Risk Measures
(MSc MCF, Oxford)

Samuel N. Cohen

University of Oxford

Hilary Term

Outline

1. Introduction
2. Loss distributions
3. Risk Measurement
   - Value at Risk
     - Problems with VaR
   - Expected Shortfall
   - Other common risk measures
4. Convex and Coherent Risk Measures
   - Importance sampling
   - Dual Representations
5. Dynamic Consistency
6. Optimal Risk Transfer
7. Euler Principle and Capital Allocation

Introduction

Introduction

- We are interested in mathematical methods of describing, quantifying and managing risk.
- Basic Idea:
  - take a model for the assets or risks,
  - use data to calibrate this model to the real world,
  - assign a 'risk value' to the assets.
- We will consider the last of these steps.
- References:
  - McNeil, Frey and Embrechts, *Quantitative Risk Management*, Princeton, 2005
  - Föllmer and Schied, *Stochastic Finance: An introduction in Discrete Time*, de Gruyter, 2002.

Loss distributions

Loss distributions

- For a fixed period $[t, t+\Delta]$, consider the losses $L = -(V_{t+\Delta} - V_t)$.
- Our aim is to describe the distribution of L.

Example

- Consider a portfolio consisting of a single European call option, written on a stock following a simple Brownian motion with drift $\mu = 0.04$ and volatility $\sigma = 0.2$.
- Suppose the strike of the option is $K = \$10$, the expiry is in one year, the stock is currently trading at $S_t = \$12$, the risk-free interest rate is $r = 0.02$ and the current option price is the Black-Scholes price (\$2.3742).
- The loss distribution is the distribution of $L = -(V_{t+1} - V_t) = -(S_{t+1} - 10)^+ + 2.3742$. Note that S_{t+1} has a lognormal distribution, with mean $12e^{\mu+\sigma^2/2} = 12.742$ and variance $12(e^{\sigma^2} - 1)e^{2\mu+\sigma^2} = 0.5522$.

Risk Measurement

Risk Measurement

- We wish to reduce the amount of information we are dealing with – we don't want to have to deal with the whole distribution every time we want to consider our portfolio.
- We shall therefore consider methods which take the loss distribution and assign it a numerical value, representing the 'riskiness'.
- Some people may think this is an oversimplification, however it is good to remember that often all our analysis comes down to a yes–no question
  - is this position acceptable?
  - do I want to buy this stock?
  - which of these portfolios is a better investment?

Purposes of Risk Measurement

- What exactly we want this number to represent depends on our application.
- Risk measures can be used for many things, but three of the most common [MFE, p34] are:
  - Determination of risk capital and capital adequacy. How much capital do I need to store as a buffer against future losses on my portfolio? This may be to satisfy a regulator, for margin requirements on an exchange, or similar.
  - Management tool. Risk measures used internally to limit the amount of risk a firm may take, eg Traders told that 95% Value at risk cannot exceed a certain bound.
  - Insurance premiums. Risk measures determine the appropriate compensation(premium) that an insurer needs in exchange for bearing the risk of insured claims.
- The exact nature of our application determines what makes a 'good' or 'bad' risk measure.

Value at Risk

- We now move to one of the most common measures of risk, ‘Value at Risk’ (VaR).
- This measure of risk is part of the Basel/Solvency capital-adequacy frameworks and has also been used as a management tool (the JPMorgan 4:15 report).
- Suppose, from some method, we know $F_L = P(L \leq l)$. We wish to define a statistic based on F_L which describes the severity of the risk of holding our portfolio until the horizon.
- A candidate might be the worst (maximum) possible loss.

Example

Suppose our decision maker is long in the option described in Example 2.1. Then the worst possible loss would equal \$2.3742. However, if we considered instead being short in the option, then our worst possible loss is infinite (as losses are unbounded).

Value at Risk

- Because of this type of problem, we define VaR_α by replacing *maximum loss* with *maximum loss which will only be exceeded with some small probability* $1-\alpha$. Here α is called the confidence level.
- Mathematically, this means

$$\mathrm{VaR}_\alpha = \inf\{l \in \mathbb{R} : P(L > l) \leq 1-\alpha)\} = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\},$$

or in terms of the profit V,

$$\mathrm{VaR}_\alpha = \sup\{v \in \mathbb{R} : P(V < v) \leq 1-\alpha)\} = \sup\{v \in \mathbb{R} : F_V(v) \leq \alpha\}.$$

- VaR is simply a quantile of the loss distribution. Typical values for α are 0.95 and 0.99; and in market risk management the horizon Δ is usually either 1 or 10 days, in credit risk it is usually one year.
- We shall write $\mathrm{VaR}(L)$ for the value at risk associated with a (loss) position L.

