Financial Optimization
ISE 347/447

Lecture 24

Dr. Ted Ralphs

Reading for This Lecture

- C&T Chapter 17

Risk Measures

- A*risk measure* ρ, as formally defined in the literature, is just a real-valued funtion of a random variable (not a very useful definition).
  - The random variable we are primarily interested in is the random vector of returns of a given portfolio.
  - Intuitively, the risk measure can be thought of as the opposite of a "utility function."
  - With utility, bigger is better; with risk, smaller is better.
- To make intuitive sense, the function should have certain properties, including
  - Monotonicity: $Y \geq X \Rightarrow \rho(Y) \leq \rho(X)$ (adding more assets to one's portfolio should not increase risk).
  - Convexity: $\rho((1-\lambda)X + \lambda Y) \leq (1-\lambda)\rho(X) + \lambda\rho(Y)$ (the risk of two separate portfolios is at least as much as the risk of one combined portfolio).
- It is also intuitive that a risk measure should not be "symmetric" (the disutility of a loss does not equal the utility of a similar gain).
- Note that a "risk measure" may not only be an evaluation of the risk in the intuitive way we think of it.

Value at Risk (VaR)

- The only risk measure we have considered so far is the variance of the return of a given portfolio.
- This measure is convex, but is not monotonic and is also symmetric.
- *Value at risk* is a risk measure developed at J.P.Morgan, which is in wide-spread use across the finance industry.
- VaR_α is defined as the smallest level of loss for which the probability of experiencing a loss above this level is smaller than $1 - \alpha$.
- In other words, the loss will exceed VaR_α with probability at most $1 - \alpha$.
- This measure is "asymmetric," since it is only positive when there is a loss.
- Let us now define this notion more formally.

Definitions

- Consider an investment decision represented by the vector $x \in \mathbb{R}^n$.
- Let the "loss" over an investment period under the outcome $\omega \in \Omega$ be $L(x, \omega)$.
- For fixed x, the loss function $L(x, \cdot)$ is a random variable that takes positive values when a loss is incurred, and negative ones when a gain occurs.
- For any fixed value of x let

$$\Psi(x, \gamma) := \mathrm{P}\left[L(x, \cdot) \leq \gamma\}\right] = F_{L(x,\cdot)}(\gamma)$$

be the cumulative distribution function of the loss function $L(x, \cdot)$ associated with holding the investment x.

Definitions (cont.)

For any $\alpha \in [0,1]$ (typically, $\alpha = .95$ is chosen), the value at risk on the confidence level α is defined by

$$\begin{aligned} \mathrm{VaR}_\alpha(x) := \min_{\gamma \in \mathbb{R}} \ & \gamma \\ \text{s.t.} \quad & \Psi(x,\gamma) \geq \alpha. \end{aligned}$$

- Since Ψ is typically a nonlinear function, computing the value at risk is a nonlinear programming problem.
- Note that if the loss function is continuous, then $\mathrm{VaR}_\alpha(x)$ is such that

$$\Psi(x, \mathrm{VaR}_\alpha(x)) = \alpha$$

Example 1

- A set of risky assets $S^1, \dots, S^n$ have multivariate normal returns $R \sim \mathrm{N}(\mu, Q)$ over the investment period $[0, 1]$.
- Suppose we want to find the portfolio x^* that minimizes the value at risk on the confidence level α.
- If the total value of the invested capital is w, then the loss incurred by the portfolio x over the investment period is $-wR^\top x$.
- Therefore, we have to solve a *bilevel* optimization problem (see next slide).

Portfolio Optimization with VaR

$$\text{(VM1)} \quad x^* = \arg\min_{x \in \mathbb{R}^n} \; \mathrm{VaR}_\alpha(x)$$
$$\text{s.t.} \quad Ax \geq a, \; Bx = b,$$

where the objective function

$$\mathrm{VaR}_\alpha(x) = \min_{\gamma \in \mathbb{R}} \; \gamma$$
$$\text{s.t.} \quad \int_{\{r : -wr^\top x \leq \gamma\}} \frac{\exp\left\{-\frac{1}{2}(r-\mu)^\top Q^{-1}(r-\mu)\right\}}{\sqrt{(2\pi)^n \det(Q)}} \, dr \geq \alpha$$

is itself the optimal solution to an optimization problem.

