Assignment 11

1. **Bayesian updating (20 points)** Consider a return $R = \mu + \epsilon$ where $\epsilon \sim N(0, \sigma_\epsilon^2)$. Further, we assume we are not sure what the value for μ is. Instead we think that $\mu = \mu_0 + \nu_0$ where $\nu_0 \sim N(0, \upsilon_0^2)$. In addition, we receive a signal that $\mu + \nu_1 = \mu_1$ where $\nu_1 \sim N(0, \upsilon_1^2)$.

   (a) Using the Gaussian projection theorem, show that our posterior estimate of μ given the additional signal μ_1 is normally distributed, $N(\hat{\mu}, \hat{\upsilon})$ with posterior mean $\hat{\mu} = E[\mu \,|\, \mu + \nu_1 = \mu_1] = \mu_0 + \beta(\mu_1 - \mu_0)$ and with posterior variance $\hat{\upsilon}^2 = V[\mu \,|\, \mu + \nu_1 = \mu_1] = \upsilon_0^2 - \beta^2(\upsilon_0^2 + \upsilon_1^2)$, where β is for you to determine.

   (b) Prove that the posterior mean and variance can be rewritten as:

$$\begin{aligned} \hat{\mu} &= \frac{\frac{1}{\upsilon_0^2}\mu_0 + \frac{1}{\upsilon_1^2}\mu_1}{\frac{1}{\upsilon_0^2} + \frac{1}{\upsilon_1^2}} \\ \hat{\upsilon}^2 &= \frac{1}{\frac{1}{\upsilon_0^2} + \frac{1}{\upsilon_1^2}} \end{aligned}$$

   Interpret this formula. Note, in particular, that the prior and signal act symmetrically on the posterior distribution.

   (c) Conclude that if you have N signals of the form $\mu + \nu_i = \mu_i$ where $\nu_i \sim N(0, \upsilon_i^2)$, $\forall i = 0, \ldots, n$, and with all ν_i independent from each other, then the posterior distribution of μ is normal $N(\hat{\mu}, \hat{\upsilon}^2)$ with

$$\begin{aligned} \hat{\mu} &= \frac{\sum_{i=0}^{n} \frac{1}{\upsilon_i^2}\mu_i}{\sum_{i=0}^{n} \frac{1}{\upsilon_i^2}} \\ \hat{\upsilon}^2 &= \frac{1}{\sum_{i=0}^{n} \frac{1}{\upsilon_i^2}} \end{aligned}$$

   What happens in the limit when you get a very large number of signals, i.e., $n \to \infty$? Interpret. *(hint: use an inductive argument and your previous results.)*

Solution:

(a) The Gaussian projection theorem gives

$$\begin{aligned}
\hat{\mu} &= \mu_0 + \upsilon_0^2(\upsilon_0^2+\upsilon_1^2)^{-1}(\mu_1-\mu_0) \\
&= \mu_0 + \beta(\mu_1-\mu_0) \\
\hat{\upsilon}^2 &= \upsilon_0^2 - \upsilon_0^2(\upsilon_0^2+\upsilon_1^2)^{-1}\upsilon_0^2 \\
&= \upsilon_0^2 - \left(\frac{\upsilon_0^2}{\upsilon_0^2+\upsilon_1^2}\right)^2 \frac{\upsilon_0^2+\upsilon_1^2}{\upsilon_0^2}\upsilon_0^2 \\
&= \upsilon_0^2 - \beta^2(\upsilon_0^2+\upsilon_1^2).
\end{aligned}$$

where

$$\beta \equiv \frac{\upsilon_0^2}{\upsilon_0^2+\upsilon_1^2}.$$

(b) For the expected value, we have

$$\begin{aligned}
\hat{\mu} &= \mu_0 + \frac{\upsilon_0^2}{\upsilon_0^2+\upsilon_1^2}(\mu_1-\mu_0) \\
&= \frac{\upsilon_1^2}{\upsilon_0^2+\upsilon_1^2}\mu_0 + \frac{\upsilon_0^2}{\upsilon_0^2+\upsilon_1^2}\mu_1 \\
&= \frac{\frac{1}{\upsilon_0^2}\mu_0 + \frac{1}{\upsilon_1^2}\mu_1}{\frac{1}{\upsilon_0^2}+\frac{1}{\upsilon_1^2}}.
\end{aligned}$$