Value at Risk

Note that

- VaR is very easy to calculate from the loss distribution, as it is simply a quantile.
- VaR does not give any indication of the scale of gains/losses away from the α-quantile.
- for any constant c, $\mathrm{VaR}(c) = -c$.
- $\mathrm{VaR}(L + c) = \mathrm{VaR}(L) - c$ for any position L, any constant c.
- $\mathrm{VaR}(\lambda L) = \lambda \mathrm{VaR}(L)$ for any $\lambda \geq 0$.
- If I have two positions L, L' and $L \leq L'$ with probability one, then $\mathrm{VaR}(L) \geq \mathrm{VaR}(L')$.
- the probability of a loss exceeding VaR_α is $\leq \alpha$.
- Suppose F_L is continuous and has an inverse, then $\mathrm{VaR}_\alpha = F_L^{-1}(\alpha)$.

Problems with VaR: Model Risk

- VaR is always prone to model risk, that is, that the risk it describes is always the risk that is given by the model.
- If our model is incorrect, inaccurate or in some other way misleading, then VaR will also be misleading.
- This is less a problem with VaR than a problem with how it is used – appropriate backtesting of our model is needed to ensure that our estimated VaR corresponds to the 'real-world' VaR.
- (Note that in some contexts, 'VaR' is used not only to refer to the mathematical risk measure, but to the whole process of statistical estimation and application within a firm's risk management system, further complicating this issue.)
- It is also important to consider this when managing extreme or correlated risks, for which very little data is available, and so good models are hard to find.

Problems with VaR: Choice of parameters and horizon

- Whenever we work with risk measures based on loss distributions, in addition to questions of model risk, we have to decide on a horizon Δ, and on a risk parameter (here α).
- Δ needs to correspond to the the time period over which the firm is committed to hold its portfolio.
- Liquidity and legal constraints can influence this, as the ability of the firm to change its position may be limited in practice.
- If an incorrect horizon is used, significant problems can arise. Similarly, the choice of α is debatable, as different applications will call for different levels.

Problems with VaR: Liquidity risk

- This is related to the earlier two points, however it is important to highlight it separately.
- While our position may have a notional value V_{t+1} in our model, if we are unable to liquidate our position for this value, our interpretation of VaR becomes more difficult.

Problems with VaR: Non-subadditivity

- From a mathematical perspective this is probably the most significant flaw, and the one which is best addressed by other measures of risk.
- It is possible to find portfolios L and L' such that $\mathrm{VaR}(L) + \mathrm{VaR}(L') < \mathrm{VaR}(L + L')$.
- This means that the diversified portfolio $L + L'$ has a *higher* risk than the two individual portfolios L and L', when considered using VaR.
- Hence, a firm using VaR may find it preferable not to diversify (which normally we would think of as reducing the risk).
- Furthermore, if a firm is able to split up its position into multiple sub-entities, and take the VaR of each separately, then by doing so it may be possible for them to lower their capital requirements, without changing the net position.

Problems with VaR

Example

Consider a position $W \sim N(0,1)$. What is the $\mathrm{VaR}_{0.95}(W)$? Now consider

$$L = \begin{cases} W & \text{if } W \in [\mathrm{VaR}_{0.95}(W) - 0.001, 0] \\ 0 & \text{otherwise} \end{cases}$$

Is $\mathrm{VaR}_{0.95}(L)$ more or less than $\mathrm{VaR}_{0.95}(W)$? What is $\mathrm{VaR}_{0.95}(W - L)$?

- If all assets under consideration have losses which follow normal distributions, then VaR is convex.
- If all assets have losses which are 'nearly normal', it is then reasonable to believe that VaR will not be too far from being convex.
- However, if we begin to consider derivatives, then typical assets no longer have distributions which are similar to normal distributions.

Expected Shortfall

- Due to the subadditivity of VaR, alternatives have been proposed.
- The alternative 'closest' to VaR is the Expected shortfall 'ES'.
- Note that this has various names, including 'expected shortfall', 'conditional VaR', 'tail VaR' and 'average VaR'. We shall use the term ES.
- This is defined as

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F_L)du = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u du$$

where q_u is the quantile function $q_u(F) = \inf(x \in \mathbb{R} : F(x) \geq \alpha)$.

