Lecture 6. Portfolio Theory Beyond Markowitz

Kevin Mo

Stevens Institute of Technology

steve.yang@stevens.edu

10/14/2021

Outline

Estimating the Inputs Used in Mean-Variance Optimization

- ▶ Quantitative techniques for forecasting security expected returns and risk most often rely on historical data. Therefore, it is important to keep in mind that we are implicitly assuming that the past can predict the future.

- ▶ It is well known that expected returns exhibit significant time variation (nonstationarity) and that realized returns are strongly influenced by changes in expected returns.

- ▶ In practice, if portfolio managers believe that the inputs that rely on the historical performance of an asset class are not a good reflection of the future expected performance of that asset class, they may alter the inputs objectively or subjectively.

► Given the historical returns of two securities i and j, and $R_{i,t}, R_{j,t}$, where $t = 1, ..., T$, the sample mean and covariance are given by

$$\bar{R}_i = \frac{1}{T} \sum_{t=1}^{T} R_{i,t}$$

$$\bar{R}_j = \frac{1}{T} \sum_{t=1}^{T} R_{j,t}$$

$$\sigma_{ij} = \frac{1}{T-1} \sum_{t=1}^{T} (R_{i,t} - \bar{R}_i)(R_{j,t} - \bar{R}_j)$$

In the case of N securities, the covariance matrix can be expressed directly in matrix form:

$$\Sigma = \frac{1}{N-1} \mathbf{X}\mathbf{X}'$$

where

$$\mathbf{X} = \begin{bmatrix} R_{11} & \dots & R_{1T} \\ \vdots & \ddots & \vdots \\ R_{N1} & \dots & R_{NT} \end{bmatrix} - \begin{bmatrix} \bar{R}_1 & \dots & \bar{R}_1 \\ \vdots & \ddots & \vdots \\ \bar{R}_N & \dots & \bar{R}_N \end{bmatrix}$$

Under the assumption that security returns are independent and identically distributed (iid), it can be demonstrated that Σ is the maximumlikelihood estimator of the population covariance matrix and that this matrix follows Wishart distribution with $N-1$ degrees of freedom.

▶ When using a longer history, it is common that historical security returns are first converted into excess returns, $R_{i,t} - R_{f,t}$, and thereafter the expected return is estimated using

$$\bar{R}_i = R_{f,t} + \frac{1}{T} \sum_{t=1}^{T} (R_{i,t} - R_{f,t})$$

- ▶ Unfortunately, for financial time series, the sample mean is a poor estimator for the expected return. The sample mean is the *best linear unbiased estimator* (BLUE) of the population mean for distributions that are not heavy-tailed.

  Furthermore, financial time series are typically not stationary, so the mean is not a good forecast of expected returns. The resulting estimator has large estimation error (as measured by the standard error), which significantly influences the mean-variance portfolio allocation process.

- ▶ Sample covariance matrix estimator can be improved using weighted data

$$\sigma_{ij} = \frac{1-d}{1-d^T} \sum_{t=1}^{T} d^{T-t} (R_{i,t} - \bar{R}_i)(R_{j,t} - \bar{R}_j)$$

  where T is large enough. The weighting (decay) parameter d can be estimated by ML estimation.

- The sample covariance matrix is a nonparametric (unstructured) estimator. An alternative is to make assumptions on the structure of the covariance matrix during the estimation process.

  However, structured estimators can suffer from specification error, that is, the assumptions made may be too restrictive for accurate forecasting of reality.

- More robust or stable (lower estimation error) estimates of expected return and covariances should be used. One approach is to impose more structure on the estimator. Most commonly, practitioners use some form of factor model to produce the expected return forecasts. Another possibility is to use Bayesian/Black-Litterman or shrinkage estimators.

Practical Problems Encountered in Mean-Variance Optimization

The simplicity and the intuitive appeal of portfolio construction using modern portfolio theory have attracted significant attention. Unfortunately, in real world applications there are many problems with it. The issues include:

1. Sensitivity to estimation error: large estimation errors in expected returns and/or covariances introduce errors in the optimized portfolio weights.