For the variance, we have

$$\begin{aligned}
\hat{\upsilon}^2 &= \upsilon_0^2 - \frac{\upsilon_0^2}{\upsilon_0^2+\upsilon_1^2}\upsilon_0^2 \\
&= \frac{\upsilon_0^2\upsilon_1^2}{\upsilon_0^2+\upsilon_1^2} \\
&= \frac{1}{\frac{1}{\upsilon_0^2}+\frac{1}{\upsilon_1^2}}.
\end{aligned}$$

This result shows that our updated μ is normally distributed with mean $\hat{\mu}$ and variance $\hat{\upsilon}^2$. Notice how the posterior mean is a weighted combination of the prior mean and the signal realization. The prior variance and the variance of the signal affect the weighting. If the variance of the signal is large, then the prior mean has considerable weight in the posterior; if the prior variance is large, the signal realization has considerable weight in the posterior. Finally, our updated variance parameter reflects our uncertainty about μ. This estimate is smaller than both the prior variance and the variance of the signal.

(c) *Derivation (induction):*

Given the *prior* mean and variance $\hat{\mu}_{n-1}$ and $\hat{v}_{n-1}^2$, suppose that your *posterior* mean and variance $\hat{\mu}_n$ and $\hat{v}_n^2$ after you observed the *signal* μ_n are given by

$$\begin{aligned}
\hat{\mu}_n &= \frac{\frac{1}{\hat{v}_{n-1}^2}\hat{\mu}_{n-1} + \frac{1}{v_n^2}\mu_n}{\frac{1}{\hat{v}_{n-1}^2} + \frac{1}{v_n^2}} \\
\hat{v}_n^2 &= \frac{1}{\frac{1}{\hat{v}_{n-1}^2} + \frac{1}{v_n^2}}
\end{aligned}$$

As we have seen above, this is true for $n = 1$. Then,

$$\begin{aligned}
\hat{v}_n^2 &= \frac{1}{\frac{1}{\hat{v}_{n-1}^2} + \frac{1}{v_n^2}} \\
&= \frac{1}{\frac{1}{\hat{v}_{n-2}^2} + \frac{1}{v_{n-1}^2} + \frac{1}{v_n^2}} \\
&= \frac{1}{\frac{1}{\hat{v}_{n-3}^2} + \frac{1}{v_{n-2}^2} + \frac{1}{v_{n-1}^2} + \frac{1}{v_n^2}} = \cdots = \frac{1}{\sum_{i=0}^n \frac{1}{v_i^2}}
\end{aligned}$$

and

$$\begin{aligned}
\hat{\mu}_n &= \frac{\frac{1}{\hat{v}_{n-1}^2}\hat{\mu}_{n-1} + \frac{1}{v_1^2}\mu_n}{\frac{1}{\hat{v}_{n-1}^2} + \frac{1}{v_n^2}} \\
&= \frac{\frac{1}{\hat{v}_{n-2}^2}\hat{\mu}_{n-2} + \frac{1}{v_{n-1}^2}\mu_{n-1} + \frac{1}{v_1^2}\mu_n}{\frac{1}{\hat{v}_{n-2}^2} + \frac{1}{v_{n-1}^2} + \frac{1}{v_n^2}} \\
&= \frac{\frac{1}{\hat{v}_{n-3}^2}\hat{\mu}_{n-3} + \frac{1}{v_{n-2}^2}\mu_{n-2} + \frac{1}{v_{n-1}^2}\mu_{n-1} + \frac{1}{v_1^2}\mu_n}{\frac{1}{\hat{v}_{n-3}^2} + \frac{1}{v_{n-2}^2} + \frac{1}{v_{n-1}^2} + \frac{1}{v_n^2}} = \cdots = \frac{\sum_{i=0}^n \frac{1}{v_i^2}\mu_i}{\sum_{i=0}^n \frac{1}{v_i^2}}.
\end{aligned}$$

Interpretation:

Let τ_i denote the *precision* of signal μ_i, i.e., the inverse of its variance. Similarly, let $\tau_0 = 1/v_0^2$ denote the prior precision, and $\hat{\tau}_n = 1/\hat{v}_n^2$. the posterior precision.