Expected Shortfall

- We can show that

$$\mathrm{ES}_\alpha = \alpha^{-1} E[(L - \mathrm{VaR}_\alpha)^+] + \mathrm{VaR}_\alpha.$$

- When the loss is continuous, an even easier definition is possible:

$$\mathrm{ES}_\alpha = E(L | L \geq \mathrm{VaR}_\alpha)$$

  which is also often quite easy to calculate.
- A key advantage of ES over VaR is that it *is* subadditive. (See [FS] for proof.)

Other common risk measures

- Variance. Since the early work of Markowitz, this has been the standard measure of risk in many areas. However, it is often not appropriate in finance as it does not distinguish between up- and down-side risks. This causes particular problems as payoffs are often significantly skewed.
- Sharpe ratios. For assessing investments, a common quantity is the Sharpe ratio $E[-L]/Sd[L]$. This is connected to the market price of risk $(r - E[L])/Sd[L]$, and to the slope of the efficient frontier in a Markowitz framework. All the problems of variance are present in the Sharpe ratio.

Other common risk measures

- Partial moments. This is a useful general family of risk measures. We define the Upper Partial moment UPM(k, q) by

$$\mathrm{UPM}(k, q) = \int_q^\infty (l - q)^k dF_L(l) \geq 0.$$

  If $k = 0$, we obtain $P(L \geq q)$. For $k = 1$, we obtain $E((L - q)^+)$, for $k = 2$ and $q = E(L)$, we obtain the upper semivariance of L. Higher values of k make this more conservative.
- Extensions of partial moments are also common. The most basic is the Sortino ratio $E[-L]/UPM(2, E[L])$ which is a simple variation on the Sharpe ratio using the downside risk/semivariance.
- Another extension is the Omega ratio: $E[(L - q)^-]/E[(L - q)^+]$

A credit example

Example

- Consider a bank lending \$1 to 5000 clients, who are independent and have the same default probability $p = 1\%$.
- For each client i, we write $Z_i = 0$ if he/she does not default, $Z_i = 1$ otherwise.
- So $Z = \sum Z_i$ represents the total number of defaults (and hence the total loss), and is a binomial random variable $Z \sim Bin(5000, 0.01)$, that is

$$P(Z = x) = \binom{5000}{x} 0.012^x (0.988)^{5000-x}$$

- With $\alpha = 99\%$, we have $\mathrm{VaR}_\alpha(Z) = 67$, and $\mathrm{ES}_\alpha(Z) = 93.293$. (Either analytically or from the Monte-Carlo method below.)

Example

- Now suppose instead of being independent, these individuals are affected by an economy-wide factor. With probablity $q = 0.5\%$, the economy enters a significant downturn, and the probability of default jumps to $p_1 = 2.5\%$.
- If the economy does not turn, the probability of default is $p_2 = 0.9925\%$, (giving an overall rate of default 1%). We denote the total number of defaults in this model as Z'.
- We can use a Monte-Carlo method to calculate the risks.
- From here we estimate $\mathrm{VaR}_\alpha(Z') = 69$, but $\mathrm{ES}_\alpha(Z') = 182.131$.
- The reason for this difference is that ES observes the long tail of the distribution of losses in the second model, while VaR does not.

Convex and Coherent Risk Measures

Coherent Risk Measures

- To give a more general and theoretically defensible theory of risk, Artzner, Delbaen, Eber and Heath proposed the following axioms for a risk measure.
- A risk measure is a map $\rho : L^\infty(\mathcal{F}_T) \to \mathbb{R}$, that is it maps bounded random variables known at time T to the reals.
- Intuitively, it assigns a numerical value to each random outcome.
- It should satisfy the following properties
  - Monotonicity. If $X \geq Y$ with probability one, then $\rho(X) \leq \rho(Y)$.
  - Cash translability. For any constant c, $\rho(X + c) = \rho(X) - c$.
  - Positive homogeneity. If $\lambda \geq 0$, then $\rho(\lambda X) = \lambda\rho(X)$.
  - Subadditivity. $\rho(X + Y) \leq \rho(X) + \rho(Y)$.
- A map satisfying these assumptions will be called a coherent risk measure.

