Risk Sensitive Portfolio Optimization

Łukasz Stettner, Institute of Mathematics Polish Acad. Sci.

October 1, 2008

What is risk? Portfolio risk? - historical approach

$S_i(t)$ the price of the i-th asset at time t $(i = 1, 2, \ldots, d)$
static approach $t = 0, 1$
random rate of return of the i-th asset

$$\zeta_i := \frac{S_i(1) - S_i(0)}{S_i(0)} = \frac{S_i(1)}{S_i(0)} - 1$$

expected rate of return of the i-th asset $\mu_i := E\zeta_i$
$\mu = [\mu_1, \ldots, \mu_d]^T$ vector of the expected rates of return,
$\zeta := [\zeta_1, \ldots, \zeta_d]^T$, vector of random rates of return
Σ covariance matrix
$\Sigma = E\left\{(\zeta - \mu)(\zeta - \mu)^T\right\} = (\Sigma^{ij})$
T stands for transponse

What is risk? Portfolio risk? - historical approach (cont.)

$\theta := [\theta_1, \ldots, \theta_d]^T$ vector of portfolio strategies (at time 0): portions of capital invested in assets: θ_i the portion of capital invested in i-th asset

random portfolio rate of return $R(\theta)$ is equal to $\theta^T \zeta$ since: if x is an initial capital

$\frac{x\theta_i}{S_i(0)}$ is the number of i-th assets at time 0

$\frac{x\theta_i}{S_i(0)} S_i(1)$ is the value of the portfolio located in i-th asset at time 1,

$R(\theta) = \frac{\sum_{i=1}^d \frac{x\theta_i}{S_i(0)} S_i(1) - x}{x} =$
$\sum_{i=1}^d \frac{\theta_i}{S_i(0)} S_i(1) - 1 = \sum_{i=1}^d \frac{\theta_i}{S_i(0)} (S_i(1) - S_i(0)) = \sum_{i=1}^d \theta_i \zeta_i$

Expected value of the portfolio

$$E\left\{R(\theta)\right\} = \theta^T \mu \tag{1}$$

variance of the portfolio rate of return

$$Var(R(\theta)) = \theta^T \Sigma \theta. \tag{2}$$

Explanation:

$Var(R(\theta)) = E\left\{\left(\sum_{i=1}^{d} \theta_i(\zeta_i - \mu_i)\right)^2\right\} =$
$E\left\{(\theta^T(\zeta - \mu))((\zeta - \mu)^T\theta)\right\} = \theta^T E\left\{(\zeta - \mu)(\zeta - \mu)^T\right\}\theta$

risk measure (historical approach) variance of the portfolio

Harry Markowitz, born Chicago 1927, Nobel prize 1990

The basic concepts of portfolio theory came to me one afternoon in the library while reading John Burr Williams's Theory of Investment Value. Williams proposed that the value of a stock should equal the present value of its future dividends. Since future dividends are uncertain, I interpreted Williams's proposal to be to value a stock by its expected future dividends. But if the investor were only interested in expected values of securities, he or she would only be interested in the expected value of the portfolio; and to maximize the expected value of a portfolio one need invest only in a single security. This, I knew, was not the way investors did or should act. Investors diversify because they are concerned with risk as well as return. Variance came to mind as a measure of risk. The fact that portfolio variance depended on security covariances added to the plausibility of the approach. Since there were two criteria, risk and return, it was natural to assume that investors selected from the set of Pareto optimal risk-return combinations.

Markowitz theory - foundations

We maximize expected portfolio rate of return taking into account the risk (minimizing it?)
What is risk?
portfolio risk function $Risk(R(\theta))$
in Markowitz theory $Risk(R(\theta)) = Var(R(\theta))$
two criterion (minimax) problem
Markowitz order $\succeq$.
the strategy θ (portfolio rate of return $R(\theta)$ is better than θ' (rate of return of $R(\theta')$), we write as $\theta \succeq \theta'$ lub $R(\theta) \succeq R(\theta')$, if $E\{R(\theta)\} \geq E\{R(\theta')\}$, and $Risk(R(\theta)) \leq Risk(R(\theta'))$.

Markowitz theory - foundations (cont.)

the strategy θ is better than θ' if the rate of return of θ is greater tan that of θ' and the risk corresponding to the strategy θ is not greater than the risk corresponding to θ'

With each strategy one can associate a point on the plane R^2, $(Risk(R(\theta)), E\{R(\theta)\})$

The strategy θ is maximal, if there are no strategy θ' different than θ such that $\theta' \succeq \theta$

The maximal strategies form on the plane R^2 the set called efficient frontier

Classical Markowitz theory - analytic approach

Convex analysis problem: to minimize

$$\theta^T \Sigma \theta \tag{3}$$

under the constraints
$\theta^T \mu = \mu_p$ i $\theta^T \mathcal{J} = 1$,
where $\mathcal{J} = [1, \ldots, 1]^T$, while μ_p is the fixed portfolio expected rate of return. For fixed portfolio rate of return we minimize risk understood as portfolio variance,
Markowitz, H., Portfolio Selection Efficient Diversification of Investments, Wiley, 1959.
When the matrix Σ is nonsingular, the problem is solved using Lagrange multipliers.

Classical Markowitz theory - analytic approach (cont.)

$G(\theta, \kappa) = \theta^T \Sigma \theta + \kappa_1(\theta^T \mu - \mu_p) + \kappa_2(\theta^T \mathcal{J} - 1)$
necessary condition for optimality:
$\frac{\partial G}{\partial \theta_i} = 0 \quad \frac{\partial G}{\partial \kappa_j} = 0$ dla $j = 1, 2$
Hence
Theorem 1. If μ does not have the same coordinates and Σ is nonsingular then

$$\inf_\theta Var(R(\theta)) = B^T H^{-1} B, \tag{4}$$

where $H = A^T \Sigma^{-1} A$, $A = [\mu, \mathcal{J}]$, $B = \begin{bmatrix} \mu_p \\ 1 \end{bmatrix}$.
Proof:

$$2\Sigma\theta + A\kappa = 0 \quad A^T\theta = B \tag{5}$$

where $\kappa := [\kappa_1, \kappa_2]^T$
we solve the first equation with respect to θ:

Classical Markowitz theory - analytic approach (cont.)

$\theta = \frac{-1}{2}\Sigma^{-1}A\kappa$
From the second equation from (5) we have
$A^T\Sigma^{-1}A\kappa = -2B$
i.e.
$\kappa = -2(A^T\Sigma^{-1}A)^{-1}B := -2H^{-1}B$
Note that H is symmetric:
$H = A^T\Sigma^{-1}A$ therefore $H^T = A^T\Sigma^{-1T}A = H$ since Σ and Σ^{-1} sạ symmetric. We are now in position to calculate the portfolio variance using the obtained formulae for θ i κ
$\theta^T\Sigma\theta = \frac{-1}{2}\theta^T\Sigma\Sigma^{-1}A\kappa = \theta^T AH^{-1}B = (A^T\theta)^T H^{-1}B = B^T H^{-1}B$
which completes the proof, assuming that H is invertible.