Define

$$\overline{\tau}_n = \sum_{i=1}^{n} \tau_i \qquad\qquad \overline{\mu}_n = \sum_{i=1}^{n} \frac{\tau_i}{\overline{\tau}_n} \mu_i$$

$\overline{\mu}_n$ is a weighted mean of the n signal realizations. The weight on the signal realization μ_i is greater, the more precise the i'th signal is, i.e., the lower its variance υ_i^2 or equivalently the higher its precision $\tau_i = 1/\upsilon_i^2$. With this notation, the previous results rewrite

$$\begin{aligned} \hat{\tau}_n &= \tau_0 + \overline{\tau}_n \\ \hat{\mu}_n &= \frac{\tau_0 \mu_0 + \overline{\tau}_n \overline{\mu}_n}{\tau_0 + \overline{\tau}_n}. \end{aligned}$$

The weighted mean is unbiased

$$E[\overline{\mu}_n|\mu] = \sum_{i=1}^{n} \frac{\tau_i}{\overline{\tau}_n} E[\mu_i|\mu] = \sum_{i=1}^{n} \frac{\tau_i}{\overline{\tau}_n} E[\mu + \nu_i|\mu] = \mu.$$

Moreover, as long as the precision of each signal is bounded away from zero (i.e., the variance is finite), then $\overline{\tau}_n \to \infty$ as $n \to \infty$. Thus, the weighted mean is consistent since

$$Var[\overline{\mu}_n|\mu] = \sum_{i=1}^{n} \left(\frac{\tau_i}{\overline{\tau}_n}\right)^2 Var[\mu_i|\mu] = \frac{1}{\overline{\tau}_n^2} \sum_{i=1}^{n} \tau_i^2 \upsilon_i^2 = \frac{1}{\overline{\tau}_n^2} \sum_{i=1}^{n} \tau_i = \frac{1}{\overline{\tau}_n^2} \overline{\tau}_n = \frac{1}{\overline{\tau}_n}.$$

In the limit when you get a very large number of signals, then you put all the weight on the weighted mean of the signal realizations ($\overline{\tau}_n/(\tau_0 + \overline{\tau}_n) \to 1$), and none on your prior ($\tau_0/(\tau_0 + \overline{\tau}_n) \to 0$). Furthermore, since the weighted mean is a consistent estimator of μ, the uncertainty about μ disappears.

Derivation 2 (Multivariate Gaussian projection theorem):

By the multivariate Gaussian projection theorem, if $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ is a multivariate normal vector of random variables (here X_1 is 1-dimensional and X_2 is n-dimensional) with expected return vector $M = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}$ and covariance matrix $\Omega = \begin{bmatrix} \Omega_{12} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$, then the vector of random variable X_1 conditional on $X_2 = q$

is Gaussian $(X_1|X_2 = q) \sim N(\hat{\mu}, \hat{\upsilon}_n^2)$, with

$$\begin{aligned}\hat{\mu} &= M_1 + \Omega_{12}\Omega_{22}^{-1}(q - M_2) \\ \hat{\upsilon}_n^2 &= \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}.\end{aligned}$$

To apply the theorem, denote with

$$\boldsymbol{\mu}_n = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}.$$

the vector of the n signals received by the investor, and let

$$X = \begin{bmatrix} \mu \\ \boldsymbol{\mu}_n \end{bmatrix}.$$

We have $M_1 = E[\mu] = \mu_0$ and $M_2 = E[\boldsymbol{\mu}_n] = \mu_0 \mathbf{1}_n$. Moreover, since

$$\begin{aligned}Var[\mu] &= E[\upsilon_0^2] && = \upsilon_0^2 \\ Cov[\mu, \mu_i] &= Var[\mu] + Cov[\nu_0, \nu_i] && = \upsilon_0^2 && \forall i = 1, ..., n \\ Var[\mu_i] &= Var[(\nu_0 + \nu_i)^2] && = \upsilon_0^2 + \upsilon_i^2 && \forall i = 1, ..., n \\ Cov[\mu_i, \mu_j] &= Cov[\nu_0 + \nu_i, \nu_0 + \nu_j] && = \upsilon_0^2 && \forall i, j = 1, ..., n,\ i \neq j.\end{aligned}$$