Example 2

- The return vector R of a set of risky assets $S^1, \ldots, S^n$ takes the values $r^1, \ldots, r^k \in \mathbb{R}^n$ with probability $1/k$ each.
- Find the vector x^* of relative wealth allocation weights that minimizes the value at risk on the confidence level α.

$$\text{(VM2)} \quad x^* = \arg\min_{x \in \mathbb{R}^n} \ \mathrm{VaR}_\alpha(x)$$
$$\text{s.t.} \quad Ax \geq a, \ Bx = b,$$

with

$$\mathrm{VaR}_\alpha(x) = \min_{\gamma \in \mathbb{R}} \ \gamma$$
$$\text{s.t.} \quad \sum_{\{i : -wx^\top r^i \leq \gamma\}} \frac{1}{k} \geq \alpha.$$

Drawbacks

VaR_α-minimization has a number of serious drawbacks:

- The objective function VaR_α in Example 2 is nonlinear and nonsmooth with many local minimizers.
- Both (VM1) and (VM2) are bilevel optimization problems, which are generally computationally difficult to solve..
- VaR_α is not *convex*!
  - Suppose we buy two bonds for $100 each, each of which will default with a probability of 4%.
  - $\text{VaR}_{0.95}$ is 0 for each bond independently, since the probability of losing nothing is more than 95% for each bond individually.
  - The combination of the two bonds has $\text{VaR}_{.95}$ equal to 100, however (the probability of losing nothing on either bond is $.96^2 < 0.95$).
  - So $\text{VaR}_{.95}$ is bigger for the combination!
- VaR_α pays no attention to the *magnitude* of losses when the rare extremal event of experiencing a loss above the level VaR_α occurs.

Conditional Value at Risk (CVaR)

- To overcome this drawback, the notion of *conditional value at risk* (CVaR) has been developed.
- This is the same as *mean expected loss*, *mean shortfall*, *expected shortfall risk* and *tail-VaR*.
- As before, let $\alpha \in [0, 1]$ be a given confidence level.
- Then we define

$$\mathrm{CVaR}_\alpha(x) := \frac{1}{1-\alpha} \int_{\{\omega : L(x,\omega) \geq \mathrm{VaR}_\alpha(x)\}} L(x, \omega)\, \mathrm{P}[d\omega].$$

- The intuitive basis for this definition is that when the loss function is continuous, we have

$$\mathrm{CVaR}_\alpha(x) = \mathbb{E}\left[L(x, \omega) \,\|\, L(x, \omega) \geq \mathrm{VaR}_\alpha(x)\right]$$

Example 3

- A given investment generates losses of $L(j) = j - 80 \; (j = 1, \ldots, 100)$ each with probability 1%.
- We have
$$\mathrm{VaR}_{0.95} = \min_{j=1,\ldots,100} L(j)$$
$$\text{s.t.} \quad \sum_{i=1}^{j} \frac{1}{100} \geq 0.95.$$
- The constraint is satisfied for $j = 95, \ldots, 100$. Therefore,
$$\mathrm{VaR}_{\alpha} = \min_{j=95,\ldots,100} (j - 80) = 15.$$
- The expected shortfall risk is
$$\mathrm{CVaR}_{0.95} = \frac{1}{0.05} \sum_{j=95}^{100} \frac{j-80}{100} = 17.5.$$

Comparing VaR and CVaR

- Note that

$$\begin{aligned}\mathrm{CVaR}_\alpha(x) &\geq \frac{1}{1-\alpha}\int_{\{\omega : L(x,\omega)\geq \mathrm{VaR}_\alpha(x)\}} \mathrm{VaR}_\alpha(x)\,\mathrm{P}[d\omega] \\ &= \frac{\mathrm{VaR}_\alpha(x)}{1-\alpha}\mathrm{P}\left[L(x,\omega)\geq \mathrm{VaR}_\alpha(x)\right] \\ &\geq \mathrm{VaR}_\alpha(x),\end{aligned}$$

so minimizing CVaR_α also makes VaR_α small, but the opposite may not be true.

- $\mathrm{CVaR}_\alpha(x)$ can now be used as a risk measure in investment decision problems that take the form

$$\text{(CVM)} \quad x^* = \arg\min_{x\in\mathbb{R}^n} \mathrm{CVaR}_\alpha(x)$$
$$\text{s.t.} \quad x \in \mathcal{F},$$

where $\mathcal{F}$ is some set of feasible investments defined by a set of constraints.

