v} \ \sim \ \mathsf{MVN}_n(\boldsymbol{\mu}_{mar}, \boldsymbol{\Sigma}_{mar}) \tag{16}$$

where

$$\boldsymbol{\mu}_{mar} \ := \ \boldsymbol{\pi} \ + \ \boldsymbol{\Sigma}\,\mathbf{P}^\top \left(\mathbf{P}\boldsymbol{\Sigma}\,\mathbf{P}^\top \ + \ \boldsymbol{\Omega}\right)^{-1} (\mathbf{v} - \mathbf{P}\boldsymbol{\pi}) \tag{17}$$

$$\boldsymbol{\Sigma}_{mar} \ := \ \boldsymbol{\Sigma} \ - \ \boldsymbol{\Sigma}\,\mathbf{P}^\top(\mathbf{P}\boldsymbol{\Sigma}\,\mathbf{P}^\top \ + \ \boldsymbol{\Omega})^{-1}\,\mathbf{P}\boldsymbol{\Sigma}. \tag{18}$$

Note that as $\boldsymbol{\Omega} \to \infty$, we see that $\mathbf{X}|\mathbf{v} \to \mathsf{MVN}_n(\boldsymbol{\pi}, \boldsymbol{\Sigma})$, the reference model.

Scenario Analysis

Scenario analysis corresponds to the situation where $\boldsymbol{\Omega} \to \mathbf{0}$ in which case the user has complete confidence in his view. A better interpretation, however, is that the user simply wants to understand the posterior distribution when he conditions on certain outcomes or *scenarios*. In this case it is immediate from (17) and (18) that

$$\mathbf{X}|\mathbf{V} = \mathbf{v} \ \sim \ \mathsf{MVN}_n(\boldsymbol{\mu}_{scen}, \boldsymbol{\Sigma}_{scen})$$

$$\boldsymbol{\mu}_{scen} \ := \ \boldsymbol{\pi} \ + \ \boldsymbol{\Sigma}\,\mathbf{P}^\top \left(\mathbf{P}\boldsymbol{\Sigma}\,\mathbf{P}^\top\right)^{-1} (\mathbf{v} - \mathbf{P}\boldsymbol{\pi})$$

$$\boldsymbol{\Sigma}_{scen} \ := \ \boldsymbol{\Sigma} \ - \ \boldsymbol{\Sigma}\,\mathbf{P}^\top(\mathbf{P}\boldsymbol{\Sigma}\,\mathbf{P}^\top)^{-1}\,\mathbf{P}\boldsymbol{\Sigma}.$$

Exercise 10 *What happens to $\boldsymbol{\mu}_{scen}$ and $\boldsymbol{\Sigma}_{scen}$ when $\mathbf{P}$ is the identity matrix? Is this what you would expect?*

[5] If you prefer, just consider the case where $\boldsymbol{\Omega} = \mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^\top/c$ and $c \to 0$.

Example 1 (From Meucci[6] 2009)
Consider a trader of European options who has certain views on the options market and wishes to incorporate these views into his portfolio optimization problem. We assume the current time is t and that the investment horizon has length τ. Let $C(\mathbf{X}, \mathcal{I}_t)$ denote the price of a generic call[7] option at time $t+\tau$ as a function of all time t information, $\mathcal{I}_t$, and the changes, $\mathbf{X}$, in the risk factors between times t and $t+\tau$. We can then write

$$C(\mathbf{X}, \mathcal{I}_t) \; = \; C_{bs}\left(y_t e^{X_y}, \; h\left(y_t e^{X_y}, \sigma_t + X_\sigma, K, T-t\right); \; K, \; T-\tau, \; r\right) \tag{19}$$

where $C_{bs}(y, \sigma, K, T, r)$ is the Black-Scholes price of a call option on a stock with current price y, implied volatility σ, strike K, time-to-maturity T and risk-free rate r. The function $h(\cdot)$ represents a *skew* mapping and for a fixed time-to-maturity, T, it returns the implied volatility at some fixed strike K as a function of the current stock price and the current *ATM* implied volatility. For example, we could take

$$h(y, \sigma; K, T) \; := \; \sigma \; + \; a\frac{\ln(y/K)}{\sqrt{T}} \; + \; b\left(\frac{\ln(y/K)}{\sqrt{T}}\right)^2$$

where a and b are fixed scalars that depend on the underlying security and that can be fitted empirically using historical data for example. Returning to (19), we see that y_t is the current price of the underlying security and $X_y := \ln(y_{t+\tau}/y_t)$ is the log-change in the underlying security price. σ_t is the current ATM implied volatility and $X_\sigma := \sigma_{t+\tau} - \sigma_t$ is the change in the ATM implied volatility.