we have

$$\begin{aligned}\Omega_{11} &= Var(\mu) && = \upsilon_0^2 \\ \Omega_{12} &= Cov(\mu, \boldsymbol{\mu}_n^\top) && = \upsilon_0^2 \mathbf{1}_n^\top \\ \Omega_{21} &= \Omega_{12}^\top && = \upsilon_0^2 \mathbf{1}_n \\ \Omega_{22} &= Cov(\boldsymbol{\mu}_n, \boldsymbol{\mu}_n^\top) && = \text{diag}(\boldsymbol{\upsilon}_n^2) + \upsilon_0^2 \mathbf{1}_n \mathbf{1}_n^\top,\end{aligned}$$

where $\mathbf{1}_n$ is a $n \times 1$ vector of one, $\boldsymbol{\upsilon}_n$ is a $n \times 1$ vector with the variances of the n signals as its components, and $\text{diag}(\boldsymbol{\upsilon}_n^2)$ is a diagonal matrix with $\boldsymbol{\upsilon}_n^2$ as its diagonal elements.

Note that

$$\begin{aligned}
\mathbf{1}_n^\top \mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1}\mathbf{1}_n &= \sum_{i=1}^n \frac{1}{\upsilon_i^2} = \sum_{i=1}^n \tau_i \quad = \overline{\tau}_n \\
\mathbf{1}_n^\top \mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1}\boldsymbol{\mu}_n &= \sum_{i=1}^n \frac{\mu_i}{\upsilon_i^2} = \sum_{i=1}^n \tau_i\mu_i = \overline{\tau}_n\overline{\mu}_n
\end{aligned}$$

By the Sherman-Morrison formula,

$$\begin{aligned}
\Omega_{22}^{-1} &= \mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1} - \upsilon_0^2 \frac{\mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1}\mathbf{1}_n\mathbf{1}_n^\top \mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1}}{1+\upsilon_0^2\mathbf{1}_n^\top \mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1}\mathbf{1}_n} \\
&= \mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1} - \frac{\mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1}\mathbf{1}_n\mathbf{1}_n^\top \mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1}}{\tau_0+\overline{\tau}_n}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\hat{\upsilon}_n^2 &= \upsilon_0^2 - (\upsilon_0^2)^2\overline{\tau}_n + (\upsilon_0^2)^2\frac{\overline{\tau}_n^2}{\tau_0+\overline{\tau}_n} \\
&= \frac{1-\upsilon_0^2\overline{\tau}_n + \upsilon_0^2\overline{\tau}_n - (\upsilon_0^2)^2\overline{\tau}_n^2}{\tau_0+\overline{\tau}_n} + (\upsilon_0^2)^2\frac{\overline{\tau}_n^2}{\tau_0+\overline{\tau}_n} \\
&= \frac{1}{\tau_0+\overline{\tau}_n} \\
\hat{\mu} &= \mu_0 + \upsilon_0^2\mathbf{1}_n^\top\left(\mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1} - \frac{\mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1}\mathbf{1}_n\mathbf{1}_n^\top \mathrm{diag}(\boldsymbol{\upsilon}_n^2)^{-1}}{\tau_0+\overline{\tau}_n}\right)(\boldsymbol{\mu}_n - \mathbf{1}_n\mu_0) \\
&= \mu_0 + \upsilon_0^2\overline{\tau}_n\overline{\mu}_n - \upsilon_0^2\frac{\overline{\tau}_n^2\overline{\mu}_n}{\tau_0+\overline{\tau}_n} - \mu_0\upsilon_0^2\overline{\tau}_n + \mu_0\upsilon_0^2\frac{\overline{\tau}_n^2}{\tau_0+\overline{\tau}_n} \\
&= \frac{\tau_0\mu_0 + \overline{\tau}_n\overline{\mu}_n - \mu_0\overline{\tau}_n + \mu_0\overline{\tau}_n + \upsilon_0^2\overline{\tau}_n^2\overline{\mu}_n - \mu_0\upsilon_0^2\overline{\tau}_n^2}{\tau_0+\overline{\tau}_n} - \upsilon_0^2\frac{\overline{\tau}_n^2\overline{\mu}_n}{\tau_0+\overline{\tau}_n} + \mu_0\upsilon_0^2\frac{\overline{\tau}_n^2}{\tau_0+\overline{\tau}_n} \\
&= \frac{\tau_0\mu_0 + \overline{\tau}_n\overline{\mu}_n}{\tau_0+\overline{\tau}_n}.
\end{aligned}$$