Classical Markowitz theory - analytic approach (cont.)

Let $H = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$, and $\upsilon = detH$.

If matrix Σ is nonsingular and all coordinates of μ are nonidentical and $\mu \neq 0$ then $\upsilon > 0$.

From the definition of H we have

$a = \mu^T \Sigma^{-1} \mu$, $b = \mu^T \Sigma^{-1} \mathcal{J} = \mathcal{J}^T \Sigma^{-1} \mu$, $c = \mathcal{J}^T \Sigma^{-1} \mathcal{J}$

also $\upsilon = ac - b^2$

since Σ is positive definite ($y^T \Sigma y > 0$ when $y \neq 0$), the matrix Σ^{-1} is also positive definite i.e. $a > 0$ whenever $\mu \neq 0$ and $c > 0$

Furthermore

$(b\mu - a\mathcal{J})^T \Sigma^{-1} (b\mu - a\mathcal{J}) = bba - abb - abb - aac =$

$a(ac - b^2) = a\upsilon > 0$

whenever μ does not have the same coordinates. The proof of Theorem 1 is completed.

Classical Markowitz theory - analytic approach (cont.)

(4) characterizes the efficient frontier for variance as a measure of risk. This will be the upper part of the parabola in coordinate system $(Risk(R(\theta)), E\left\{R(\theta)\right\})$
Using matrix $H = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ we have

$$\min_{\theta} Var(R(\theta)) = \frac{1}{\upsilon}(c\mu_p^2 - 2b\mu_p + a) \tag{6}$$

which give us formula for the above mentioned parabola. The coordinates of the origin of the parabola are
$\mu_{mv} = \frac{b}{c}$ and $Var(R(\theta_{mv})) = \frac{1}{c}$.

Optimal strategy

$$\theta = \frac{-1}{2}\Sigma^{-1}A\kappa = \Sigma^{-1}AH^{-1}B \tag{7}$$

also

$$H^{-1} = \frac{1}{\upsilon}\begin{bmatrix} c & -b \\ -b & a \end{bmatrix} \tag{8}$$

hence (since $A = [\mu, \mathcal{J}]$, $B = \begin{bmatrix} \mu_p \\ 1 \end{bmatrix}$).

$\theta_{opt} = \frac{1}{\upsilon}\Sigma^{-1}A\begin{bmatrix} c\mu_p - b \\ -b\mu_p + a \end{bmatrix} =$
$\frac{1}{\upsilon}\Sigma^{-1}\left(\mu(c\mu_p - b) + \mathcal{J}(-b\mu_p + a)\right) =$
$\frac{1}{\upsilon}\Sigma^{-1}\left((a\mathcal{J} - b\mu) + (c\mu - b\mathcal{J})\mu_p\right)$
which given the formula for an optimal strategy.

Minimal variance portfolio

$$\theta_{opt} = \frac{1}{\upsilon}\Sigma^{-1}\left((a\mathcal{J} - b\mu) + (c\mu - b\mathcal{J})\mu_p\right) \tag{9}$$

if now $\mu_p = \frac{b}{c} = \mu_{mv}$ then
$\theta_{opt} = \frac{1}{\upsilon}\Sigma^{-1}\left(a - \frac{b^2}{c}\right)\mathcal{J} = \Sigma^{-1}\mathcal{J}\frac{1}{c}$
is the strategy minimizing variance we denote θ_{mv} for which we have $Var(R(\theta_{mv})) = \frac{1}{c}$.

Sharpe coefficient - tangent portfolio

(William F. Sharpe, born Boston 1934, Nobel prize 1990)
Minimal variance strategy usually provides relatively small portfolio rate of return. Hence it is quite natural to look for other portfolios from the efficient frontier.
One opportunity is to maximize so called Sharp coefficient

$$\frac{\mu_p}{\sqrt{Var(R(\theta))}}, \tag{10}$$

We are looking for the greatest intersection coefficient of the line starting from the origin with efficient frontier (upper part of hiperbola) i.e. for linear part of the line tangent to efficient frontier (upper part of hiperbola) starting from the origin (we tacitly assume here that $\mu_{mv} > 0$)
In fact,
$Var(R(\theta_{opt})) = \frac{1}{v}(c\mu_p^2 - 2b\mu_p + a)$

Sharpe coefficient - tangent portfolio (cont.)

Hence $\sqrt{Var(R(\theta_{opt}))} = \sqrt{\frac{1}{\upsilon}(c\mu_p^2 - 2b\mu_p + a)}$

To find optimal tangent portfolio we calculate

$$\frac{d\mu_p}{d\sqrt{Var(R(\theta_{opt}))}} = \left(\frac{d\sqrt{Var(R(\theta_{opt}))}}{d\mu_p}\right)^{-1} =$$

$$\left(\frac{1}{2\sqrt{\frac{1}{\upsilon}(c\mu_p^2 - 2b\mu_p + a)}}\frac{1}{\upsilon}(2c\mu_p - 2b)\right)^{-1} = \frac{\upsilon\sqrt{\frac{1}{\upsilon}(c\mu_p^2 - 2b\mu_p + a)}}{c\mu_p - b}$$

We would like to have $\frac{d\mu_p}{d\sqrt{Var(R(\theta_{opt})}} = \frac{\mu_p}{\sqrt{Var(R(\theta_{opt}))}}$, i.e.

$\mu_p(c\mu_p - b) = \upsilon Var(R(\theta_{opt})) = c\mu_p^2 - 2b\mu_p + a$

which gives $b\mu_p = a$ and

$$\mu_p = \frac{a}{b} = \mu_{tg} \tag{11}$$

Sharpe coefficient - tangent portfolio (cont.)

If $\mu_{tg} = \frac{a}{b}$ we have
$Var(R(\theta_{tg})) = \frac{1}{\upsilon}(c\left(\frac{a}{b}\right)^2 - 2b\frac{a}{b} + a) = \frac{1}{\upsilon}(\frac{ca^2}{b^2} - a) = \frac{a}{b^2}$
Using (9)

$$\theta_{tg} = \frac{1}{\upsilon}\Sigma^{-1}\left(a\mathcal{J} - b\mu\right. \\ \left. +(c\mu - b\mathcal{J})\frac{a}{b}\right) = \Sigma^{-1}\frac{\mu}{b} \tag{12}$$

We looked for the line tangent to $\sqrt{Var(R(\theta_{opt}))}$. One could look for the line tangent to the graph of $Var(R(\theta_{opt}))$.

