Quantitative Investment

S. G. Kou
Questrom School of Business
Boston University

This version Spring 2020

Chapter 3

Bayesian Analysis and An Estimation Problem with the Mean Variance Analysis

3.1 What is Probability

So far we have focused on point estimation and hypothesis testing via frequentist methods. The frequentist method is based on the following assumptions.

F1. Probability is equivalent to limiting relative frequencies, and hence are objective properties.

F2. Unknown parameters are fixed, deterministic constants.

F3. Inference procedures should always be interpreted via long run averages. For example, a 95% c.i. should has a 95% limiting coverage frequency if we repeat the same procedure many times.

However, there is a different school of statistical inference, called Bayesian inference, which is based on the following assumptions.

B1. Probability relates to degree of belief, not frequencies. With this interpretation, we can have wider applications of probability. For example, we can say that with probability 0.55 an apple from a tree did drop to the head of Isaac Newton. This statement reflects a subjective belief, not a limiting frequency.

B2. Unknown parameters are uncertain and therefore can be modeled as random variables.

B3. Inference means giving a updated prediction about the distribution of the unknown parameters.

Clearly the Bayesian approach is subjective; this attributes to the popularity of the frequentist approach, as people in general likes objective methods. However, in finance, there is a growing support of Bayesian approach, mainly because estimation in many financial problems are very hard without Bayesian approaches. For example, since it is very difficult to estimate the true unknown returns of stocks, it makes sense to model the returns as random variables, and use Bayesian approaches to give estimators of returns by combining subjective views of the traders and the empirical data.

This is the view adapted in Black-Litterman's asset allocation method, which we will cover later.

3.2 The Bayesian Method

The Bayesian inference requires two inputs.

1. A prior distribution $f(\theta)$ about the unknown parameter θ before we see the data.

2. A likelihood function, or a model, that links θ and the data $X = (X_1, ..., X_n)$. More precisely, we need to specify the conditional joint density $f(X|\theta)$.

After this, we can calculate the posterior distribution of $f(\theta|X)$ by using Bayes' formula from elementary probability. More precisely,

$$f(\theta|X) = \frac{f(\theta, X)}{f(X)} = \frac{f(X|\theta)f(\theta)}{\int f(\theta, X)d\theta} = \frac{f(X|\theta)f(\theta)}{\int f(X|\theta)f(\theta)d\theta}.$$

We can write this as

$$f(\theta|X) \propto f(X|\theta)f(\theta),$$

because the normalizing constant

$$C(X) = \int f(X|\theta)f(\theta)d\theta$$

does not depend on θ. Typical $C(X)$ can be either got analytically (by using the fact that total probability must be one), or by numerical integration.

A recent revolution in Bayesian analysis is that very often we do not need to do any calculation at all, as algorithms may be available to draw samples from $f(\theta|X)$ directly so that we can simply do Monte Carlo simulation.

3.3 Point Estimation

After getting the posterior distribution $f(\theta|X)$, we can perform the point estimation easily by using the center of $f(\theta|X)$, such as mean, mode, median of $f(\theta|X)$. We can construct a c.i. by looking at the quantiles of $f(\theta|X)$; more precisely a $1-\alpha$ c.i. can be given by (a, b) such that

$$\int_{-\infty}^{a} f(\theta|X)d\theta = \frac{\alpha}{2}, \quad \int_{b}^{\infty} f(\theta|X)d\theta = \frac{\alpha}{2}.$$

Example 1. Bernoulli distribution. Suppose X_1,..., X_n are i.i.d. Bernoulli random variables with a unknown parameter p. To estimate p, we first use a uniform prior. More precisely, the prior for p is uniform on $[0,1]$ with a density $f(p)=1,\ 0\le p\le 1$. Now the likelihood is

$$\begin{aligned} f(X_1,...,X_n|p) &= p^{X_1}(1-p)^{1-X_1}\cdot p^{X_2}(1-p)^{1-X_2}\cdots p^{X_n}(1-p)^{1-X_n} \\ &= p^{Y_n}(1-p)^{n-Y_n}, \end{aligned}$$

where

$$Y_n=\sum_{i=1}^n X_i.$$

By Bayes' formula the posterior has the form

$$f(p|X)\propto f(X|p)f(p)=p^{Y_n}(1-p)^{n-Y_n}.$$

To find out what exactly the posterior $f(p|X)$ is, we recall the density of a Beta distribution with parameters α and β is given by

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}.$$

Therefore, $f(p|X)$ is the density of a Beta distribution with parameters Y_n+1 and $n-Y_n+1$. Since the mean of a Beta distribution is $\alpha/(\alpha+\beta)$, we can estimate p by the mean of the posterior $f(p|X)$:

$$\hat p=\frac{Y_n+1}{(Y_n+1)+(n-Y_n+1)}=\frac{Y_n+1}{n+2}.$$

A 95% c.i. can be constructed by looking at the 97.5% and 2.5% quantiles of the Beta distribution with parameters Y_n+1 and $n-Y_n+1$.