Derivation 3 (Bayes' Law):[1]

Recall that Bayes' Theorem, expressed in terms of probability distributions, appears as:

$$f(\theta|\mathrm{data}) = \frac{f(\mathrm{data}|\theta)f(\theta)}{f(\mathrm{data})},$$

where $f(\theta|\mathrm{data})$ is the posterior distribution for the parameter θ, $f(\mathrm{data}|\theta)$ is the sampling density for the data —which is proportional to the Likelihood function, only differing by a constant that makes it a proper density function— $f(\theta)$ is the prior distribution for the parameter, and $f(\text{data})$ is the marginal probability of the data. For a continuous sample space, this marginal probability is computed as:

$$f(\theta|\text{data}) = \frac{f(\text{data}|\theta)f(\theta)}{f(\text{data})},$$

the integral of the sampling density multiplied by the prior over the sample space for θ. This quantity is sometimes called the marginal likelihood for the data and acts as a normalizing constant to make the posterior density proper. Because this denominator simply scales the posterior density to make it a proper density, and because the sampling density is proportional to the likelihood function, Bayes' Theorem for probability distributions is often stated as:

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}.$$

Let X denote the $N \times 1$ vector of signals: $X = \begin{bmatrix} \mu_1 & \cdots & \mu_n \end{bmatrix}$. Our goal is ultimately to update our knowledge of μ with the signals $\mu_1, \cdots, \mu_n$. In other words, we wish to find $f(\mu|X)$. Bayes' Theorem tells us that:

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

where $f(X|\mu)$ is the likelihood function for the current data, and $f(\mu)$ is the prior. Since the signals are normally distributed with a mean equal to μ and precision τ_i, then our likelihood function for X is:

$$f(X|\mu) \propto L(\mu|X) \prod_{i=1}^{n} \sqrt{\frac{\tau_i}{2\pi}} \exp\left\{-\frac{\tau_i(\mu_i - \mu)^2}{2}\right\}$$

Our prior distribution for μ is:

$$f(\mu) = \sqrt{\frac{\tau_0}{2\pi}} \exp\left\{-\frac{\tau_0(\mu - \mu_0)^2}{2}\right\},$$

where in this expression, μ is the random variable, with μ_0 as the prior mean, and τ_0 reflects the variation of μ around μ_0.

Our posterior is the product of the likelihood and prior, which gives us:

$$f(\mu|X) \propto \left(\prod_{i=0}^{n} \sqrt{\tau_i}\right) \exp\left\{-\sum_{i=0}^{n} \frac{\tau_i(\mu_i-\mu)^2}{2}\right\}$$

This posterior can be reexpressed as a normal distribution for μ, but it takes some algebra in order to see this. First, since the terms outside the exponential are simply normalizing constants with respect to μ, we can drop them and work with the terms inside the exponential function. Second, let's expand the quadratic components and the summations. For the sake of simplicity, I temporarily drop the exponential function in this expression:

$$\begin{aligned}
&(-1/2) \quad \left[\sum_{i=0}^{n} \frac{\mu_i^2 - 2\mu\mu_i}{\upsilon_i^2} + \mu^2 \sum_{i=0}^{n} \tau_i\right] \\
= \; &(-1/2) \quad \left(\sum_{i=0}^{n} \tau_i\right) \left[\frac{\sum_{i=0}^{n} \tau_i \mu_i^2}{\sum_{i=0}^{n} \tau_i} - 2\mu \frac{\sum_{i=0}^{n} \tau_i \mu_i}{\sum_{i=0}^{n} \tau_i} + \mu^2\right] \\
= \; &(-1/2) \quad \left[\frac{\frac{\sum_{i=0}^{n} \tau_i \mu_i^2}{\sum_{i=0}^{n} \tau_i} - 2\mu \frac{\sum_{i=0}^{n} \tau_i \mu_i}{\sum_{i=0}^{n} \tau_i} + \mu^2}{\frac{1}{\sum_{i=0}^{n} \tau_i}}\right].
\end{aligned}$$