2. The effects of uncertainty in the inputs in the optimization process: the classic portfolio optimization problem is solved as a deterministic problem - completely ignoring the uncertainty in the inputs.

3. The large data requirement: it is simply unreasonable for the portfolio manager to produce good estimates of all the inputs required in classical portfolio theory.

Shrinkage Estimation

- It is well known since Stein's seminal work that biased estimators often yield better parameter estimates than their generally preferred unbiased counterparts. In particular, it can be shown that if we consider the problem of estimating the mean of an N-dimensional multivariate normal variable $(N > 2)$, $X \in N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with known covariance matrix $\boldsymbol{\Sigma}$, the sample mean $\hat{\boldsymbol{\mu}}$ is not the best estimator of the population mean $\boldsymbol{\mu}$ in terms of the quadratic loss function

$$L(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})$$

  For example, the so-called *James-Stein shrinkage estimator*

$$\hat{\boldsymbol{\mu}}_{JS} = (1 - w)\hat{\boldsymbol{\mu}} + w\mu_0 \boldsymbol{I}$$

  has a lower quadratic loss than the sample mean, where $w = min\left(\frac{N-2}{T(\hat{\boldsymbol{\mu}} - \mu_0 \boldsymbol{I})' \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\mu}} - \mu_0 \boldsymbol{I})} \right)$.

- The vector $\mu_0 \boldsymbol{I}$ and the weight w are referred to as the shrinkage target and the shrinkage intensity/shrinkage factor.

  In effect, shrinkage is a form of averaging different estimators. The shrinkage estimator typically consists of three components: (1) an estimator with little or no structure (like the sample mean above); (2) an estimator with a lot of structure (the shrinkage target); and (3) the shrinkage intensity.

  The *shrinkage target* is chosen with the requirements: (1) it should have only a small number of free parameters (robust and with a lot of structure). (2) it should have some of the basic properties in common with the unknown quantity.

  The *shrinkage intensity* can be chosen based on theoretical properties or simply by numerical simulation.

- The most well-known shrinkage estimator used to estimate expected returns in the financial literature is the one proposed by Jorion, where the shrinkage target is given by $\mu_g \mathbf{l}$ with

$$\mu_g = \frac{\mathbf{l}'\Sigma^{-1}\hat{\mu}}{\mathbf{l}'\Sigma^{-1}\mathbf{l}}$$

and

$$w = \frac{N+2}{N+2+T(\hat{\mu}-\mu_g\mathbf{l})'\Sigma^{-1}(\hat{\mu}-\mu_g\mathbf{l})}$$

- Several studies documented that for the mean-variance framework:

  (1) the variability in the portfolio weights from one period to the next decreases; and (2) the out-of-sample risk-adjusted performance improves significantly when using a shrinkage estimator as compared to the sample mean.

- We can also apply the shrinkage techniques for covariance matrix estimation. This involves shrinking an unstructured covariance estimator toward a more structured covariance estimator.

- In practice the single-factor model and the constant correlation model yield similar results, but the constant correlation model is much easier to implement. In the case of the constant correlation model, the shrinkage estimator for the covariance matrix takes the form

$$\hat{\Sigma}_{LW} = w\hat{\Sigma}_{CC} + (1 - w)\hat{\Sigma}$$

where $\hat{\Sigma}$ is the sample covariance matrix, and $\hat{\Sigma}_{CC}$ is the sample covariance matrix with constant correlation.

First, we decompose the sample covariance matrix according to $\hat{\Sigma} = \Lambda C \Lambda'$, where Λ is a diagonal matrix of the volatilities of returns and C is the sample correlation matrix.

▶ The covariance matrix C can be written as

$$C = \begin{bmatrix} 1 & \hat{\rho}_{12} & \dots & \hat{\rho}_{1N} \\ \hat{\rho}_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\rho}_{N-1N} \\ \hat{\rho}_{N1} & \dots & \hat{\rho}_{NN-1} & 1 \end{bmatrix}$$

Second, we replace the sample correlation matrix with the constant correlation matrix

$$C_{CC} = \begin{bmatrix} 1 & \hat{\rho} & \dots & \hat{\rho} \\ \hat{\rho} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\rho} \\ \hat{\rho} & \dots & \hat{\rho} & 1 \end{bmatrix}$$

where $\hat{\rho}$ is the average of all the sample correlations, in other words

$$\hat{\rho} = \frac{2}{(N-1)N} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \hat{\rho}_{ij}$$

The Black-Litterman Model

- In the Black-Litterman model an estimate of future expected returns is based on combining market equilibrium (e.g. the CAPM equilibrium) with an investor's views.