Suppose we consider a total of N different options on a total of $m < N$ underlying securities and $s < N$ times-to-maturity. Using w_i to denote the quantity of the i^{th} option in our portfolio we see that the profit, $\Pi_{\mathbf{w}}$, at time $t+\tau$ satisfies

$$\Pi_{\mathbf{w}} \; = \; \sum_{i=1}^{N} w_i \left(C_i(\mathbf{X}, \mathcal{I}_t) - c_{i,t}\right) \tag{20}$$

where $c_{i,t}$ denotes the current price of the i^{th} option. At this point we could use standard statistical techniques to estimate the mean and covariance matrix of

$$\mathbf{X} = (X_y^{(1)}, X_{\sigma_1}^{(1)}, \; \ldots, \; X_y^{(m)}, X_{\sigma_s}^{(m)}) \tag{21}$$

where $X_{\sigma_j}^{(i)}$ denotes the ATM implied volatility for the i^{th} security and the j^{th} time-to-maturity. We could then solve a portfolio optimization problem by maximizing the expected value of (20) subject to various constraints. For example, we may wish to impose delta-neutrality on our portfolio and this is easily modeled (how?) as a linear constraint on $\mathbf{w}$. We could also impose risk-constraints by limiting some measure of downside risk such as VaR[8] or CVaR. Alternatively we could subtract a constant times the variance or CVaR from the objective function and control our risk in this manner. Numerical methods and / or simulation techniques can then be used if necessary to solve the problem.

Including Subjective Views

Suppose now that the trader has her own subjective views on the risk factors in (21) and possibly some *other* risk factors as well. These additional risk factors do not *directly* impact the option prices but they can impact the option prices *indirectly* via their influence on the posterior distribution of $\mathbf{X}$. These other risk factors might represent macro-economic factors such as inflation or interest rates or the general state of the economy. The trader could express these views as in (15) and then calculate the posterior distribution as given by (16). She could now solve her portfolio optimization where she again maximizes $\Pi_{\mathbf{w}}$ subject to the same risk constraints but now using the posterior distribution of the risk factors instead of the reference distribution. ∎

[6] "Enhancing the Black-Litterman and Related Approaches: Views and Stress-test on Risk Factors", Working paper, 2009.
[7] To simplify the discussion we will only consider call options but note that put options are just as easily handled.
[8] We will see in Section 3, however, that using VaR to control risk in a portfolio optimization problem can be a very bad idea.

Remark 2 *In Example 1 the posterior distribution of the risk factors was assumed to be multivariate normal. This of course is consistent with the model specification of (14) and (15). However it is worth mentioning that statistical and numerical techniques are also available for other models where there is freedom to specify the multivariate probability distributions.*

---

3 Mean-VaR Portfolio Optimization

In this section we consider the situation where an investor maximizes the expected return on her portfolio subject to a constraint on the VaR of the portfolio. Given the failure of VaR to be subadditive, it is perhaps not surprising that such an exercise could result in a very unbalanced and undesirable portfolio. We demonstrate this using the setting of Example 2 of the *Risk Measures, Risk Aggregation and Capital Allocation* lecture[9] notes.