Using this expression, any term that does not include μ can be viewed as a proportionality constant, can be factored out of the exponent, and can be dropped. Finally, all we need to do is to complete the square in μ and discard any remaining constants to obtain:

$$(-1/2) \left[\frac{\left(\mu - \frac{\sum_{i=0}^{n} \tau_i \mu_i}{\sum_{i=0}^{n} \tau_i}\right)^2}{\frac{1}{\sum_{i=0}^{n} \tau_i}}\right].$$

This result shows that our updated μ is normally distributed with mean and variance

$$\frac{\sum_{i=0}^{n} \tau_i \mu_i}{\sum_{i=0}^{n} \tau_i} \qquad\qquad \frac{1}{\sum_{i=0}^{n} \tau_i}.$$

2. **Black Litterman (40 points)**

We will replicate the results of He-Litterman (1992) to better understand how to apply the Black-Litterman formula. We are considering the optimal asset allocation to seven country equity index returns with correlation matrix given on table 1 page 21 of the lecture notes and with volatility and relative market capitalization weights given in table 2 of page 21 of the lecture notes.

(a) Assume an investor has a risk-aversion coefficient $\gamma = 3$ and no uncertainty about his estimate of the mean vector μ_0. Compute the expected return vector μ_0 that would have him hold a portfolio equal to the market portfolio with weights w_{eq} given in table 2.

(b) Assume another investor with risk-aversion $\gamma = 2.5$ views returns as $R = \mu + \epsilon$ where $\epsilon \sim N(0, \Sigma)$. He starts with a prior that $\mu \sim N(\mu_0, \tau\Sigma)$, where Σ is the empirical covariance matrix of returns. Suppose that $\tau = 0.05$. Derive his optimal portfolio w_0 and compare how it deviates from the equilibrium market weights w_{eq}.

(c) Assume that same investor obtains two additional views on the relative performance of different country returns from two different analysts. The first analyst thinks that Germany will outperform a market value weighted basket of France and UK equities by 6%. The investor's confidence in this view is $\Omega_{11} = 0.021 \times \tau$. The second analyst thinks that the canadian equity market will outperform the US market by 2% on average. The investor's confidence in that view is $\Omega_{22} = 0.017 \times \tau$. He considers both signal to be independent as he obtained them from different analysts. Using the Black-Litterman formula, derive the posterior distribution of the mean return $\mu \sim N(\bar{\mu}, \bar{\Omega})$ as a function of the prior and the views. Verify numerically that the two sets of equations for $\bar{\mu}$ and $\bar{\Omega}$ on page 11 of the lecture notes indeed give the same answers.

(d) Given his signals the investor sees returns as $R = \mu + \epsilon$ where $\epsilon \sim N(0, \Sigma)$ and $\mu \sim N(\bar{\mu}, \bar{\Omega})$. Derive his optimal unconstrained mean-variance portfolio w^*. Compare it to his prior portfolio w_0 and to the market weights w_{eq}.

(e) Show that the optimal portfolio w^* can be decomposed into the prior portfolio and an 'overlay' of view portfolios. That is we can rewrite $w^* = w_0 + \lambda_1 P_1^\top + \lambda_2 P_2^\top$ where P_i denotes the $i^t h$ row of the view portfolio matrix P. Find the view-weights λ_1, λ_2.

(f) In addition the investor has an absolute view that the Japanese stock market will outperform the equilibrium view. In particular he thinks that the Japanese market equity return will be 4.5%. His uncertainty about the view is $\Omega_{33}/\tau = 0.03$. Derive the new optimal portfolio and the weights on the three views $\lambda_1, \lambda_2, \lambda_3$. Discuss how the portfolio and the weights change as his uncertainty becomes smaller, e.g., $\Omega_{33}/\tau = 0.01$.