Maximization of the coefficient

$$\frac{\mu_p}{Var(R(\theta))} \tag{13}$$

i.e. tangency coefficient to the upper part of efficient frontier parabola

$c\mu_p^2 - 2b\mu_p + a - \upsilon Var(R(\theta_{opt})) = 0$

The upper part of the parabola is given by

$\mu_p = \frac{2b+\sqrt{\Delta}}{2c}$

with $\Delta = 4b^2 - 4c(a - \upsilon Var(R(\theta_{opt}))) = 4\upsilon(cVar(R(\theta_{opt})) - 1)$

the tangency coefficient should satisfy

$\frac{d\mu_p}{dVar(R(\theta_{min}))} = \frac{1}{2c}\frac{1}{2\sqrt{\Delta}}4\upsilon c = \frac{\upsilon}{\sqrt{\Delta}}$

Maximization of the coefficient (cont.)

the tangent line is of the form $\mu = \frac{\upsilon}{\sqrt{\Delta}} Var + z$
it intersect the origin when $z = 0$. We look for μ and Var from the upper parabola part
i.e. $2c\mu = 2b + \sqrt{\Delta}$ and $\mu\sqrt{\Delta} = \upsilon Var$, which gives
$2c\mu^2 = 2b\mu + \upsilon Var$. Therefore $c\mu^2 + a - 2c\mu^2 = 0$, and
$\mu = \sqrt{\frac{a}{c}} = \mu_{st}$
Hence
$Var(R(\theta_{st})) = \frac{1}{\upsilon}(c\left(\frac{a}{c}\right)^2 - 2b\frac{a}{c} + a) = \frac{1}{\upsilon}(\frac{a^2}{c} - 2\frac{ba}{c} + a) = \frac{1}{\upsilon c}a(a - 2b + c)$
and
$\theta_{st} = \frac{1}{\upsilon}\Sigma^{-1}\left((a\mathcal{J} - b\mu) + (c\mu - b\mathcal{J})\frac{a}{c}\right) = \frac{1}{\upsilon}\Sigma^{-1}\left(a(1 - \frac{b}{c})\mathcal{J} + (a - b)\mu\right)$

The role of minimal variance and tangent portfolios

$\theta_{opt} = \frac{1}{\upsilon}\Sigma^{-1}\left((a\mathcal{J} - b\mu) + (c\mu - b\mathcal{J})\mu_p\right)$
$\theta_{tg} = \Sigma^{-1}\frac{\mu}{b}$
$\theta_{mv} = \Sigma^{-1}\mathcal{J}\frac{1}{c}$
Hence

$$\theta_{opt} = \frac{-b^2 + bc\mu_p}{\upsilon}\theta_{tg} + \frac{ac - bc\mu_p}{\upsilon}\theta_{mv} \qquad (14)$$

which means that optimal strategy is a linear combination of tangent and minimal variance strategies.

Value functionals

replace two criterions problem by one criterion

we maximize

$$F\left(E\left\{R(\theta)\right\}, Risk(R(\theta))\right) \tag{15}$$

over all admissible portfolio strategies θ. The choice of the point of efficient frontier is replaced by the choice of risk parameter λ
The function F should be increasing with respect to the first coordinate and decreasing with respect to the second. The risk aversion is measured by the parameter $\lambda \geq 0$. The most natural form of the function F is

$$F(x, y) = x - \frac{1}{2}\lambda y. \tag{16}$$

Value functionals (cont.)

We maximize
$F\left(E\left\{R(\theta)\right\}, Var(R(\theta))\right) = E\left\{R(\theta)\right\} - \frac{1}{2}\lambda Var(R(\theta)) =$
$\theta^T \mu - \frac{1}{2}\lambda\theta^T\Sigma\theta$
with respect to such theta θ that $\theta^T \mathcal{J} = 1$
Again we use Lagrange multipliers method. We form the function
$G(\theta, \kappa) = \theta^T \mu - \frac{1}{2}\lambda\theta^T\Sigma\theta + \kappa(\theta^T \mathcal{J} - 1)$
Necessary condition of optimality is:
$\frac{\partial G}{\partial \theta_i} = 0 \quad \frac{\partial G}{\partial \kappa} = 0$
hence we obtain
Theorem 2. If μ does not have the same coordinates and matrix Σ is nonsingular then

$$\theta_{opt} = \frac{1}{\lambda}\Sigma^{-1}\left(\mu + \mathcal{J}\frac{\lambda - b}{c}\right) \tag{17}$$

Proof

from the necessary condition we get $\mu - \lambda\Sigma\theta + \mathcal{J}\kappa = 0 \quad \mathcal{J}^T\theta = 1$
we solve the first equation with respect to θ
$\theta = \frac{1}{\lambda}\left(\Sigma^{-1}\left(\mu + \mathcal{J}\kappa\right)\right)$
and substitute to the second equation
$\frac{1}{\lambda}\mathcal{J}^T\Sigma^{-1}\left(\mu + \mathcal{J}\kappa\right) = 1$
Hence (using nonsingularity of Σ)
$\kappa = \frac{\lambda - \mathcal{J}^T\Sigma^{-1}\mu}{\mathcal{J}^T\Sigma^{-1}\mathcal{J}} = \frac{\lambda - b}{c}$
(recall that $c = \mathcal{J}^T\Sigma^{-1}\mathcal{J}$ i $b = \mathcal{J}^T\Sigma^{-1}\mathcal{J}$)
$\theta_{opt} = \frac{1}{\lambda}\Sigma^{-1}\left(\mu + \mathcal{J}\frac{\lambda - b}{c}\right) = \frac{b}{\lambda}\theta_{tg} + (1 - \frac{b}{\lambda})\theta_{mv}$
since $\theta_{tg} = \Sigma^{-1}\mu\frac{1}{b}$ and $\theta_{mv} = \Sigma^{-1}\mathcal{J}\frac{1}{c}$

$$\theta_{opt} = \frac{b}{\lambda}\theta_{tg} + (1 - \frac{b}{\lambda})\theta_{mv} \tag{18}$$

which is an analogy (14) for Markowitz model.
When $\lambda = b$ ($b = \mathcal{J}^T\Sigma^{-1}\mu$) we have $\theta_{opt} = \theta_{tg}$
and when $\lambda \to \infty$ we have $\theta_{opt} \to \theta_{mv}$.

Proof (cont.)

$\mu_{opt} = \theta_{opt}^T \mu = \frac{1}{\lambda} \mu^T \Sigma^{-1} \mu + \frac{1}{\lambda} \mathcal{J}^T \Sigma^{-1} \mu \frac{\lambda - b}{c} = \frac{a}{\lambda} + \frac{b}{\lambda} \frac{\lambda - b}{c} = \frac{a}{\lambda} + \frac{b}{c} - \frac{b^2}{c\lambda} = \frac{\upsilon}{c\lambda} + \mu_{mv}$
since $a = \mu^T \Sigma^{-1} \mu$, $b = \mathcal{J}^T \Sigma^{-1} \mu$, zaś $\mu_{mv} = \frac{b}{c}$. Furthermore (since $c = \mathcal{J}^T \Sigma^{-1} \mathcal{J}$ and $Var(R(\theta_{mv})) = \frac{1}{c}$)
$Var(R(\theta_{opt})) = \theta_{opt}^T \Sigma \theta_{opt} = \frac{1}{\lambda}(\mu^T + \mathcal{J}^T \frac{\lambda - b}{c}) \Sigma^{-1} \Sigma \frac{1}{\lambda} \Sigma^{-1} (\mu + \mathcal{J} \frac{\lambda - b}{c}) = \frac{1}{\lambda^2}(a + 2b \frac{\lambda - b}{c} + (\frac{\lambda - b}{c})^2 c) = \frac{1}{\lambda^2}(a + \frac{2b\lambda - 2b^2 + \lambda^2 - 2b\lambda + b^2}{c}) = \frac{1}{\lambda^2}(a + \frac{\lambda^2 - b^2}{c}) = \frac{\upsilon}{c\lambda^2} + \frac{1}{c} = \frac{\upsilon}{c\lambda^2} + Var(R(\theta_{mv}))$