Comparing VaR and CVaR: Simple Example

- Suppose again that we buy two bonds for \$100 each, each of which will default with a probability of 4%.
- $\mathrm{CVaR}_{0.95}$ is $80 = (.04 \times 100)/0.05$ for each bond independently
- The combination of the two bonds has $\mathrm{CVaR}_{.95}$ equal to $(200 \times .04^2 + 100 \times (.05 - .04^2))/.05 = 103$.
- Note that in this example, we do not have $\mathrm{P}[L(x,\omega) \geq \mathrm{VaR}_\alpha(x)] = 1-\alpha$ (we will come back to this).
- This means that the risk of the two bonds together is now less than the sum of the risks of the individual bonds ($103 < 160$).
- In fact, we can show that CVaR_α is both monotonic and convex.

Example 4

In Example 1, if we had proposed to find an investment that minimizes CVaR_α, we would have had to solve

$$\text{(CVM1)} \quad x^* = \arg\min_{x\in\mathbb{R}^n} \ \mathrm{CVaR}_\alpha(x)$$
$$\text{s.t.} \quad Ax \geq a, \ Bx = b,$$

where

$$\mathrm{CVaR}_\alpha(x) = \frac{-w}{1-\alpha}\int_{\{r:-wr^\top x\geq \mathrm{VaR}_\alpha(x)\}} \frac{r^\top x\cdot \exp\left\{-\frac{1}{2}(r-\mu)^\top Q^{-1}(r-\mu)\right\}}{\sqrt{(2\pi)^n\det(Q)}}\,dr$$

and

$$\mathrm{VaR}_\alpha(x) = \min_{\gamma\in\mathbb{R}} \ \gamma$$
$$\text{s.t.} \quad \int_{\{r:-wr^\top x\leq\gamma\}} \frac{\exp\left\{-\frac{1}{2}(r-\mu)^\top Q^{-1}(r-\mu)\right\}}{\sqrt{(2\pi)^n\det(Q)}}\,dr \geq \alpha$$

Example 5

In Example 2, if we had proposed to find an investment that minimizes CVaR_α, we would have had to solve

$$\text{(CVM2)} \quad x^* = \arg\min_{x \in \mathbb{R}^n} \ \mathrm{CVaR}_\alpha(x)$$
$$\text{s.t.} \quad Ax \geq a, \ Bx = b,$$

where

$$\mathrm{CVaR}_\alpha(x) = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} -wx^\top r^i,$$
$$\mathcal{I} = \{i : -wx^\top r^i \geq \mathrm{VaR}_\alpha(x)\},$$

and

$$\mathrm{VaR}_\alpha(x) = \min_{\gamma \in \mathbb{R}} \ \gamma$$
$$\text{s.t.} \quad \sum_{\{i : -wx^\top r^i \leq \gamma\}} \frac{1}{k} \geq \alpha.$$

Computing CVaR

- These examples illustrate that computing $\mathrm{CVaR}_\alpha(x)$ generally requires the computation of $\mathrm{VaR}_\alpha(x)$.
- This suggests that the CVaR_α-minimization problem

$$\text{(CVM)} \quad x^* = \arg\min_{x\in\mathbb{R}^n} \mathrm{CVaR}_\alpha(x) \quad \text{s.t.} \quad x \in \mathcal{F}$$

might be even harder than the VaR_α-minimization problem

$$\text{(VM)} \quad x^* = \arg\min_{x\in\mathbb{R}^n} \mathrm{VaR}_\alpha(x) \quad \text{s.t.} \quad x \in \mathcal{F}.$$

- It thus comes as a surprise that under quite reasonable modeling assumptions, the opposite is true.

Let $\beta(x) = \mathrm{P}[L(x,\omega) \geq \mathrm{VaR}_\alpha(x)]$ and consider the auxiliary function

$$F_\alpha(x,\gamma) := \gamma + \int_\Omega \frac{(L(x,\omega) - \gamma)_+}{\beta(x)} \; \mathrm{P}[d\omega].$$

Theorem 1.

i) For any fixed x, the function $\gamma \mapsto F_\alpha(x,\gamma)$ is convex.

ii) $\mathrm{VaR}_\alpha(x)$ is a minimizer of the problem $\min_\gamma F_\alpha(x,\gamma)$.

iii) $F_\alpha(x, \mathrm{VaR}_\alpha(x)) = \mathrm{CVaR}_\alpha(x)$.