Solution: *Decomposition of optimal portfolio into prior portfolio and an 'overlay' of view portfolios.*
We seek constant λ and a K-vector Λ such that:

$$\begin{aligned} w^* &= \lambda w_0 + P^\top \Lambda \\ \gamma^{-1}\hat{\Sigma}^{-1}\hat{\mu} &= \lambda\gamma^{-1}(\underbrace{\Omega_0 + \Sigma}_{=\Sigma_0})^{-1}\mu_0 + P^\top\Lambda \end{aligned}$$

Substituting the definition of $\hat{\mu}$ and $\hat{\Sigma}$, we get:

$$\begin{aligned} &\gamma^{-1}\left[\Sigma + \left(\Omega_0^{-1} + P^\top\Omega^{-1}P\right)^{-1}\right]^{-1}\left(\Omega_0^{-1} + P^\top\Omega^{-1}P\right)^{-1}\left(\Omega_0^{-1}\mu_0 + P^\top\Omega^{-1}q\right) \\ &\quad = \lambda\gamma^{-1}(\Omega_0 + \Sigma)^{-1}\mu_0 + P^\top\Lambda. \end{aligned}$$

We can rearrange and simplify to obtain:

$$\begin{aligned} & \gamma^{-1}\left[\left(\Omega_0^{-1} + P^\top\Omega^{-1}P\right)\left(\Sigma + \left(\Omega_0^{-1} + P^\top\Omega^{-1}P\right)^{-1}\right)\right]^{-1}\left(\Omega_0^{-1}\mu_0 + P^\top\Omega^{-1}q\right) \\ & \qquad = \lambda\gamma^{-1}(\Omega_0 + \Sigma)^{-1}\mu_0 + P^\top\Lambda \\ \Longleftrightarrow \quad & \gamma^{-1}\left[\left(\Omega_0^{-1} + P^\top\Omega^{-1}P\right)\Sigma + I\right]^{-1}\left(\Omega_0^{-1}\mu_0 + P^\top\Omega^{-1}q\right) \\ & \qquad = \lambda\gamma^{-1}(\Omega_0 + \Sigma)^{-1}\mu_0 + P^\top\Lambda \\ \Longleftrightarrow \quad & \left(\tau^{-1}\Sigma^{-1}\mu_0 + P^\top\Omega^{-1}q\right) \\ & \qquad = \lambda(1+\tau)^{-1}\left[\left(\tau^{-1}\Sigma^{-1} + P^\top\Omega^{-1}P\right)\Sigma + I\right]\Sigma^{-1}\mu_0 \\ & \qquad + \gamma\left[\left(\tau^{-1}\Sigma^{-1} + P^\top\Omega^{-1}P\right)\Sigma + I\right]P^\top\Lambda \\ \Longleftrightarrow \quad & \left(\tau^{-1} - \lambda(1+\tau)^{-1}(1+\tau^{-1})\right)\Sigma^{-1}\mu_0 \\ & \qquad = P^\top\Omega^{-1}\left[\lambda(1+\tau)^{-1}P\mu_0 - q + \gamma P\Sigma P^\top\Lambda + \gamma(1+\tau^{-1})\Omega\Lambda\right] \end{aligned}$$

This condition is satisfied if

$$
\begin{aligned}
0 &= \tau^{-1} - \lambda(1+\tau)^{-1}(1+\tau^{-1}) \\
\mathbf{0} &= \lambda(1+\tau)^{-1}P\mu_0 - q + \gamma\left[P\Sigma P^\top + (1+\tau^{-1})\Omega\right]\Lambda.
\end{aligned}
$$

The first equation gives

$$
\begin{aligned}
0 &= -\lambda(1+\tau)^{-1}(1+\tau^{-1}) \\
0 &= (1+\tau)\tau^{-1} - \lambda(1+\tau^{-1}) \\
\lambda &= 1.
\end{aligned}
$$

The second equation gives

$$
\begin{aligned}
\gamma\left[P\Sigma P^\top + (1+\tau^{-1})\Omega\right]\Lambda &= q - (1+\tau)^{-1}P\mu_0 \\
\Lambda &= \gamma^{-1}\left[P\Sigma P^\top + \left(1+\frac{1}{\tau}\right)\Omega\right]^{-1}\left(q - \frac{1}{1+\tau}P\mu_0\right).
\end{aligned}
$$