Form of the variance corresponding to the optimal strategy

We have $\mu_{opt} = \frac{\upsilon}{c\lambda} + \mu_{mv}$, $\mu_{mv} = \frac{b}{c}$ i $Var(R(\theta_{mv})) = \frac{1}{c}$
$Var(R(\theta_{opt})) = \frac{\upsilon}{c\lambda^2} + Var(R(\theta_{mv})) = \frac{(\mu_{opt}-\mu_{mv})^2 c}{\upsilon} + Var(R(\theta_{mv})) =$
$\frac{1}{\upsilon}\left(c\mu_{opt}^2 - 2b\mu_{opt} + \frac{b^2}{c} + \frac{\upsilon}{c}\right) = \frac{1}{\upsilon}\left(c\mu_{opt}^2 - 2b\mu_{opt} + a\right)$
This is an analogy to (6) ($\min_\theta Var(R(\theta)) = \frac{1}{\upsilon}(c\mu_p^2 - 2b\mu_p + a)$) obtained for Markowitz model.

Summary:Using optimal strategies for $F\left(E\left\{R(\theta)\right\}, Var(R(\theta))\right) = E\left\{R(\theta)\right\} - \frac{1}{2}\lambda Var(R(\theta)) = \theta^T\mu - \frac{1}{2}\lambda\theta^T\Sigma\theta$ for different λ we obtain the whole part of the efficient frontier form the origin of the parabola (minimal risk point) to the Sharpe point (tangency point with the effective frontier hiperbola)

Comments and remarks

1. the methodology considered (both Markowitz and one criterion aim functional) does not impose any constraints on the portfolio strategies (we admit both short selling and short borrowing). Vector θ can admit arbitrary values (both positive and negative but $\theta^T \mathcal{J} = 1$); If we are looking for nonnegative strategies we have to study (in a number of models) further part of efficient frontier (using continuity of the aim functional)

2. Function $F_\lambda \left(E\left\{R(\theta)\right\}, Var(R(\theta))\right) = E\left\{R(\theta)\right\} - \frac{1}{2}\lambda Var(R(\theta))$ describes a constant risk aversion - the derivative of this function with respect to λ is constant. An alternative approach leads to study risk sensitive functionals of the form $F_\lambda(R(\theta)) = \frac{-1}{\lambda} \ln E\left\{\exp\left\{-\lambda R(\theta)\right\}\right\}$

Risk sensitive functional - motivation:

Let $g(\lambda) = \ln E\{\exp\{-\lambda X\}\}$ - using Taylor expansion we have
$g(\lambda) = g(0) + \lambda g'(0) + \frac{\lambda^2}{2} g''(0) + remainder(\lambda)$
$g'(\lambda) = \frac{E\{-Xe^{-\lambda X}\}}{E\{e^{-\lambda X}\}}$
$g''(\lambda) = \frac{E\{X^2 e^{-\lambda X}\}E\{e^{-\lambda X}\} - (E\{-Xe^{-\lambda X}\})^2}{(E\{e^{-\lambda X}\})^2}$
hence
$\frac{-1}{\lambda} g(\lambda) = E\{X\} - \frac{1}{2}\lambda Var(X) + remainder(\lambda)$, i.e.
$F_\lambda(R(\theta)) = \frac{-1}{\lambda} \ln E\{\exp\{-\lambda R(\theta)\}\} =$
$E\{R(\theta)\} - \frac{1}{2}\lambda Var(R(\theta)) + remainder(\lambda)$

References:

De Giorgi, E., – A Note on Portfolio Selection under Various Risk Measures, Working paper, University of Zurich 2002.

Elton, E. J., Gruber, M. J. – Modern Portfolio Theory and Investment Analysis, Wiley, 1981.

Markowitz, H., – Portfolio Selection Efficient Diversification of Investments, Wiley, 1959.

Sharpe, W. F., – Portfolio Theory and Capital Markets, McGraw-Hill, 1970.

Szegö, G. P., – Portfolio Theory with Application to Bank Asset Management, Academic Press, 1980.

Whittle, P., – Risk-sensitive Optimal Control, Wiley, 1990.

Risk sensitive stationary portfolio

$F_\lambda(R(\theta)) = \frac{-1}{\lambda} \ln E \{\exp\{-\lambda R(\theta)\}\}$
Fact 1. $\theta \mapsto F_\lambda(R(\theta))$ is (strictly) concave (Hölder inequality).
There is at most one maximum point of $\theta \mapsto F_\lambda(R(\theta))$ $\theta^T \mathcal{J} = 1$.
Fact 2.(Jensen inequality) $F_\lambda(R(\theta)) \leq E\left\{\theta^T \zeta\right\}$.
Lagrange multiplier's method ?
$G(\theta, \kappa) = \frac{-1}{\lambda} \ln E\{\exp\{-\lambda R(\theta)\}\} + \kappa(\theta^T \mathcal{J} - 1)$
Necessary condition of optimality is:
$\frac{\partial G}{\partial \theta_i} = 0 \quad \frac{\partial G}{\partial \kappa} = 0$
We obtain
$\frac{-1}{\lambda} \frac{E\left\{\exp\left\{-\lambda\theta^T\zeta\right\}(-\lambda)\zeta^i\right\}}{E\left\{\exp\left\{-\lambda\theta^T\zeta\right\}\right\}} + \kappa = 0$
$\mathcal{J}^T \theta = 1$
How to find $\theta(\kappa)$ satisfying the first equation? (is it possible?)

Counterexample

$d = 2$, $\theta_2 = 1 - \theta_1$
$\zeta_1 \in \{a_1, b_1\}$,
$\zeta_2 \in \{a_2, b_2\}$
$-1 < a_1 < 0 < b_1$, $-1 < a_2 < 0 < b_2$ (no arbitrage opportunity)
$a_1 < a_2$, $b_1 > b_2$
$-\lambda\theta_1(b_2 - a_1) \preceq F_\lambda(R(\theta)) \preceq \lambda\theta_1(b_1 - a_2)$
summary: $\sup_\theta F_\lambda(R(\theta)) = \infty$,
Lagrange multipliers method can not be used.
When we restrict ourselves to $\theta_i \geq 0$ everything is fine.
We have however to use approximate methods to find an optimal (unique portfolio).

Markowitz problem with semivariance

$$Risk(R(\theta)) = SVar(R(\theta)) = E\left\{\left(\sum_{i=1}^{d}\theta_i(\zeta_i-\mu_i)\right)^{-2}\right\} \to \min$$

under $\theta^T\mu = \mu_p$ and $\theta^T\mathcal{J} = 1$.