Proof: i) Since $(L(x,\omega)-\gamma)_+$ is a convex function in γ, it is true that for any γ_1,γ_2 and $\tau\in[0,1]$,

$$F_\alpha(x,\tau\gamma_1+(1-\tau)\gamma_2)$$

$$\le \tau\gamma_1+(1-\tau)\gamma_2 + \int_\Omega \left(\tau\frac{(L(x,\omega)-\gamma_1)_+}{\beta(x)} + (1-\tau)\frac{(L(x,\omega)-\gamma_2)_+}{\beta(x)}\right) \mathrm{P}[d\omega]$$

$$= \tau F_\alpha(x,\gamma_1)+(1-\tau)F_\alpha(x,\gamma_2).$$

This shows that $F_\alpha(x,\gamma)$ is convex in γ.

ii) Since the problem of minimizing $F_\alpha(x,\gamma)$ with respect to γ is convex, the KKT conditions are sufficient for optimality, i.e., we only need to check that the $F_\alpha(x,\gamma)$ is stationary at $\gamma = \mathrm{VaR}_\alpha(x)$.

For any set $\mathcal{S} \subset \Omega$ let $\chi_{\mathcal{S}}$ be the associated indicator function

$$\chi_{\mathcal{S}}(\omega) = \begin{cases} 1 & \text{if } \omega \in \mathcal{S}, \\ 0 & \text{otherwise.} \end{cases}$$

With this notation we have

$$\begin{aligned} \frac{\partial}{\partial \gamma} F_\alpha(x, \mathrm{VaR}_\alpha(x)) &= 1 - \int_\Omega \frac{\chi_{\{\omega : L(x,\omega) \geq \mathrm{VaR}_\alpha(x)\}}(\omega)}{\beta(x)} \mathrm{P}[d\omega] \\ &= 1 - \frac{\mathrm{P}\left[L(x,\omega) \geq \mathrm{VaR}_\alpha(x)\right]}{\beta(x)} = 0 \end{aligned}$$

iii) We have

$$F_\alpha(x, \mathrm{VaR}_\alpha(x)) = \mathrm{VaR}_\alpha(x) + \int_\Omega \frac{(L(x,\omega) - \mathrm{VaR}_\alpha(x))_+}{\beta(x)} \, \mathrm{P}[d\omega]$$

$$= \mathrm{VaR}_\alpha(x) + \int_{\{\omega : L(x,\omega) \geq \mathrm{VaR}_\alpha(x)\}} \frac{L(x,\omega)}{\beta(x)} \mathrm{P}[d\omega] - \mathrm{VaR}_\alpha(x) \frac{\mathrm{P}[L(x,\omega) \geq \mathrm{VaR}_\alpha(x)]}{\beta(x)}$$

$$= \mathrm{VaR}_\alpha(x) + \mathrm{CVaR}_\alpha(x) - \mathrm{VaR}_\alpha(x).$$

Minimizing CVaR_α

- Theorem 1 now implies that the CVaR_α-minimization problem

$$\text{(MCV)} \quad \min_{x\in\mathbb{R}^n} \mathrm{CVaR}_\alpha(x) \quad \text{s.t.} \quad x \in \mathcal{F}$$

can be reformulated as the single-level optimization problem

$$\text{(MCV')} \quad \min_{(x,\gamma)\in\mathbb{R}^{n+1}} F_\alpha(x,\gamma) \quad \text{s.t.} \quad x \in \mathcal{F}.$$

- In applications, it is often the case that F_α is convex in x as well, and $\mathcal{F}$ is a convex set.
- In this case (MCV') is a convex minimization problem and can generally be well solved.

Example 6

- Problem (CVM1) from Example 4 is equivalent to

$$\text{(CVM1')} \quad \min_{x,\gamma} \gamma + \frac{1}{1-\alpha} \int_{\mathbb{R}^n} \left(-wr^\top x - \gamma\right)_+ \frac{\exp\left\{-\frac{1}{2}(r-\mu)^\top Q^{-1}(r-\mu)\right\}}{\sqrt{(2\pi)^n \det(Q)}}\, dr$$

$$\text{s.t.} \quad Ax \geq a, \; Bx = b.$$

- Since $(-wr^\top x - \gamma)_+$ is convex in x, the objective function of (CVM1') is a positive combination of convex functions and hence also convex in x.
- By Theorem 1 the objective function is also convex in γ.