Example 2 (From *QRM* by MFE)
Consider an investor who has a budget of V that she can invest in $n = 100$ defaultable corporate bonds. The probability of a default over the next year is identical for all bonds and is equal to 2%. We assume that defaults of different bonds are independent from one another. The current price of each bond is 100 and if there is no default, a bond will pay 105 one year from now. If the bond defaults then there is no repayment. Let

$$\Lambda_V \; := \; \{\boldsymbol{\lambda} \in \mathbb{R}^{100} \; : \; \boldsymbol{\lambda} \geq \mathbf{0}, \; \sum_{i=1}^{100} 100\lambda_i \; = \; V\} \tag{22}$$

denote the set of all possible portfolios with current value V. Note that (22) implicitly rules out short-selling or the borrowing of additional cash. Let $L(\boldsymbol{\lambda})$ denote the loss on the portfolio one year from now and suppose the objective is to solve

$$\max_{\boldsymbol{\lambda} \in \Lambda_V} \; \mathrm{E}\left[-L(\boldsymbol{\lambda})\right] \; - \; \beta \, \mathsf{VaR}_\alpha \left(L(\boldsymbol{\lambda})\right) \tag{23}$$

where $\beta > 0$ is a measure of risk aversion. This problem is easily solved: $\mathrm{E}\left[L(\boldsymbol{\lambda})\right]$ is identical for every $\boldsymbol{\lambda} \in \Lambda_V$ since the individual loss distributions are identical across all bonds. Selecting the optimal portfolio therefore amounts to choosing the vector $\boldsymbol{\lambda}$ that minimizes $\mathsf{VaR}_\alpha \left(L(\boldsymbol{\lambda})\right)$. When $\alpha = .95$ we have already[10] seen that it was "better" (from the perspective of minimizing VaR) to invest all funds, i.e. V, into just one bond rather than investing equal amounts in all 100 bonds. Clearly this "better" portfolio is not a well diversified portfolio. ∎

As MFE point out, problems of the form (23) occur frequently in practice, particularly in the context of *risk adjusted performance measurement*. For example the performance of a trading desk might be measured by the ratio of profits earned to the risk capital required to support the desk's operations. If the risk capital is calculated using VaR, then the traders face a similar problem to that given in (23).

Given our earlier result on the subadditivity of VaR when the risk-factors are elliptically distributed, it should not be too surprising that we have a similar result when using the *Mean-VaR* criterion to select portfolios.

Proposition 2 (Proposition xxx in *MFE*)
Suppose $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ *with* $\mathrm{Var}(X_i) < \infty$ *for all* i *and let* $\mathcal{W} := \{\mathbf{w} \in \mathbb{R}^n \; : \; \sum_{i=1}^n w_i \; = \; 1\}$ *denote the set of possible portfolio weights. Let* V *be the current portfolio value so that* $L(\mathbf{w}) = V \sum_{i=1}^n w_i X_i$ *is the (linearized) portfolio loss and let* $\Theta := \{\mathbf{w} \in \mathcal{W} \; : \; -\mathbf{w}^\top \boldsymbol{\mu} \; = \; m\}$ *be the subset of portfolios giving expected return* m. *Then if* ϱ *is any positive homogeneous and translation invariant real-valued risk measure that depends only on the distribution of risk we have*

$$\underset{\mathbf{w} \in \Theta}{\operatorname{argmin}} \; \varrho(L(\mathbf{w})) \; = \; \underset{\mathbf{w} \in \Theta}{\operatorname{argmin}} \; \mathrm{Var}(L(\mathbf{w}))$$

where $\mathrm{Var}(\cdot)$ *denotes*[11] *variance.*

---

[9] That example was in turn taken from *Quantitative Risk Management* by McNeil, Frey and Embrechts.
[10] See Example 2 of the *Risk Measures, Risk Aggregation and Capital Allocation* lecture notes.
[11] As opposed to Value-at-Risk, VaR.

Before giving a proof of Proposition 2, we recall an earlier definition.