  The Black-Litterman expected return is a shrinkage estimator where market equilibrium is the shrinkage target and the shrinkage intensity is determined by the portfolio manager's confidence in the model inputs.

- The Black-Litterman expected return is calculated as a weighted average of the market equilibrium and the investor's views. The weights depend on (1) the volatility of each asset and its correlations with the other assets and (2) the degree of confidence in each forecast. It can be interpreted as a Bayesian model.

-

▶ In the classical mean-variance optimization framework an investor is required to provide estimates of the expected returns and covariances of all the securities in the investment universe considered. This is of course a humongous task, given the number of securities available today.

▶ Furthermore, many trading strategies used today cannot easily be turned into forecasts of expected returns and covariances. In particular, not all trading strategies produce views on *absolute* return, but rather a *relative* rankings of securities that are predicted to outperform/underperform other securities.

For example, considering two stocks, A and B, instead of the absolute view, "the one-month expected return on A and B are 1.2% and 1.7% with a standard deviation of 5% and 5.5%, respectively", while a relative view may be of the form "B will outperform A with half a percent over the next month" or simply "B will outperform A over the next month".

Step 1: Basic Assumptions and Starting Point

▶ One of the basic assumptions underlying the Black-Litterman model is that the expected return of a security should be consistent with market equilibrium unless the investor has a specific view on the security. In other words, an investor who does not have any views on the market should hold the market. Our starting point is the CAPM model:

$$E(R_i) - R_f = \beta_i(E(R_M) - R_f)$$

where $E(R_i), E(R_M), and R_f$ are the expected return on security i, the expected return on the market portfolio, and the risk-free rate, respectively. Furthermore,

$$\beta_i = \frac{cov(R_i, R_M)}{\sigma_M^2}$$

where σ_M^2 is the variance of the market portfolio.

- Let's denote by $\mathbf{w}_b = (w_{b1}, ..., w_{bN})'$ the market capitalization or benchmark weights, so that with an asset universe of N securities the return on the market can be written as

$$R_M = \sum_{j=1}^{N} w_{bj} R_j$$

Then by the CAPM, the expected excess return on asset $i, \Pi_i = E(R_i) - R_f$ becomes

$$\Pi_i = \beta_i (E(R_M) - R_f)$$
$$= \frac{cov(R_i, R_M)}{\sigma_M^2} (E(R_M) - R_f)$$
$$= \frac{E(R_M) - R_f}{\sigma_M^2} \sum_{j=1}^{N} cov(R_i, R_j) w_{bj}$$

We can also express this in matrix-vector form as $\mathbf{\Pi} = \delta \mathbf{\Sigma} \mathbf{w}$ where we define the market price of risk as

$$\delta = \frac{E(R_M) - R_f}{\sigma_M^2}$$

- The true expected returns $\boldsymbol{\mu}$ of the securities are unknown. However, we assume that our previous equilibrium model serves as a reasonable estimate of the true expected returns in the sense that

$$\boldsymbol{\Pi} = \boldsymbol{\mu} + \boldsymbol{\epsilon_\Pi}, \boldsymbol{\epsilon_\Pi} \sim N(0, \tau\Sigma)$$

  for some small parameter $\tau << 1$. We can think about $\tau\Sigma$ as our confidence in how well we can estimate the equilibrium expected returns.

- In other words, a small τ implies a high confidence in our equilibrium estimates and vice versa. According to portfolio theory, because the market portfolio is on the efficient frontier, as a consequence of the CAPM an investor will be holding a portfolio consisting of the market portfolio and a risk-free instrument earning the risk-free rate.