This should be compared with the frequentist's MLE estimator, which is

$$\tilde p=\frac{Y_n}{n}.$$

Note that

$$\hat p=\frac{n}{n+2}\tilde p+(1-\frac{n}{n+2})\frac{1}{2}=\frac{n}{n+2}\tilde p+(1-\frac{n}{n+2})p_0,$$

where p_0 is the mean of the prior distribution, i.e. uniform on 0 and 1. In short the Bayesian estimator is a weighted average of MLE and the prior mean.

Although for large sample size n, the difference between $\hat p$ and $\tilde p$ may be quite small, for small sample size the difference is quite significant. In the extreme case that only one sample is available, i.e. $n=1$, then

$$\hat p=\left\{\begin{array}{ll}\frac{1}{3}, & \text{if } X_1=0\\ \frac{2}{3}, & \text{if } X_1=1\end{array}\right\},\quad \tilde p=\left\{\begin{array}{ll}0, & \text{if } X_1=0\\ 1, & \text{if } X_1=1\end{array}\right\}.$$

Therefore, the Bayesian estimator $\hat p$ makes more sense for small sample sizes.

Even with a large sample size, the Bayesian estimator $\hat p$ sometimes may make more sense. For example, suppose in the past 10 years or 2500 (10*250=2500) trading days, there is no financial crisis (or a terrorist attack), and we want to estimate the chance of a financial crisis (or a terrorist attack) will happen tomorrow. Then with $n = 2500$ and $Y_n = 0$, we have two estimators

$$\hat{p} = \frac{0+1}{2500+2} = \frac{1}{2502}, \quad \tilde{p} = \frac{0}{2500} = 0.$$

As we all know the financial losses attributed to a financial crisis (or a terrorist attack) could be enormous. Thus, if you are a trader and want to reserve money for the uncertainty of a financial crisis (or a terrorist attack), then clearly the Bayesian estimator $\hat{p}$ is more useful than the MLE $\tilde{p}$.

Of course, the Bayesian estimator depends on the prior distribution; different priors lead to different posterior distributions.

Example 1 (continued). Now suppose that the prior distribution is Beta with parameters α and β. Then

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1},$$

$$f(p|X) \propto f(X|p)f(p) \propto p^{Y_n}(1-p)^{n-Y_n} p^{\alpha-1}(1-p)^{\beta-1} = p^{Y_n+\alpha-1}(1-p)^{n-Y_n+\beta-1}.$$

Therefore, $f(p|X)$ has a density Beta$(Y_n+\alpha, n-Y_n+\beta)$. The Bayesian estimator becomes

$$\hat{p} = \frac{Y_n+\alpha}{(Y_n+\alpha)+(n-Y_n+\beta)} = \frac{Y_n+\alpha}{n+\alpha+\beta},$$

and

$$\hat{p} = \frac{Y_n}{n+\alpha+\beta} + \frac{\alpha}{n+\alpha+\beta} = \frac{n}{n+\alpha+\beta}\tilde{p} + \frac{\alpha+\beta}{n+\alpha+\beta}p_0,$$

where p_0 is the mean of the prior distribution, which equals to $\alpha/(\alpha+\beta)$.

Of course, the previous prior of uniform[0,1] is a special case of Beta distribution with $\alpha = \beta = 1$.

In this example both prior and posterior have Beta distribution. When this happens, the prior is said to be conjugate with respect to the likelihood. Conjugate priors are widely used in Bayesian inference due to its analytical tractability. Here is another example of conjugate priors.