Definition 1 *We say two random variables V and W are of the same type if there exist scalars $a > 0$ and b such that $V \sim aW + b$ where $\sim$ denotes "equal in distribution".*

Proof: Recall that if $\mathbf{X} \sim \mathsf{E}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ then $\mathbf{X} = \mathbf{AY} + \boldsymbol{\mu}$ where $\mathbf{A} \in \mathbb{R}^{n \times k}$, $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\mathbf{Y} \sim S_k(\psi)$ is a spherical random vector. For any $\mathbf{w} \in \Theta$ the loss $L(\mathbf{w})$ can therefore be represented as

$$\begin{aligned} L(\mathbf{w}) \;=\; V\mathbf{w}^\top \mathbf{X} \;&=\; V\mathbf{w}^\top \mathbf{AY} \;+\; V\mathbf{w}^\top \boldsymbol{\mu} \\ &\sim\; V\|\mathbf{w}^\top \mathbf{A}\|\, Y_1 \;+\; V\mathbf{w}^\top \boldsymbol{\mu} \end{aligned} \tag{24}$$

where (24) follows from part 3 of Theorem 2 in the *Multivariate Distributions* lecture notes. Therefore $L(\mathbf{w})$ is of the same type for every $\mathbf{w} \in \Theta$ and it follows (why?) that $\varrho\left((L(\mathbf{w}) + mV)/\sqrt{\mathrm{Var}(L(\mathbf{w}))}\right) \;=\; k$ for all such $\mathbf{w} \in \Theta$ and where k is some constant. Positive homogeneity and translation invariance then imply

$$\varrho(L(\mathbf{w})) \;=\; k\sqrt{\mathrm{Var}(L(\mathbf{w}))} - mV. \tag{25}$$

It is clear from (25) that minimizing $\varrho(L(\mathbf{w}))$ over $\mathbf{w} \in \Theta$ is equivalent to minimizing $\mathrm{Var}(L(\mathbf{w}))$ over $\mathbf{w} \in \Theta$.
□

Note that Proposition 2 states that when risk factors are elliptically distributed then we can replace variance in the Markowitz / mean-variance approach with any positive homogeneous and translation invariant risk measure. This of course includes Value-at-Risk.

---

4 Mean-CVaR Portfolio Optimization

We now consider[12] the problem where the decision maker needs to strike a balance between CVaR and expected return. This could be formulated as either minimizing CVaR subject to a constraint on the expected return or equivalently, maximizing expected return subject to a constraint on CVaR. We begin by describing the problem setting and then state some important results regarding the optimization of CVaR.

Suppose there are a total of N securities with initial price vector, $\mathbf{p}_0$, at time $t = 0$. Let $\mathbf{p}_1$ denote the random security price vector at date $t = 1$ and let f denote[13] its probability density function. We use $\mathbf{w}$ to denote the vector of portfolio holdings that are chosen at date $t = 0$. The *loss function*[14], $l(\mathbf{w}, \mathbf{p}_1)$, then satisfies

$$l(\mathbf{w}, \mathbf{p}_1) \;=\; \mathbf{w}^\top (\mathbf{p}_0 - \mathbf{p}_1)\,.$$

Let $\Psi(\mathbf{w}, \alpha)$ denote the probability that the loss function does not exceed some threshold, α. Recalling that $\mathsf{VaR}_\beta(\mathbf{w})$ is the quantile of the loss distribution with confidence level β, we see that

$$\Psi\left(\mathbf{w}, \mathsf{VaR}_\beta(\mathbf{w})\right) \;=\; \beta.$$

We also have the familiar definition of CVaR_β as the expected portfolio loss conditional on the portfolio loss exceeding VaR_β. In particular

$$\mathsf{CVaR}_\beta(\mathbf{w}) \;=\; \frac{1}{1-\beta} \int_{l(\mathbf{w},\mathbf{p}_1) > \mathsf{VaR}_\beta(\mathbf{w})} l(\mathbf{w}, \mathbf{p}_1)\, f(\mathbf{p}_1)\, d\mathbf{p}_1. \tag{26}$$

It is difficult to optimize CVaR (with respect to $\mathbf{w}$) using (26) due to the presence of VaR_β on the right-hand-side. The key contribution of Rockafeller and Uryasev (2000) was to define the simpler function

$$F_\beta(\mathbf{w}, \alpha) \;:=\; \alpha \;+\; \frac{1}{1-\beta} \int_{l(\mathbf{w},\mathbf{p}_1) > \alpha} \left(l(\mathbf{w}, \mathbf{p}_1) - \alpha\right)\; f(\mathbf{p}_1)\, d\mathbf{p}_1 \tag{27}$$

which can be used instead of CVaR due to the following proposition.