$r_i = \zeta_i - \mu_i$, $Er_i = 0$

$\theta_1 = 1 - \sum_{i=1}^{d}\theta_i$,

$$E\left\{\left(\sum_{i=1}^{d}\theta_i r_i\right)^{-2}\right\} = E\left\{\left(r_1 + \sum_{i=2}^{d}\theta_i(r_i - r_1)\right)^{-2}\right\} \to \min$$

under

$\sum_{i=2}^{d}\theta_i(\mu_i-\mu_1) = \mu_p - \mu_1$

Two cases: 1. $\mu_i = \mu_1$, for each i then $\mu = \mu_p$, the problem becomes

$$\min_{(\theta_2,\ldots,\theta_d)} E\left\{\left(r_1 + \sum_{i=2}^{d}\theta_i(r_i - r_1)\right)^{-2}\right\} \to \min$$

An auxiliary lemma

Problem (A): $\min_{x \in R^m} E\left\{(A + B^T x)^-\right\}^2$
where $B = (B_1, \ldots, B_m)^T$, $EB_i^2 < \infty$, $EB_i = 0$ and $EA^2 < \infty$.
Lemma. The problem (A) admits an optimal solution.
Proof of Lemma.
Assume first that $B_1, \ldots, B_m$ are linearly independent
$P(\sum_{i=1}^m \alpha_i B_i = 0) = 1$ implies $\alpha_1 = \ldots = \alpha_m = 0$.
Let $S = \left\{(k, y) \in R^{m+1} : k \in [0, 1], \|y\| = 1\right\}$
$c = \inf_{(k,y) \in S} E[(kA + B^T y)^-]^2$
$(k, y) \to E[(kA + B^T y)^-]^2$ is continuous so there is (k^*, y^*) such that $c = E[(k^* A + B^T y^*)^-]^2$. If $c = 0$ we have $k^* > 0$ since otherwise $E[(B^T y^*)^-]^2 = 0$ and $P(B^T y^* \geq 0) = 1$, $EB^T y^* = 0$ and finally $P(B^T y^* = 0) = 1$, $y^* = 0$ but $\|y^*\| = 1$ a contradiction.

Proof of Lemma cont.

$c = 0 \Rightarrow k^* > 0$
then $\frac{y^*}{k^*}$ is an optimal solution to the Problem (A).
When $c > 0$ then for $\|x\| \geq 1$
$E[(A + B^T x)^-]^2 = \|x\|^2 E[(\frac{A}{\|x\|} + B^T \frac{x}{\|x\|})^-]^2 \geq \|x\|^2 c$ so that we have coercivity of
$x \to E\left\{(A + B^T x)^-\right\}^2$, and since it is a continuous function it admits a minimizer x^*.
Assume now that $\{B_1, \ldots, B_m\}$ are not independent. If $P(B = 0) = 1$ and $x \in R^m$ is a minimizer. If $P(B \neq 0) > 0$ there is a subset D of $\{B_1, \ldots, B_m\}$ whose elements are linearly independent and every element in this set is a linear combination of D. Suppose that such subset is $\{B_1, \ldots, B_k\}$ and let
$\tilde{B} = (B_1, \ldots, B_k)^T$.
By the the proof there is a minimizer $\tilde{x}^*$ for
$x \to E\left\{(A + \tilde{B}^T x)^-\right\}^2$ and $x^* = (\tilde{x}^*, 0, \ldots, 0)^T$ is a minimizer.

Semivariance main result

Theorem There is a minimizer to

$$E\left\{\left(\sum_{i=1}^d \theta_i r_i\right)^{-2}\right\} = E\left\{\left(r_1 + \sum_{i=2}^d \theta_i(r_i - r_1)\right)^{-2}\right\} \to \min$$

under

$$\sum_{i=2}^d \theta_i(\mu_i - \mu_1) = \mu_p - \mu_1$$

Case 1. $\mu_i = \mu_1$ (continuation) by Lemma we have a minimizer to

$$\min_{(\theta_2,\ldots,\theta_d)} E\left\{\left(r_1 + \sum_{i=2}^d \theta_i(r_i - r_1)\right)^{-2}\right\}$$

Case 2. There is i such that $r_1 \neq r_i$. For simplicity let $i = 2$. We have

$$\theta_2 = \frac{\mu_p - \mu_1}{\mu_2 - \mu_1} - \sum_{i=3}^d \theta_i \frac{(\mu_i - \mu_1)}{\mu_2 - \mu_1}$$

and therefore the problem is reduced to min over $(\theta_3, \ldots, \theta_d)$ of

$$E\left\{\left(r_1 + \frac{\mu_p - \mu_1}{\mu_2 - \mu_1}(r_2 - r_1) + \sum_{i=2}^d \theta_i(r_i - r_1) - (r_2 - r_1)\frac{(\mu_i - \mu_1)}{\mu_2 - \mu_1}\right)^{-2}\right\}$$

and by Lemma there is a minimizer.

Semivariance result generalizations

H. Jin, H. Markowitz, XY Zhou, A note on semivariance, Math. Fin. 16 (2006), 53-61

general downside risk function: $f(x) = 0$ for $x \geq 0$, and $f(x) > 0$ for $x < 0$

example: $f(x) = (x^-)^p$ with $p > 0$

$E\left\{ f\left(\left(\sum_{i=1}^d \theta_i(\zeta_i - \mu_i) \right)^- \right) \right\} \to \min$

Assumptions: f l.s.c, $f(kx) \geq g(k)f(x)$ and $\lim_{x\to\infty} g(x) = \infty$.

Monetary Measures of Risk

financial position $X : \Omega \to R$ net worth position at maturity
$\mathcal{X}$ the class of financial positions
$\rho : \mathcal{X} \to R$ is monetary measure of risk when:
- if $X \leq Y$ we have $\rho(X) \geq \rho(Y)$ (monotonicity)
- if $m \in R$ then $\rho(X + m) = \rho(X) - m$ (translation invariance)

properties:

$$\rho(X + \rho(X)) = 0$$

($\rho(0) = 0$ (normalization))
$|\rho(X) - \rho(Y)| \leq \|X - Y\|$
monetary measure of risk is convex measure whenever
$\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$ for $\lambda \in [0, 1]$
(diversification does not increase the risk)

Coherent measures of risk

convex measure of risk is called a coherent measure of risk when
if $\lambda \geq 0$ we have $\rho(\lambda X) = \lambda\rho(X)$ (positive homogeneity)
properties of the coherent measures of risk:
$\rho(0) = 0$
$\rho(X + Y) \leq \rho(X) + \rho(Y)$

Acceptance sets

ρ - monetary measure of risk $\mathcal{A}_\rho := \{X \in \mathcal{X} | \rho(X) \leq 0\}$ - acceptance set

ρ convex iff $\mathcal{A}_\rho$ is convex

ρ homogeneous iff $\mathcal{A}_\rho$ is a cone.

$\rho(X) = \inf\{m \in R | m + X \in \mathcal{A}\}$

worst case measure ρ_{max}

$\rho_{max}(X) = -\inf_\omega X(\omega)$

is a coherent measure of risk

for every monetary measure of risk ρ we have

$\rho(X) \leq \rho(\inf_\omega X(\omega)) = \rho_{max}(X)$

$\rho(X) = -EX$ is also a coherent measure of risk.