Example 2. (Normal distribution with unknown μ). Let $X_1, ..., X_n$ be i.i.d. from $N(\mu, \sigma^2)$. For simplicity, assume that the variance σ^2 is known. We want to provide a Bayesian estimator of μ. Consider a prior $N(a, b^2)$. Then the posterior density is

$$\begin{aligned}
f(\mu|X) &\propto f(\mu)f(X|\mu) \\
&= \frac{1}{b\sqrt{2\pi}}\exp\left\{-\frac{(\mu-a)^2}{2b^2}\right\}\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(X_i-\mu)^2}{2\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{(\mu-a)^2}{2b^2}-\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{2\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{b^2}+\frac{1}{\sigma^2/n}\right)\mu^2+\frac{1}{2}\left(\frac{a}{b^2}+\frac{1}{\sigma^2/n}\cdot\frac{1}{n}\sum_{i=1}^{n}X_i\right)2\mu\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\frac{1}{b^2}+\frac{1}{\sigma^2/n}\right)\left[\mu^2-\frac{\left(\frac{a}{b^2}+\frac{1}{\sigma^2/n}\bar{X}\right)}{\left(\frac{1}{b^2}+\frac{1}{\sigma^2/n}\right)}2\mu\right]\right\}.
\end{aligned}$$

Letting

$$\begin{aligned}
\tau^2 &= \left(\frac{1}{b^2}+\frac{1}{\sigma^2/n}\right)^{-1} = \frac{b^2\sigma^2}{\sigma^2+nb^2}, \\
w &= \frac{1}{\sigma^2/n}\left(\frac{1}{b^2}+\frac{1}{\sigma^2/n}\right)^{-1} = \frac{nb^2}{\sigma^2+nb^2},
\end{aligned}$$

the above equation becomes

$$\begin{aligned}
f(\mu|X) &\propto \exp\left\{-\frac{1}{2\tau^2}\left[\mu^2-\left(w\bar{X}+(1-w)a\right)2\mu\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau^2}\left[\mu-\left(w\bar{X}+(1-w)a\right)\right]^2\right\}.
\end{aligned}$$

Therefore, the posterior density $f(\mu|X)$ is normal with mean $w\bar{X}+(1-w)a$ and variance τ^2, i.e.

$$\mu|X \sim N(w\bar{X}+(1-w)a,\ \tau^2).$$

Thus, the Bayesian estimator is given by the mean of the posterior density

$$\hat{\mu} = w\bar{X}+(1-w)a,$$

which is a weighted average of the MLE $\bar{X}$ and the prior mean a. The $1-\alpha$ c.i. for μ is given by

$$\hat{\mu}\pm z_{\alpha/2}\tau = \hat{\mu}\pm z_{\alpha/2}\left(\frac{1}{b^2}+\frac{1}{\sigma^2/n}\right)^{-1/2}.$$

This should be compared with the frequentist c.i.

$$\bar{X}\pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Example 3. (Normal distribution with unknown σ^2). Let $X_1, ..., X_n$ be i.i.d. from $N(0, \sigma^2)$. We want to estimate σ^2. A conjugate prior for σ^2 can be chosen as an inverse gamma distribution, IG(α, β). More precisely, if X has a gamma distribution with parameters α and β, $X \sim G(\alpha, \beta)$, then $Y = 1/X$ has an inverse gamma distribution with parameters α and β, i.e.

$$Y \sim IG(\alpha, \beta).$$

In particular, the density of Y is given by

$$f(y) = \frac{\beta^\alpha e^{-\beta/y}}{\Gamma(\alpha) y^{\alpha+1}}, \quad y > 0,$$

and

$$E[Y] = \left\{ \begin{array}{ll} \frac{\beta}{\alpha-1} & \text{if } \alpha > 1 \\ \infty, & \text{if } 0 < \alpha \leq 1 \end{array} \right\},$$

$$Var(Y) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \quad \alpha > 2.$$

The posterior density for σ^2 is

$$\begin{aligned} f(\sigma^2|X) &\propto f(\sigma^2) f(X|\sigma^2) \\ &= \frac{\beta^\alpha e^{-\beta/\sigma^2}}{\Gamma(\alpha)\sigma^{2(\alpha+1)}} \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(X_i - 0)^2}{2\sigma^2}\right\} \\ &\propto \frac{1}{\sigma^{2(\alpha+1)}\sigma^n} \exp\left\{-\frac{1}{\sigma^2}\left(\beta + \sum_{i=1}^n \frac{{X_i}^2}{2}\right)\right\} \\ &= \frac{1}{(\sigma^2)^{(\tilde{\alpha}+1)}} \exp\left\{-\frac{1}{\sigma^2}\tilde{\beta}\right\}, \end{aligned}$$

where

$$\tilde{\alpha} = \alpha + \frac{n}{2}, \; \tilde{\beta} = \left(\beta + \frac{1}{2}\sum_{i=1}^n X_i^2\right)$$