Step 2: Expressing an Investor's Views

▶ Formally, K views in Black-Litterman model are expressed as a K-dimensional vector q with

$$\mathbf{q} = \mathbf{P}\boldsymbol{\mu} + \boldsymbol{\epsilon_q}, \boldsymbol{\epsilon_q} \sim N(0, \Omega)$$

where $\mathbf{P}$ is a $K \times N$ matrix (explained in the following example) and Ω is a $K \times K$ matrix expressing the confidence in the views.

▶ Let us assume that the asset universe that we consider have five stocks ($N = 5$) and that an investor has the following two views: (1) Stock 1 will have a return of 1.5%. (2) Stock 3 will outperform Stock 2 by 4%. Mathematically, we express the two views together as

$$\left[\begin{array}{c} 1.5\% \\ 4\% \end{array} \right] = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \end{array} \right] \left[\begin{array}{c} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{array} \right] + \left[\begin{array}{c} \epsilon_1 \\ \epsilon_2 \end{array} \right]$$

- We also remark at this point that the error terms ϵ_1, ϵ_2 do not explicitly enter into the Black-Litterman model - but their variances do. Although in some case they are directly available as a by-product of the view or the strategy, in other cases they need to be estimated separately. For example,

$$\mathbf{\Omega} = \left[\begin{array}{cc} 1\%^2 & 0 \\ 0 & 1\%^2 \end{array} \right]$$

corresponds to a higher confidence in the views, and conversely,

$$\mathbf{\Omega} = \left[\begin{array}{cc} 5\%^2 & 0 \\ 0 & 7\%^2 \end{array} \right]$$

represents a much lower confidence in views. The off-diagonal elements of Σ are typically set to zero. The reason for this is that the error terms of the individual views are most often assumed to be independent of one another.

Step 3: Combining an Investor's Views with Market Equilibrium

► Having specified the market equilibrium and an investor's views separately, we are now ready to combine the two together. There are two different approaches that can be used to arrive to the Black-Litterman model. We will discuss a derivation called mixed estimation technique.

Let us first recall the specification of market equilibrium

$$\mathbf{\Pi} = \mu + \epsilon_{\mathbf{\Pi}}, \epsilon_{\mathbf{\Pi}} \sim N(0, \tau\Sigma)$$

and the one for the investor's views

$$\mathbf{q} = \mathbf{P}\mu + \epsilon_{\mathbf{q}}, \epsilon_{\mathbf{q}} \sim N(0, \Omega)$$

We can stack these two equations together in the form

$$\mathbf{y} = \mathbf{X}\mu + \epsilon, \epsilon \sim N(0, (V))$$

- We observe that this is just a standard linear model for the expected returns $\boldsymbol{\mu}$. Calculating the Generalized Least Squares (GLS) estimator for $\boldsymbol{\mu}$, we obtain

$$\hat{\boldsymbol{\mu}}_{BL} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

$$= \left(\begin{bmatrix} \mathbf{I} & \mathbf{P}' \end{bmatrix} \begin{bmatrix} (\tau\Sigma)^{-1} & \\ & \Omega^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{\Pi} \end{bmatrix} \right)^{-1}$$

$$\begin{bmatrix} \mathbf{I} & \mathbf{P}' \end{bmatrix} \begin{bmatrix} (\tau\Sigma)^{-1} & \\ & \Omega^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{\Pi} \\ \mathbf{q} \end{bmatrix}$$

$$= \left(\begin{bmatrix} \mathbf{I} & \mathbf{P}' \end{bmatrix} \begin{bmatrix} (\tau\Sigma)^{-1} \\ \Omega^{-1}\mathbf{P} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{P}' \end{bmatrix} \begin{bmatrix} (\tau\Sigma)^{-1}\mathbf{\Pi} \\ \Omega^{-1}\mathbf{q} \end{bmatrix}$$

$$= [(\tau\Sigma)^{-1} + \mathbf{P}'\Omega^{-1}\mathbf{P}]^{-1}[(\tau\Sigma)^{-1}\mathbf{\Pi} + \mathbf{P}'\Omega^{-1}\mathbf{q}]$$

The last line in the above formula is the Black-Litterman expected returns that blend the market equilibrium with the investor's views.