[12]Much of our discussion follows Conditional Value-at-Risk: Optimization Algorithms and Applications, in *Financial Engineering News*, Issue 14, 2000 by S. Uryasev.

[13]We don't need to assume that $\mathbf{p}_1$ has a probability density but it makes the exposition easier.

[14]Since $\mathbf{p}_0$ is fixed, we consider the loss function to be a function of $\mathbf{w}$ and $\mathbf{p}_1$ only.

Proposition 3 *The function $F_\beta(\mathbf{w},\alpha)$ is convex with respect to α. Moreover, minimizing $F_\beta(\mathbf{w},\alpha)$ with respect to α gives $CVaR_\beta(\mathbf{w})$ and $VaR_\beta(\mathbf{w})$ is a minimum point. That is*

$$CVaR_\beta(\mathbf{w}) \;=\; F_\beta\left(\mathbf{w}, VaR_\beta(\mathbf{w})\right) \;=\; \min_{\alpha} F_\beta(\mathbf{w},\alpha).$$

Proof: See Rockafeller and Uryasev (2000[15]). □

We can understand the result in Proposition 3 as follows. Let $\mathbf{y} := (y_1,\ldots,y_N)$ denote N samples and let $y_{(i)}$ for $i=1,\ldots,N$ denote the corresponding *order statistics*. Then $\text{CVaR}_\beta(\mathbf{y}) = \left(\sum_{l=N-K+1}^{N} y_{(l)}\right)/K$ where $K := \lceil (1-\beta)N \rceil$. It is in fact easy to see that $\text{CVaR}_\beta(\mathbf{y})$ is the solution to the following linear program:

$$\max_{\mathbf{q}=(q_1,\ldots,q_N)} \sum_{i=1}^{N} q_i y_i$$

$$\text{subject to} \qquad \mathbf{1}^\top \mathbf{q} = 1$$
$$0 \le q_i \le 1/K, \quad \text{for } i=1,\ldots,N.$$

Exercise 11 *Formulate the dual of this linear program and relate it to the function F_β as defined in (27).*

Note that we can now use Proposition 3 to optimize CVaR_β over $\mathbf{w}$ and to simultaneously calculate VaR_β. This follows since

$$\min_{\mathbf{w}\in\mathcal{W}} \text{CVaR}_\beta(\mathbf{w}) \;=\; \min_{\mathbf{w}\in\mathcal{W},\alpha} F_\beta(\mathbf{w},\alpha) \tag{28}$$

where $\mathcal{W}$ is some subset of $\mathbb{R}^N$ denoting the set of feasible portfolios. For example if there is no short-selling allowed then $\mathcal{W} = \{\mathbf{w}\in\mathbb{R}^N \,:\, w_i \ge 0 \;\text{ for } i=1,\ldots N\}$. We can therefore optimize CVaR_β and compute the corresponding VaR_β by minimizing $F_\beta(\mathbf{w},\alpha)$ with respect to both variables. Moreover, because the loss function, $l(\mathbf{w},\mathbf{p}_1)$, is a linear and therefore convex function of the decision variables, $\mathbf{w}$, it can be shown that $F_\beta(\mathbf{w},\alpha)$ is also convex with respect to $\mathbf{w}$. Therefore if the constraint set $\mathcal{W}$ is convex, we can minimize CVaR_β by solving a smooth convex optimization problem. While very efficient numerical techniques are available for solving these problems, we will see below how linear programming formulations can also be used to solve approximate or discretized versions of these problems.

Note that the problem formulation in (28) includes the problem where we minimize CVaR subject to a constraint on the mean portfolio return. We will now see how to solve an approximation to this problem.