This implies that the posterior $f(\sigma^2|X)$ is again an inverse gamma density with

$$f(\sigma^2|X) \sim IG(\tilde{\alpha}, \tilde{\beta}).$$

In particular, the Bayesian estimator for σ^2 is the mean of $IG(\tilde{\alpha}, \tilde{\beta})$, which is (as long as $\tilde{\alpha} > 1$),

$$\hat{\sigma}^2 = \frac{\tilde{\beta}}{\tilde{\alpha} - 1} = \frac{\beta + \frac{1}{2}\sum_{i=1}^n X_i^2}{\alpha + \frac{n}{2} - 1} = \frac{2\beta + \sum_{i=1}^n X_i^2}{n + 2\alpha - 2}.$$

This should be compared with the frequentist MLE estimator

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n X_i^2.$$

Indeed,

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_{i=1}^n X_i^2}{n}\frac{n}{n+2\alpha-2} + \frac{2\alpha-2}{n+2\alpha-2}\frac{2\beta}{(2\alpha-2)} \\ &= \tilde{\sigma}^2\frac{n}{n+2\alpha-2} + \frac{2\alpha-2}{n+2\alpha-2}\sigma_0^2,\end{aligned}$$

where $\sigma_0^2 = \frac{\beta}{\alpha-1}$ is the expectation of the prior distribution $\mathrm{IG}(\alpha, \beta)$ for σ^2.

Remarks. 1. It can be shown that all Bayesian estimators are biased. However, bias is only part of the criterion to compare estimators. For example, one may want to compare mean square error etc. This will be done later.

2. Besides Bayesian and frequentist inference, there is another one, called empirical Bayesian inference, which means that the prior is estimated from the data. We will study empirical Bayes later.

3.4 Problems with the empirical means and covariances

Although the mean variance analysis is a major breakthrough in finance, relatively few people in practice use it directly. Some main problems are

(1) Sometimes it is hard to estimate the covariance matrix $\mathbf{C}$, although it is not a big problem if the number of assets involved is small.

(2) The result is very sensitive to the mean vector $\mathbf{R}$, which is very difficult to be estimated.

(3) The portfolio weights are sensitive to estimation errors in $\mathbf{C}$ and $\mathbf{R}$.

(4) The whole theory is based on the quadratic utility function, which itself is problematic.

The literature is quite rich on the estimation risk topic: Jobson and Korkie (1980, 1981), Frost and Savarino (1986, 1988), Jorion (1986), Michaud (1989, 2000), Best and Grauer (1991), Black and Litterman (1992), Scherer (2004); see a survey paper by Brandt (2005) and the book by Meucci (2005).

There are two main approaches to handle the difficulties, namely empirical Bayesian estimation procedures and the standard Bayesian approach to do portfolio optimization. A key difference between the two approaches is that the empirical Bayesian approaches only depend on historical data, while Bayesian approaches rely on both data and subjective views. We will study the two approaches later.

Here in the remainder of this chapter we want to focus on the difficulties of just using the empirical mean and empirical variance and covariance matrices.

Suppose that we observe a vector of N stock returns $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_n$ during n periods. The mean vector is $\mathbf{R}_{1\times N}$, and covariance is $\mathbf{C}_{N\times N}$. One way to estimate the mean vector $\mathbf{R}$ is to simply use the sample mean vector $\hat{\mathbf{R}}$,

$$\hat{\mathbf{R}} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i.$$

Now assuming that $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_n$ are i.i.d. random vector with the multivariate normal distribution. Then

$$\mathbf{S}^2 = \sum_{i=1}^{n} (\mathbf{X}_i - \hat{\mathbf{R}})^\top (\mathbf{X}_i - \hat{\mathbf{R}})$$

has a Wishart distribution with

$$E\left[\mathbf{S}^2\right] = (n-1)\mathbf{C}.$$

To estimate $\mathbf{C}$ one can use the sample covariance

$$\hat{\mathbf{C}} = \frac{1}{n-1}\mathbf{S}^2.$$

The commend in Splus for $\hat{\mathbf{C}}$ is `var`.