Acceptable positions

- With these assumptions, a random variable X is called *acceptable* if $\rho(X) \leq 0$. The set of all acceptable random variables will be denoted $\mathcal{A}$. Note that these have nice interpretations.
  - Monotonicity says more is preferable to less.
  - Cash translability means that $\rho(X)$ is the minimum amount of cash that needs to be added to a position to make it acceptable (as $\rho(X + \rho(X)) = 0$).
  - Positive homogeneity means that the scale of a position is unimportant, or equivalently, that the risk can be expressed in any equivalent units (dollars, cents, millions of dollars, etc) without a problem.
  - Subadditivity implies that diversification is good.

Convex risk measures

- Positive homogeneity is useful, but does not model liquidity risk well.
- Föllmer and Schied and Frittelli and Rosazza-Gianin (simultaneously) proposed a new assumption
  - Convexity. For $\lambda \in [0, 1]$,
$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y).$$
- Positive homogeneity and subadditivity imply convexity, but not vice-versa.

Examples

- **Expected Shortfall** As discussed earlier, the Expected Shortfall $\rho(L) = \mathrm{ES}_\alpha(L)$ is a coherent risk measure.
- **Expectations** If we pick any probability measure Q, then $\rho(L) = E_Q[-L]$ is a coherent risk measure (which is linear).
- **Essential minimum** Take $\rho(X) = -\operatorname{ess\,inf}_\omega X(\omega)$. Then ρ is a coherent risk measure. Furthermore $\mathcal{A}$ is the set of all almost-surely nonnegative random variables. (This is the 'worst case risk measure').

Example

For Z, Z' as in Example 3.1, our risk is 5000 in both cases.

Examples

- **Entropic Risk** Take

$$\rho^{\text{exp}}(X) = \gamma \ln E[\exp(-X/\gamma)],$$

  for $\gamma > 0$. Then ρ^{exp} is a convex (but not coherent) risk measure. (This is the 'entropic risk').
- One might think that we could attempt to use other utility functions u and try to find risk measures of the form $\rho(X) = -u^{-1}E[u(X)]$. However, the *Kolmogorov–de Finetti theorem on associative means* implies that this is only cash translable when u is linear or exponential.

Example

Verify that ρ^{exp} satisfies the properties of a convex risk measure.

Examples

Example

For Z, Z' as defined in Example 3.1, we can simply calculate

$$\texttt{gamma} * \texttt{log}(\texttt{mean}(\texttt{exp}(\texttt{z1}/\texttt{gamma})))$$

and similarly for Z'. For $\gamma = 2$, we find

$$\rho^{\text{exp}}(Z) = 64.638, \qquad \rho^{\text{exp}}(Z') = 146.709.$$

Examples

- **Probability Transform** An interesting class of risk measures comes about when we transform the probabilities using some concave distortion.
- Precisely, we take the distribution function F_L and a nondecreasing, onto, convex map $\psi : [0,1] \to [0,1]$. If we define $F_L^\psi(x) = \psi(F_L(x))$, we obtain a new distribution function F_L^ψ.
- For $V = -L$, we can then define $\rho(V) = \int_{\mathbb{R}} x dF_L^\psi(x)$.
- Examples of this approach are the Wang transform

$$\psi(x) = \Phi(\Phi^{-1}(x) - \gamma),$$

where Φ is the standard normal distribution function; or Cherny & Madan's minmaxvar

$$\psi(x) = (1 - (1-x)^{1/(1+\gamma)})^{1+\gamma}.$$

Example

Consider Z, Z' as in Example 3.1. Take $\gamma = 2$. From a Monte-Carlo method, we can estimate

$$\rho^{\text{minmaxvar}}(Z) \approx 68.045, \qquad \rho^{\text{wang}}(Z) \approx 64.711$$
$$\rho^{\text{minmaxvar}}(Z') \approx 95.963, \qquad \rho^{\text{wang}}(Z') \approx 81.899$$

Importance sampling

- A lot of effort is being wasted in simulating data with low losses.
- We can significantly improve our efficiency by using an importance sampling technique.
- Recall that the cdf of a random variable X with density f is given by

$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

- We simulate $\{x_1, x_2, ..., x_n\}$ from this distribution, to find the empirical cdf

$$F(x) \approx \tilde{F}(x) = \sum_{\{i:x_i \leq x\}} 1/n = \frac{\{\#\text{ observations below level } x\}}{n}$$

- Then transform $\tilde{F}$, and approximate the integral using first differences, that is, by calculating distorted weights on the observations.