4.1 The Linear Programming Formulation

Suppose the density function f is not available to us or that it is simply difficult to work with. Suppose also, however, that we can easily generate samples, $\mathbf{p}_1^{(1)},\ldots,\mathbf{p}_1^{(J)}$ from f. Typically J would be on the order of thousands or tens of thousands. We can then approximate $F_\beta(\mathbf{w},\alpha)$ with $\tilde{F}_\beta(\mathbf{w},\alpha)$ defined as

$$\tilde{F}_\beta(\mathbf{w},\alpha) \;:=\; \alpha \;+\; v \sum_{j=1}^{J} \left(l(\mathbf{w},\mathbf{p}_1^{(j)}) - \alpha\right)^+. \tag{29}$$

where $v := 1/((1-\beta)J)$. Once again if the constraint set $\mathcal{W}$ is convex, then minimizing CVaR_β amounts to solving a *non-smooth* convex optimization problem. In our setup $l(\mathbf{w},\mathbf{p}_1^{(j)})$ is in fact linear (and therefore convex) so that we can solve it using linear programming methods. In fact we have the following formulation for minimizing CVaR_β subject to $\mathbf{w}\in\mathcal{W}$:

$$\min_{\alpha,\mathbf{w},\mathbf{z}} \alpha \;+\; v\sum_{j=1}^{J} z_j \tag{30}$$

[15] Optimization of Conditional Value-at-Risk, *Journal of Risk*, 2(3):21-41.

subject to

$$\begin{aligned} \mathbf{w} &\in \mathcal{W} && (31) \\ l(\mathbf{w}, \mathbf{p}_1^{(j)}) - \alpha &\leq z_j \quad \text{for } j = 1, \ldots, J && (32) \\ z_j &\geq 0 \quad \text{for } j = 1, \ldots, J. && (33) \end{aligned}$$

Exercise 12 *Convince yourself that the problem formulation in (30) to (33) is indeed correct.*

If the constraint $\mathcal{W}$ is linear then we have a linear programming formulation of the (approximate) problem and standard linear programming methods can be used to obtain a solution. These LP methods work well for a moderate number of scenarios, e.g. $J = 10,000$ but they become increasingly inefficient as J gets large. There are simple[16] non-smooth convex optimization techniques that can then be used to solve the problem much more efficiently.

4.2 Maximizing Expected Return Subject to CVaR Constraints

We now consider the problem of maximizing expected return subject to constraints on CVaR. In particular, suppose we want to maximize the expected gain (or equivalently, return) on the portfolio subject to m different CVaR constraints. This problem may be formulated as

$$\max_{\mathbf{w} \in \mathcal{W}} \mathrm{E}\left[-l(\mathbf{w}, \mathbf{p}_1)\right] \tag{34}$$

subject to

$$\mathrm{CVaR}_{\beta_i}(\mathbf{w}) \leq C_i \quad \text{for } i = 1, \ldots, m \tag{35}$$

where the C_i's are constants that constrain the CVaRs at different confidence levels. However, we can take advantage of Proposition 3 to instead formulate the problem as

$$\max_{\alpha_1, \ldots, \alpha_m, \mathbf{w} \in \mathcal{W}} \mathrm{E}\left[-l(\mathbf{w}, \mathbf{p}_1)\right] \tag{36}$$

subject to

$$F_{\beta_i}(\mathbf{w}, \alpha_i) \leq C_i \quad \text{for } i = 1, \ldots, m. \tag{37}$$

Exercise 13 *Explain why the first formulation of (34) and (35) can be replaced by the formulation of (36) and (37).*

Note also that if the i^{th} CVaR constraint in (37) is binding then the optimal α_i^* is equal to VaR_{β_i}.

Exercise 14 *Provide a linear programming formulation for solving an approximate version of the optimization problem of (36) and (37). (By "approximate" we are referring to the situation above where we can generate samples, $\mathbf{p}_1^{(1)}, \ldots, \mathbf{p}_1^{(J)}$ from f and use these samples to formulate an LP.)*