Recall that in the presence of the risk-free rate, the vector of optimal mean variance efficient portfolio weights is given by

$$w = \xi \cdot \mathbf{C}^{-1}(\mathbf{R} - r\mathbf{1})$$

To get an unbiased estimator for the portfolio weights we shall use an unbiased estimator

$$\hat{w} = \xi \cdot (n-N-2)\left(\mathbf{S}^2\right)^{-1}(\hat{\mathbf{R}} - r\mathbf{1}) = \xi \cdot \frac{n-N-2}{n-1} \cdot \left(\hat{\mathbf{C}}\right)^{-1}(\hat{\mathbf{R}} - r\mathbf{1}).$$

To show this estimator is unbiased we use the fact that $\mathbf{S}^2$ and $\hat{\mathbf{R}}$ are independent, and $\left(\mathbf{S}^2\right)^{-1}$ has an inverse Wishart distribution with mean

$$\frac{\mathbf{C}^{-1}}{n-N-2}.$$

Indeed,

$$\begin{aligned}
E[\hat{w}] &= \xi \cdot (n-N-2) \cdot E\left[\left(\mathbf{S}^2\right)^{-1}\right] \cdot E(\hat{\mathbf{R}} - r\mathbf{1}) \\
&= \xi \cdot (n-N-2) \cdot \frac{\mathbf{C}^{-1}}{n-N-2} \cdot (\mathbf{R} - r\mathbf{1}) \\
&= \xi \cdot \mathbf{C}^{-1} \cdot (\mathbf{R} - r\mathbf{1}) = w.
\end{aligned}$$

Kan and Zhou (2007) show that the above unbiased portfolios perform better (highest expected out-of-sample performance) respect to the one based on an unbiased estimate of the covariance matrix (i.e. without the term $\frac{n-N-2}{n-1}$), and the one based on the MLE (obtained using $\left(\frac{1}{n-1}\mathbf{S}^2\right)^{-1}$ to estimate $\mathbf{C}^{-1}$). However, it is still an open problem to compute $Var[\hat{w}]$.

3.4.1 A numerical Example

To see how this works in terms of real data. We shall look at the data set "stocks" in Splus which contains the daily returns of five stocks, AMOCO, FORD, HP, IBM, and MERCK, from July 3, 1962, to December 31, 1991 (7420 trading days).

The mean vector and sample covariance are given by the following commands

```
> colMeans(stocks)
 AMOCO FORD HP IBM MERCK
 0.0006770299 0.0005284341 0.0007547311 0.0004244345 0.0008023445
> var(stocks)
```

Next we shall compute the optimal weights.

```
>
> dim(stocks)
[1] 7420 5
#compute the optimal weights for stocks,assuming xi=0.43, risk-free
rate =0.
optimalWeights1 <- 0.43*((7420-5-2)/(7420-1)) *solve(var(stocks))

%*% (colMeans(stocks)- rep(0, 5))
> optimalWeights1
#compute the optimal weights for the money market account.
 >1-sum(optimalWeights1)
```

In summary here is the optimal weights

Amoco	83.76933%
Ford	19.50973%
HP	26.69266%
IBM	-20.06771%
MERCK	129.16454%
Money Market	-139.0686%

The optimal weights seem to be not reasonable at all! No one in the real world will use these highly variable weights.

Best and Grauer (1991) also note that the estimated optimal weights are not robust, meaning that even a small increase in the estimated mean of just one asset can make a significant changes of the optimal weights, even making the weights to change signs from positive to negative or vice versa, although the optimal portfolio's expected return and standard deviation are virtually unchanged (p. 325).

Note that we can also compute the resulting in-sample stock part annual returns.

```
#Computing the in-sample stock only part annualized log-returns
((rep(1, 7420) %*% stocks) %*% optimalWeights1)*(250/7420)
 [,1]
[1,] 0.4557169
```

In other words, the annual in-sample return from the estimated optimal portfolio, ignoring the money market account (as we assume risk-free rate is zero), is about 45.6% per year for over a 30 year time horizon, which is very good. Of course, in-sample returns do not mean very much. It is more important to look at out-of-sample performance.

3.4.2 Errors in the Mean and Optimal Weights

As we have learned before, it is very hard to estimate the drift of a geometric Brownian motion. Hence it should not be surprising that it is difficult to estimate the population mean return vector $\mathbf{R}_{1\times N}$. Now we try to understand why it is difficult to estimate the optimal weights w^*. In particular, why the estimation of w^* is so sensitive to the estimation errors in $\mathbf{R}_{1\times N}$.