$\rho(X) = \sup_{Q \in \mathcal{Q}} \{E_Q\{-X\}\}$ is also a coherent measure of risk

Value at Risk

$VaR_\alpha(X) = \inf\{m \in R | P\{m + X < 0\} \leq \alpha\}$

α quantiles: upper

$q_\alpha^+(X) := \inf\{x \in R : P\{X \leq x\} > \alpha\}$

lower quantile

$q_\alpha^-(X) := \inf\{x \in R : P\{X \leq x\} \geq \alpha\}$

α quantiles interval $[q_\alpha^-(X), q_\alpha^+(X)]$

$P\{-X \leq m\} \geq 1 - \alpha$

hence: $VaR_\alpha(X) = q_{1-\alpha}^-(-X)$

for continuous r.v. X

$\inf\{m : P\{m + X < 0\} \leq \alpha\} = \sup\{m : P\{m + X < 0\} > \alpha\}$

so that $VaR_\alpha(X) = -q_\alpha^+(X)$ and $P\{X \leq -VaR_\alpha(X)\} = \alpha$

Properties of the value at risk

$VaR_\alpha(X) = \inf\{m \in R | P\{m + X < 0\} \leq \alpha\}$

1. $X \geq 0$ then $VaR_\alpha(X) \leq 0$,
2. $X \geq Y$ then $VaR_\alpha(X) \leq VaR_\alpha(Y)$
3. $VaR_\alpha(\lambda X) = \lambda VaR_\alpha(X)$ for $\lambda \geq 0$
4. $VaR_\alpha(m + X) = VaR_\alpha(X) - m$

VaR_α is a monetary measure of risk

$\alpha \in [0, 0.05]$

VaR_α in general is not a convex measure of risk

Conditional value at risk

X continuous integrable r.v.
(expected shortfall, average value at risk)
$CVaR_\alpha(X) = E\{-X|X + VaR_\alpha(X) \le 0\}$
properties of $CVaR_\alpha$

1. $CVaR_\alpha(\lambda X) = \lambda CVaR_\alpha(X)$ for $\lambda \ge 0$
2. $CVaR_\alpha(m + X) = CVaR_\alpha(X) - m$

Lemma: If $X \in L_1$, $x \in R$ s.t. $P\{X \le x\} > 0$ then for any A s.t. $P(A) \ge P\{X \le x\}$ we have

$$E\{X|A\} \ge E\{X|X \le x\}$$

Proof.

$$\frac{1}{P(A)} \int_A u f_X(u) du \ge \frac{1}{P\{X \le x\}} \int_{X \le x} u f_X(u) dy$$

hint: approximate X by discrete r.v.

Conditional value at risk (cont.)

3. $X \geq Y$ implies that $CVaR_\alpha(X) \leq CVaR_\alpha(Y)$
since $P(X + VaR_\alpha(X) \leq 0) = \alpha = P(Y + VaR_\alpha(Y) \leq 0)$
hence $CVaR_\alpha(Y) \geq -E\{Y|X + VaR_\alpha(X) \leq 0\} \geq$
$-E\{X|X + VaR_\alpha(X) \leq 0\} = CVaR_\alpha(X)$
4. $CVaR_\alpha(X + Y) \leq CVaR_\alpha(X) + CVaR_\alpha(Y)$
since
$CVaR_\alpha(X + Y) = E\{-X|X + Y + VaR_\alpha(X + Y) \leq 0\} +$
$E\{-Y|X + Y + VaR_\alpha(X + Y) \leq 0\} \leq$
$E\{-X|X + VaR_\alpha(X) \leq 0\} + E\{-Y|Y + VaR_\alpha(Y) \leq 0\}$
CVaR is a coherent measure of risk
5. $CVaR_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha VaR_\alpha(X)$
$CVaR_\alpha$ is a continuous function of the parameter α
6. $\lim_{\alpha\to 0} CVaR_\alpha(X) = -essinfX$
$\lim_{\alpha\to 1} CVaR_\alpha(X) = -EX$

Measures of dispersion (deviation)

$\mathcal{X} = L^2$
$D : \mathcal{X} \to [0, \infty]$ is a measure of dispersion iff
$D(X + c) = D(X)$ for any $c \in R$
$D(0) = 0$, $D(\alpha X) = \alpha D(X)$ for $\alpha > 0$
$D(X + Y) \leq D(X) + D(Y)$
$D(X) \geq 0$, and $D(X) > 0$ for $X \neq const$
Examples: standard deviation, negative semi standard deviation
$\sigma_-(X) = \sqrt{E(\max(EX - X, 0))^2}$
positive semi standard deviation
$\sigma_+(X) = \sqrt{E(max(X - EX, 0))^2}$

Elliptic distributions

random vector $X = (X_1, \ldots, X_d)^T$ has elliptic law, if there is a vector μ, a positive definite symmetric matrix Ω, a nonnegative function g_d such that, $\int_0^\infty x^{\frac{d}{2}-1} g_d(x)dx < \infty$, and a norming constant c_d such that the density f_X of the vector X is of the form
$f_X(x) = c_d |\Omega|^{-1/2} g_d(\frac{1}{2}(x-\mu)^T \Omega^{-1}(x-\mu))$,
where $|\Omega|$ is the determinant of Ω. One can show that
$c_d = \frac{\Gamma(\frac{d}{2})}{(2\pi)^{d/2}} \left(\int_0^\infty x^{d/2-1} g_d(x)dx \right)^{-1}$.
where $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ and for positive integer d we have $\Gamma(d) = d!$, while $\Gamma(d+\frac{1}{2}) = \frac{1 \cdot 2 \cdot 3 \cdot (2d-1)}{2^d} \sqrt{\pi}$ i $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
Douglas Kelker 1970 Department of Statistics and Applied Probability, University of Alberta

Properties of elliptic distributions (I)

characteristic function of the vector X is given by
$\phi_X(t) = E\left(e^{itX}\right) = e^{it^T\mu}\psi(\frac{1}{2}t^T\Omega t)$
for a function $\psi(t)$ called characteristic generator.
Notation: $X \sim E_d(\mu, \Omega, \psi)$, $X \sim E_d(\mu, \Omega, g_d)$.
If $\int_0^\infty g_1(x)dx < \infty$ there exists EX and $EX = \mu$. If furthermore
$|\psi'(0)| < \infty$
or equivalently $\int_0^\infty \sqrt{x} g_1(x)dx < \infty$ then
$Cov(X) := E\left\{(X - EX)(X - EX)^T\right\} = -\psi'(0)\Omega$

Properties of elliptic distributions (II)

If $X \sim E_d(\mu, \Omega, g_d)$, A is $m \times d$ matrix ($m \leq d$) and b - m dim. vector then

$AX + b \sim E_m(A\mu + b, A\Omega A^T, g_m)$

linear combination of elliptic distributions with the same generator ψ is elliptic with generator ψ.