4.3 Beware of the Bias!

Suppose we minimize CVaR subject to some portfolio constraints by generating J scenarios and then solving the resulting discrete version of the problem as described above. Let $\mathbf{w}^*$ and CVaR^* denote the optimal portfolio holdings and the optimal objective function respectively. Suppose now we estimate the *out-of-sample* portfolio CVaR by running a Monte-Carlo simulation to generate portfolio losses using $\mathbf{w}^*$. Let $\widehat{\mathrm{CVaR}}(\mathbf{w}^*)$ be the estimated CVaR. How do you think $\widehat{\mathrm{CVaR}}(\mathbf{w}^*)$ will compare with CVaR^*? In particular, which of

$$\begin{aligned} (i) \quad \widehat{\mathrm{CVaR}}(\mathbf{w}^*) &= \mathrm{CVaR}^* \\ (ii) \quad \widehat{\mathrm{CVaR}}(\mathbf{w}^*) &> \mathrm{CVaR}^* \\ (iii) \quad \widehat{\mathrm{CVaR}}(\mathbf{w}^*) &< \mathrm{CVaR}^* \end{aligned}$$

[16] See Abad and Iyengar's Portfolio Selection with Multiple Spectral Risk Constraints, *SIAM Journal on Financial Mathematics*, Vol. 6, Issue 1, 2015.

would you expect to hold? After some consideration it should be clear that we would expect (ii) to hold. This *bias* that results from solving a simulated version of the problem can be severe. It is therefore always prudent to estimate $\widehat{\text{CVaR}}(\mathbf{w}^*)$.

Exercise 15 *Can you think of any ways to minimize the bias?*

It is also worth mentioning that some of the problems with mean-variance analysis that we discussed earlier also arise with mean-CVaR optimization. In particular, estimation errors can be significant and in fact these errors can be exacerbated when estimating tail quantities such as CVaR. It is not surprising that if we compute sample mean-CVaR frontiers then we will see similar results to the results of Figure 3. One must therefore be very careful when performing mean CVaR optimization to mitigate the impact of estimation errors[17].

---

5 Parameter Estimation

The various methods we've described all require good estimators of mean return vectors, covariance matrices etc. But without good estimators, the methodology we've developed in these notes is of no use whatsoever! This should be be clear from the mean-variance experiments we conducted in Section 1.4. We saw there that naive estimators typically perform very poorly generally due to insufficient data. Moreover, it is worth emphasizing that one cannot obtain more data simply by going further back in history. For example, data from 10 years ago, for example, will generally have no information whatsoever regarding future returns. Indeed this may well be true of return (or factor return) data from just 1 year ago.

It is also worth emphasizing, however, that it is much easier to estimate covariance matrices than mean return vectors. As a simple example, suppose the true weekly return data for a stock is IID $\text{N}(\mu, \sigma^2)$. (This of course is very unlikely.) Given weekly observations over 1 year it is easy to see that our estimate of the mean return, $\widehat{\mu}$, will depend only on the value of the stock at the beginning and end of the year. In particular, there is no value in knowing the value of the stock at intermediate time points. This is not true when it comes to estimating the variance and in this case our estimator, $\widehat{\sigma}^2$ will depend on all intermediate values of the stock. Thus we can see how estimating variances (and indeed covariances) is much easier than estimating means.

Covariance matrices can and should be estimated using robust estimators or shrinkage estimators. Examples of robust estimators include the estimator based on Kendall's τ that we discussed earlier in the course as well as the methodology (that we did not discuss) that we used to obtain the estimated frontier in Figure 4. Examples of shrinkage estimators include estimators arising from factor models or Bayesian estimators.

Depending on the time horizon of the investor, it may be very necessary to use time series methods in conjunction with these robust and / or shrinkage methods to construct good estimators. For example, we may want to use GARCH methods to construct a multivariate GARCH factor model which can then be used to estimate the distribution of the next period's return vector.

As mentioned above, it is very difficult to estimate mean returns and naive estimators based on historical data should not be used! Robust, shrinkage, and time series methods may again be useful together with subjective views based on other analysis for estimating mean vectors. As with the Black-Litterman approach, the subjective views can be imposed via a Bayesian framework. There is no easy way to get rich but common sense should help in avoiding investment catastrophes!

[17] Conditional Value-at-Risk in Portfolio Optimization, *Operations Research Letters*, 39: 163–171 by Lim, Shantikumar and Vahn (2011) perform several numerical experiments to emphasize how serious these estimation errors can be.