Intuitvely, the elements in the variance-covariance matrix $\mathbf{C}$ are generally small numbers, such as 0.2*0.2=0.04. Thus the elements in the inverse matrix $\mathbf{C}^{-1}$ tend to be large numbers with significant variations. Therefore, when we multiply $\mathbf{C}^{-1}$ and $\mathbf{R}$, the results tend to be very volatile.

A more precise reason is given by Best and Grauer (1991), who obtain some analytical result related to the sensitivity of mean-variance efficient portfolios to changes in asset means. In particular, assuming the covariance Σ is known, if w^* is the true unknown optimal mean-variance asset allocation given $\mathbf{R}_{1\times N}$, and $\hat{w}^*$ is the estimated optimal allocation given the estimated sample average $\hat{\mathbf{R}}$, then

$$||w^* - \hat{w}^*|| \leq \xi ||\mathbf{R} - \hat{\mathbf{R}}|| \cdot \frac{1}{\lambda_{\min}} \left(1 + \frac{\lambda_{\max}}{\lambda_{\min}}\right),$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalues of the the covariance matrix $\mathbf{C}$. Here $||\cdot||$ is the Euclidean norm. Therefore, the sensitivity of the to changes in the means may increase as the ratio $\lambda_{\max}/\lambda_{\min}$ increases. The ratio becomes large as the number of assets in the portfolio becomes large and the number of observation becomes small.

In particular, if we cannot do a good job with only 5 stocks and over 7400 data points, the situation will be even more for more stocks.

3.4.3 Possible Improvements

A significant improvement over the sample mean vector and sample covariance can be achieved by using the shrinkage method from empirical Bayes. The basic idea is to shrink the mean vector and the covariance matrix toward a plausible value (e.g. overall average for all stocks, overall market mean and covariance, that of minimal variance portfolios, etc.) Of course, by using the shrinkage we introduce some bias. But hopefully it leads to small variance and thus a reduction in the overall estimation error. Alternatively, one can use Bayesian methods to add extra subjective information by put some prior distributions on mean vector and covariances, as in the Black-Litterman approach. Both approaches will be discussed later.

There is another method, which is to put constraints on the portfolio weights so that it may not be too unreasonable. Imposing constraints on the portfolio weights such as no short selling restrictions and upper bounds, may sometimes prove useful to reduce out-of-sample volatility of the investor's optimal portfolio. Of course, whenever investors constrain their choices, they should realize that, while looking for an optimal asset allocation, they are trading off risk and return.

Indeed, Jagannathan and Ma (2003) analytically investigate the role of no short selling and other constraints, such that

$$w_i \leq w_i^{up}, \quad w_i \geq 0.$$

They show that constructing a constrained minimum variance portfolio is equivalent to construct a unconstrained minimum variance portfolio from a different covariance matrix $\tilde{\mathbf{C}}$

$$\tilde{\mathbf{C}} = \mathbf{C} + \left(\boldsymbol{\delta}\mathbf{1}^\top + \mathbf{1}\boldsymbol{\delta}^\top\right) - \left(\boldsymbol{\lambda}\mathbf{1}^\top + \mathbf{1}\boldsymbol{\lambda}^\top\right),$$

where $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$ are Langrange multipliers for the upper bound and lower bound, respectively, such that

$$\lambda_i, \delta_i \geq 0,$$
$$\delta_i = 0, \text{ if } w_i < w_i^{up}; \quad \lambda_i = 0 \text{ if } w_i > 0.$$

The matrix $\tilde{\mathbf{C}}$ is symmetric and semidefinite positive, so that it can be effectively interpreted as a covariance matrix.

Therefore, imposing no shorting selling constrains is equivalent to reduce its covariance by$\lambda_i + \lambda_j$, $j = 1, ..., N$ and its variance by $2\lambda_i$. Hence, we will have smaller estimated covariance and reduce the sampling error by imposing the no-short sales constraints.

Finally in practice many people use resampling procedure, which means resampling the historical data to compute optimal weights for each resampling data set, and then compute the average optimal weights. However, there are serious pitfalls in the portfolio resampling procedure, as it has almost no decision theoretic foundation, and it is just another heuristic. In particular, as pointed out by Kan and Zhou (2007) and Michaud (1999, 2000) all re-sampled portfolios inherit the same estimation error from the original data, and averaging may not help much.

HWK Problem.

1. Let X_1,..., X_n be random samples from an exponetial distribution with rate λ. Derive the posterior distribution of the Bayesian estimator $\hat{\lambda}$, if the prior distribution of λ is Gamma (α, β) with density

$$\frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\lambda} \lambda^{\alpha-1}.$$