Some Remarks and Observations

▶ We see that if the investor has no views (that is, $\mathbf{q} = \mathbf{\Omega} = 0$) or the confidence in the views is zero, then the Black-Litterman expected return becomes $\hat{\boldsymbol{\mu}} = \mathbf{\Pi}$. Consequently, the investor will end up holding the market portfolio as predicted by the CAPM. Intuitively speaking, the equilibrium returns in the Black-Litterman model are used to center the optimal portfolio around the market portfolio.

▶ By using $\mathbf{q} = \mathbf{P}\mu + \epsilon_{\mathbf{q}}$, we have that the investor's views alone imply the estimate of expected returns $\hat{\boldsymbol{\mu}} = (\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{q}$. Since $\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}' = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix, we can rewrite the Black-Litterman expected returns in the form

$$\hat{\boldsymbol{\mu}}_{BL} = [(\tau\Sigma)^{-1} + \mathbf{P}'\Omega^{-1}\mathbf{P}]^{-1}[(\tau\Sigma)^{-1}\mathbf{\Pi} + \mathbf{P}'\Omega^{-1}\mathbf{P}\hat{\boldsymbol{\mu}}]$$

Now we see that the Black-Litterman expected return is a confidence weighted linear combination of market equilibrium $\mathbf{\Pi}$ and the expected return $\hat{\boldsymbol{\mu}}$ implied by the investor's views.

- ▶ The two weighting matrices are give by

$$\mathbf{w}_\Pi = [(\tau\Sigma)^{-1} + \mathbf{P}'\mathbf{\Omega}^{-1}\mathbf{P}]^{-1}(\tau\Sigma)^{-1}$$
$$\mathbf{w_q} = [(\tau\Sigma)^{-1} + \mathbf{P}'\mathbf{\Omega}^{-1}\mathbf{P}]^{-1}\mathbf{P}'\mathbf{\Omega}^{-1}\mathbf{P}$$

where

$$\mathbf{w}_\Pi + \mathbf{w_q} = \mathbf{I}$$

In particular, $(\tau\Sigma)^{-1}$ and $\mathbf{P}'\mathbf{\Omega}^{-1}\mathbf{P}$ represent the confidence we have in our estimates of the market equilibrium and the views, respectively.

- ▶ It is straightforward to show that the Black-Litterman expected returns can also be written in the form

$$\hat{\boldsymbol{\mu}}_{BL} = \mathbf{\Pi} + \tau\Sigma\mathbf{P}'(\mathbf{\Omega} + \tau\mathbf{P}\Sigma\mathbf{P}')^{-1}(\mathbf{q} - \mathbf{P}\mathbf{\Pi})$$

where we now immediately see that we tilt away from the equilibrium with a vector proportional to
$\Sigma\mathbf{P}'(\mathbf{\Omega} + \tau\mathbf{P}\Sigma\mathbf{P}')^{-1}(\mathbf{q} - \mathbf{P}\mathbf{\Pi})$.

- We also mention that the Black-Litterman model can be derived as a solution to the following optimization problem:

$$\hat{\boldsymbol{\mu}}_{BL} = \arg\min_{\mu} \left\{ (\boldsymbol{\Pi} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Pi} - \boldsymbol{\mu}) + \tau(\mathbf{q} - \mathbf{P}\boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{q} - \mathbf{P}\boldsymbol{\mu}) \right\}$$

  From this formula we see that $\hat{\boldsymbol{\mu}}_{BL}$ is chosen such that it is simultaneously as close to $\boldsymbol{\Pi}$, and $\mathbf{P}\boldsymbol{\mu}$ is as close to $\mathbf{q}$ as possible. The distances are determined by $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Omega}^{-1}$. Furthermore, the relative importance of the equilibrium versus the views is determined by τ. We also see that τ is a redundant parameter as it can be absorbed into $\boldsymbol{\Omega}$.

- It is straightforward to calculate the variance of the Black-Litterman combined estimator of the expected returns by

$$var(\hat{\boldsymbol{\mu}}_{BL}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = [(\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}'\mathbf{Q}^{-1}\mathbf{P}]^{-1}$$

- Because security returns are correlated, views on just a few assets will imply changes to the expected returns on all assets.