Example 7

- Problem (MCV2) from Example 5 is equivalent to

$$\text{(MCV2')} \quad \min_{x,\gamma} \gamma + \frac{1}{\beta(x)} \sum_{i=1}^{k} \frac{(-wx^\top r^i - \gamma)_+}{k}$$
$$\text{s.t.} \quad Ax \geq a, \; Bx = b.$$

- Since $\beta(x) \approx 1 - \alpha$, Problem (MCV2) can be approximated by the convex problem

$$\text{(MCV2')} \quad \min_{x,\gamma} \gamma + \frac{1}{1-\alpha} \sum_{i=1}^{k} \frac{(-wx^\top r^i - \gamma)_+}{k}$$
$$\text{s.t.} \quad Ax \geq a, \; Bx = b.$$

Example 7 (cont.)

- Finally, Problem (MCV2') is equivalent to the following LP,

$$\text{(LMCV2')} \quad \min_{x,z,\gamma} \gamma + \frac{1}{(1-\alpha)k} \sum_{i=1}^{k} z_i$$

$$\text{s.t.} \quad z_i \geq -wx^\top r^i - \gamma, \quad (i = 1, \ldots, k)$$

$$Ax \geq a, \quad Bx = b,$$

$$z \geq 0,$$

- Note that we replaced a piecewise linear convex objective function by a linear objective by introducing extra variables and extra linear constraints.
- This is the same thing we did in the L-shaped method.

General Techniques

- Example 7 can be generalized to approximate any CVaR_α-minimization problem via LP or QP:
- For this purpose we replace the probability measure P on Ω by a finite set of equiprobable scenarios $\omega_1, \ldots, \omega_S$.
- These scenarios are typically obtained by statistical sampling.
- Next, we approximate F_α by

$$F_\alpha(x, \gamma) = \gamma + \frac{1}{(1-\alpha)S} \sum_{s=1}^{S} (L(x, \omega_s) - \gamma)_+,$$

so that the problem (MCV) can be approximated.

Approximating

The approximation is then

$$\text{(AMCV)} \quad \min_{x,\gamma} \; \gamma + \frac{1}{(1-\alpha)S} \sum_{s=1}^{S} (L(x,\omega_s) - \gamma)_+$$

$$\text{s.t.} \quad x \in \mathcal{F}.$$

Introducing artificial variables to get rid of the break points of the objective function, we replace (AMCV) by the equivalent problem

$$\text{(LAMCV)} \quad \min_{x,z,\gamma} \; \gamma + \frac{1}{(1-\alpha)S} \sum_{i=1}^{S} z_s$$

$$\begin{aligned} \text{s.t.} \quad & z_s \geq 0, \quad (s = 1, \ldots, S) \\ & z_s \geq L(x,\omega_s) - \gamma, \quad (s = 1, \ldots, S) \\ & Ax \geq b. \end{aligned}$$

Remarks

- If $L(x, \omega)$ is linear in x, then (LAMCV) is an LP.
- More generally, $L(x, \omega)$ is typically convex in x, in which case (LAMCV) is well solved via standard NLP software.
- In applications in which $L(x, \omega)$ is not convex in x, (LAMCV) is often further approximated by replacing $L(x, \omega)$ by an approximation that is convex in x.
- Typically, NLP software will do this automatically.

Further Applications

- In risk management, one is often interested in controlling the expected loss at several confidence levels.
- The following model is typical,

$$
\begin{aligned}
\text{(RM)} \quad & \max_{x} \ \mu^\top x \\
\text{s.t.} \quad & \mathrm{CVaR}_{\alpha_j}(x) \le u_{\alpha_j}, \quad (j = 1, \dots, k) \\
& Ax \ge a, \quad Bx = b.
\end{aligned}
$$

- To control the risk of the investment x, we thus require that the conditional value at risk must not exceed thresholds u_{α_j} on the confidence levels $\alpha_1, \dots, \alpha_k$.

Adapting the LAMCV

The reformulation of the finite scenario case can easily be adapted to such problems, which now become

$$
\begin{aligned}
\text{(ARM)} \quad & \max_{x,\gamma,z} \; \mu^\top x \\
\text{s.t.} \quad & \gamma + \frac{1}{(1-\alpha_j)S}\sum_{s=1}^{S} z_s \le U_{\alpha_j}, \quad (j = 1,\ldots,k) \\
& z_s \ge 0, \quad (s = 1,\ldots,S) \\
& z_s \ge L(x,\omega_s) - \gamma, \quad (s = 1,\ldots,S), \\
& Ax \ge a, \quad Bx = b.
\end{aligned}
$$