marginal law of $X \sim E_d(\mu, \Sigma, g_d)$ is elliptic

$X_k \sim E_1(\mu_k, \omega_k^2, g_1)$

where ω_k^2 is the k-th element of the diagonal of Ω, the density of X_k is of the form

$f_{X_k}(x) = \frac{c_1}{\omega_k} g_1\left(\frac{1}{2}\left(\frac{x-\mu_k}{\omega_k}\right)^2\right).$

Properties of elliptic distributions (III)

Main property

If $X \sim E_d(\mu, \Omega, g_d)$ then for
$Y = \theta_1 X_1 + \theta_2 X_2 + \ldots + \theta_d X_d = \theta^T X$ we have
$Y \sim E_1(\theta^T \mu, \theta^T \Omega \theta, g_1)$

Examples

Multidimensional normal $X \sim N_d(\mu, \Sigma)$
the density is of the form: (we identify $\Omega = \Sigma$)
$f_X(x) = \frac{c_d}{\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$
with $c_d = (2\pi)^{\frac{-d}{2}}$; the characteristic function
$\phi_X(t) = \exp\left\{it^T \mu - \frac{1}{2} t^T \Sigma t\right\}$ so that $g(u) = e^{-u}$ and $\psi(t) = e^{-t}$
(since $\psi'(0) = -1$ $\Sigma = Cov(X)$).

Examples of elliptic distributions

multidimensional Student $X \sim t_d(\mu, \Omega; p)$ with $p > \frac{d}{2}$
the density is of the form:
$f_X(x) = \frac{c_d}{\sqrt{|\Omega|}} \left[1 + \frac{(x-\mu)^T \Omega^{-1}(x-\mu)}{2k_p}\right]^{-p}$
where $c_d = \frac{\Gamma(p)}{\Gamma(p-\frac{d}{2})}(2\pi k_p)^{\frac{-d}{2}}$, and k_p is a constant dependent on p,
we have here $g_d(u) = (1 + \frac{u}{k_p})^{-p}$
in particular cases when $p = \frac{d+\nu}{2}$ i $k_p = \frac{\nu}{2}$ we have multidimensional t-Student with ν degrees of freedom and then $\Omega = \frac{\nu}{\nu-2}\Sigma$

Examples of elliptic distributions cont.

in particular case when $p = \frac{d+m}{2}$ for positive integer m i d and $k_p = \frac{m}{2}$ we have

$$f_X(x) = \frac{\Gamma(\frac{d+m}{2})}{(\pi m)^{\frac{d}{2}}\Gamma(\frac{m}{2})\sqrt{|\Sigma|}}\left[1 + \frac{(x-\mu)^T\Omega^{-1}(x-\mu)}{m}\right]^{-\frac{d+m}{2}}$$

in general case for $k_p = \frac{2p-3}{2}$ with $p > \frac{3}{2}$ we have $Cov(X) = \Omega$ then in particular for $p = \frac{d+m}{2}$

$$f_X(x) = \frac{\Gamma(\frac{d+m}{2})}{(\pi(d+m-3))^{\frac{d}{2}}\Gamma(\frac{m}{2})\sqrt{|\Omega|}}\left[1 + \frac{(x-\mu)^T\Omega^{-1}(x-\mu)}{d+m-3}\right]^{-\frac{d+m}{2}}$$

with $Cov(X) = \Omega$.

Examples of elliptic distributions cont.

$f_X(x) = \frac{c_d}{\sqrt{|\Omega|}} \left[1 + \frac{(x-\mu)^T \Omega^{-1} (x-\mu)}{2k_p} \right]^{-p}$
when $\frac{1}{2} < p \leq \frac{3}{2}$ there are no variance (heavy tails)
when $\frac{1}{2} < p \leq 1$ we have that EX does not exist. For $p = 1$ we have a multidimensional Cauchy distribution

$f_X(x) = \frac{\Gamma(\frac{d+1}{2}) \pi^{-\frac{d+1}{2}}}{\sqrt{|\Omega|}} \left[1 + (x-\mu)^T \Omega^{-1} (x-\mu) \right]^{-\frac{d+1}{2}}$

multidimensional logistic $g(u) = \frac{e^{-u}}{(1+e^{-u})^2}$, multidimensional exponential $g(u) = e^{-ru^s}$.

Portfolio analysis with elliptic rate of return

principal assumption: random rate of return ζ is $E_d(\mu, \Omega, \psi)$
portfolio rate of return $R(\theta))$ (for a strategy $\theta = (\theta_1, \ldots, \theta_d)$) is of the law
$E_1(\theta^T \mu, \theta^T \Omega \theta, \psi)$
furthermore $Var(R(\theta)) = -\psi'(0)\omega^2$, where $\omega^2 = \theta^T \Omega \theta$, and
$\frac{R(\theta) - \theta^T \mu}{\omega} \sim E_1(0, 1, \psi)$
This procedure allows standardization of the elliptic r. v.

Risk measures for elliptic rate of returns

ζ is $E_d(\mu, \Omega, \psi)$ and consequently $R(\theta)$ is $E_1(\theta^T\mu, \theta^T\Omega\theta, \psi)$
probability of the shortfall
$Risk(R(\theta)) = P\left\{R(\theta) \leq q\right\}.$
Using standardization we obtain
$Risk(R(\theta)) = F_Y(\frac{q-\theta^T\mu}{\omega}),$
where F_Y is the distribution of $E_1(0, 1, \psi)$.

Value at Risk - VaR_α

restriction on the portfolio rate of return of the form
$P\{R(\theta) \leq q\} \leq \alpha$
which leads to the following lower bound for the expected portfolio rate of return
$\theta^T \mu + \kappa_\alpha \omega \geq q,$
where κ_α is α quantile of $E_1(0,1,\psi)$
Value at Risk (VaR_α):
$VaR_\alpha(R(\theta)) = \inf\{x : P\{R(\theta) + x \leq 0\} \leq \alpha\}$
is the minimal value added to the portfolio rate of return which guarantees nonpositive rate of return with probability at most α.
We have
$VaR_\alpha(R(\theta)) = -\kappa_\alpha \omega - \theta^T \mu$
where κ_α is α quantile of $E_1(0,1,\psi)$.

Conditional VaR_α or $CVaR_\alpha$

conditional VaR_α ($CVaR_\alpha$), called also shortfall (expected shortfall)
$CVaR_\alpha(R(\theta)) = E\left\{-R(\theta) | R(\theta) + VaR_\alpha \le 0\right\}$,
which is the expected value of $-R(\theta)$ given nonpositive $R(\theta) + VaR_\alpha$.
One can show that (Föllmer Schied)
$CVaR_\alpha(R(\theta)) = \frac{1}{\alpha}\int_0^\alpha VaR_\beta d\beta = -\omega\frac{1}{\alpha}\int_0^\alpha \kappa_\beta d\beta - \theta^T\mu$.
We see that if we had no ω, where $\omega^2 = \theta^T\Omega\theta$, for a given α both $VaR_\alpha(R(\theta))$ and $CVaR_\alpha(R(\theta))$ would be a linear function of the investment strategy θ or in other words they would depend on the expected portfolio rate of return $\theta^T\mu$ only.

Risk functions for elliptic rate of return

Since $CVaR_\alpha$ is a coherent measure of risk we consider the following optimization problem
$F_\lambda(E(R(\theta), CVaR_\alpha(R(\theta)) = E(R(\theta)) - \frac{1}{2}\lambda CVaR_\alpha(R(\theta))$.
Notice that
$F_\lambda(E(R(\theta), CVaR_\alpha(R(\theta))) = (1+\frac{1}{2}\lambda)\theta^T\mu + \frac{1}{2}\lambda\sqrt{\theta^T\Omega\theta}\frac{1}{\alpha}\int_0^\alpha \kappa_\beta d\beta$.
the second term is negative since for small α (usually below 0.05) the value of $\int_0^\alpha \kappa_\beta d\beta$ is negative; We are not able to solve the problem explicitly. On can find the maximum of
$F_\lambda(E(R(\theta), CVaR_\alpha(R(\theta)))$
using approximate methods
In fact, consider Lagrange multiplier's method to the function
$F_\lambda(E(R(\theta), CVaR_\alpha(R(\theta)))$
We form
$G(\theta,\kappa) = (1+\frac{1}{2}\lambda)\theta^T\mu + \frac{1}{2}\lambda\sqrt{\theta^T\Omega\theta}\frac{1}{\alpha}\int_0^\alpha \kappa_\beta d\beta + \kappa(\theta^T\mathcal{J} - 1)$
Necessary condition for optimality is
$\frac{\partial G}{\partial \theta_i} = 0 \quad \frac{\partial G}{\partial \kappa} = 0$
Hence

Risk functions for elliptic rate of return (cont.)

$(1+\frac{1}{2}\lambda)\mu + \frac{1}{2}\lambda\frac{1}{\sqrt{\theta^T\Omega\theta}}2\Omega\theta z(\alpha) + \mathcal{J}\kappa = 0$
with $\mathcal{J}^T\theta = 1$.
We are not able to solve θ from the first equation, which earlier together with the send equation gave us κ.
The difficulties come because of the existence of the term with square root. Since $\sqrt{x} \le x$ for $x \ge 1$ one can optimize the modified risk function for elliptic rates of return
$F^m_\lambda(\theta) = (1+\frac{1}{2}\lambda)\theta^T\mu + \frac{1}{2}\lambda\theta^T\Omega\theta\frac{1}{\alpha}\int_0^\alpha \kappa_\beta d\beta,$
which is diminished the term with $\theta^T\Omega\theta$.
We use to this new problem the Lagrange multiplier's method and have
$G(\theta,\kappa) = (1+\frac{1}{2}\lambda)\theta^T\mu + \frac{1}{2}\lambda\theta^T\Omega\theta\frac{1}{\alpha}\int_0^\alpha \kappa_\beta d\beta + \kappa(\theta^T\mathcal{J} - 1)$
Necessary condition for optimality is then of the form
$\frac{\partial G}{\partial\theta_i} = 0 \quad \frac{\partial G}{\partial\kappa} = 0$
from which we obtain
$(1+\frac{1}{2}\lambda)\mu + \frac{1}{2}\lambda 2\Omega\theta z(\alpha) + \mathcal{J}\kappa = 0$
with $\mathcal{J}^T\theta = 1$.

Risk functions for elliptic rate of return (cont.)

From the first equation

$\theta = \frac{1}{\lambda z(\alpha)} \Omega^{-1} \left(-(1 + \frac{1}{2}\lambda)\mu - \mathcal{J}\kappa \right)$

substituting this to the second equation we obtain

$\mathcal{J}^T \frac{1}{\lambda z(\alpha)} \Omega^{-1} \left(-(1 + \frac{1}{2}\lambda)\mu - \mathcal{J}\kappa \right) = 1$

or

$\frac{1}{\lambda z(\alpha)} \left(-(1 + \frac{1}{2}\lambda)\mathcal{J}^T \Omega^{-1}\mu - \mathcal{J}^T \Omega^{-1} \mathcal{J}\kappa \right) = 1$

$\kappa = \left(\left(-(1 + \frac{1}{2}\lambda)\mathcal{J}^T \Omega^{-1}\mu \right) - \lambda z(\alpha) \right) \frac{1}{\mathcal{J}^T \Omega^{-1} \mathcal{J}}$

so

$\kappa = \frac{1}{\mathcal{J}^T \Omega^{-1} \mathcal{J}} \left(-(1 + \frac{1}{2}\lambda)\mathcal{J}^T \Omega^{-1}\mu \right) - \frac{\lambda z(\alpha)}{\mathcal{J}^T \Omega^{-1} \mathcal{J}}$

and finally

$\theta = \frac{-(1+\frac{1}{2}\lambda)}{\lambda z(\alpha)} \Omega^{-1}\mu + \frac{(1+\frac{1}{2}\lambda)\mu^T \Omega^{-1} \mathcal{J}}{\lambda z(\alpha) \mathcal{J}^T \Omega^{-1} \mathcal{J}} \Omega^{-1} \mathcal{J} + \frac{1}{\mathcal{J}^T \Omega^{-1} \mathcal{J}} \Omega^{-1} \mathcal{J}$

with $z(\alpha) = \frac{1}{\alpha} \int_0^\alpha \kappa_\beta d\beta$.

Further alternatives

One can consider also a function
$Risk(R(\theta)) = (CVaR_\alpha(R(\theta)) + ER(\theta))^2$.
which corresponds to the square of the CVaR (an analogy to the variance considered as a square of the standard deviation).
Then
$\tilde{F}_\lambda = \theta^T \mu - \frac{1}{2}\lambda\theta^T \Omega\theta(z(\alpha))^2$
the only difference is that $z(\alpha) = \frac{1}{\alpha}\int_0^\alpha \kappa_\beta d\beta$ in the aim function has been replaced by $(z(\alpha))^2$. The optimal strategy is of the form
$\tilde{\theta} = \frac{-(1+\frac{1}{2}\lambda)}{\lambda(z(\alpha))^2}\Omega^{-1}\mu + \frac{(1+\frac{1}{2}\lambda)}{\lambda(z(\alpha))^2}\frac{\mu^T\Omega^{-1}\mathcal{J}}{\mathcal{J}^T\Omega^{-1}\mathcal{J}}\Omega^{-1}\mathcal{J} + \frac{1}{\mathcal{J}^T\Omega^{-1}\mathcal{J}}\Omega^{-1}\mathcal{J}$.

Further references

De Giorgi, E., – A Note on Portfolio Selection under Various Risk Measures, Working paper, University of Zurich 2002.
Fang, K.T., Kotz, S., Ng, K.W., Symmetric Multivariate and Related Distributions, London, Chapman & Hall 1987.
Föllmer, H., Schied, A. – Stochastic Finance. An Introduction in Discrete Time, W. de Gruyter, 2002.
Kelker, D., – Distribution Theory of Spherical Distributions and Location-Scale Parameter Generalization, Sankhya 32 (1970), 419-430
Landsman, Z., Valdez, E. A. – Tail Conditional Expectations for Elliptical Distributions, North American Actuarial Journal 7 (2003), 55-71,