Importance sampling

- When using importance sampling, we simulate data $\{y_1, y_2, ..., y_n\}$ using a distribution with density g, and then use
$$F(x) = \int_{-\infty}^{x} \frac{f(x)}{g(x)} g(x) dx.$$
- An alternative approximation of F is given by
$$F(x) \approx \tilde{F}(x) = \sum_{\{i: y_i \leq x\}} \left(\frac{f(y_i)}{g(y_i)}\right)\left(\frac{1}{n}\right).$$
- An even better (and simpler) approximation can be obtained by using a function $k(x) \propto \frac{f(x)}{g(x)}$ and the approximation
$$F(x) \approx \tilde{F}(x) = \frac{\sum_{\{i: y_i \leq x\}} k(y_i)}{\sum_i k(y_i)}.$$
- We choose g such that more of our observations are in the relevant 'tail part' of the distribution.

Example

- Suppose we are in a similar situation to the example above, where our losses Z follow a mixture distribution, where with probability 0.90, $Z = Z_1 \sim N(0,1)$, and, with probability 0.10, $Z = Z_2 \sim N(4,4)$ (ie variance 4).
- The density of Z is then

$$f(x) = 0.90\frac{1}{\sqrt{2\pi}}\exp(-x^2/2) + 0.10\frac{1}{\sqrt{8\pi}}\exp(-(x-4)^2/8)$$

- Instead of simulating this directly, we will simulate from a distribution $Z_3 \sim N(5,4)$. Hence

$$g(x) = \tfrac{1}{\sqrt{8\pi}}\exp(-(x-5)^2/8)$$

and

$$\frac{f(x)}{g(x)} \propto k(x) = 9\exp\Big(-\frac{3x^2+10x-25}{8}\Big) + 2\exp\Big(-\frac{2x-9}{8}\Big)$$

Dual Representations

- Suppose our risk measure is continuous from above, that is, if for any sequence of random variables $X_n \downarrow X$, then we have $\rho(X_n) \uparrow \rho(X)$.
- Then we can find a useful representation of coherent and convex risk measures.
- Let $\mathcal{M}$ denote the set of all probability measures on Ω which are equivalent to our real world probability $\mathbb{P}$.

Theorem

For any coherent risk measure ρ, there exists a (weakly relatively compact) set $\mathcal{Q} \subseteq \mathcal{M}$, such that

$$\rho(X) = \max_{Q \in \mathcal{Q}} \{E_Q[-X]\}.$$

Dual Representations

Theorem

For any convex risk measure ρ, there exists a function $\alpha : \mathcal{M} \to \mathbb{R}$, such that

$$\rho(X) = \max_{Q \in \mathcal{M}} \{E_Q[-X] - \alpha(Q)\}.$$

Furthermore, we can take

$$\alpha(Q) = \sup_{X \in \mathcal{A}} E_Q[-X].$$

Dual Representations

- We can think of coherent risk measures as taking the expectation where we are uncertain about what probability we should use, and so we take the worst one. This is a form of scenario analysis, where different scenarios are considered, and the worst case used to make a decision.
- We can think of convex risk measures as being the worst of a range of expectations under different probabilities, but where we can penalise probabilities that we think are unrealistic.
- Both of these results have converses. That is, if I have anything of the form on the RHS, then it will be a coherent (resp convex) risk measure.

Dynamic Consistency

Dynamic Consistency

- We now consider what happens when we try to update our risk measure, at different points in time.
- This is important when positions are allowed to change through time, and we wish to act to minimise our risk.
- For a random outcome X known at time T, we shall write $\rho_t(X)$ for the risk, as determined at time $t \leq T$.
- We wish for ρ_t to satisfy the requirements for a risk measure at each time t. We also allow ρ_t to depend on any information available at time t, and so ρ_t is random.

Dynamic Consistency

- We have the following new assumption.

$$\rho_t(I_A X) = I_A \rho_t(I_A X)$$

  where I_A is the indicator function of any event known at time t (i.e. $A \in \mathcal{F}_t$).
- This simply ensures that ρ_t does not depend on outcomes which are excluded at time t.
- A concept of dynamic consistency can then be written as:

  If $\rho_t(X) \geq \rho_t(Y)$ with probability one,
  then $\rho_s(X) \geq \rho_s(Y)$ for all $s \leq t$.

- This is a form of 'intertemporal monotonicity', and ensures that if we will always prefer Y to X at some point in the future, then this is reflected in our choices today.

Dynamic Consistency

Theorem

Intertemporal monotonicity is equivalent to stating $\rho_s(-\rho_t(X)) = \rho_s(X)$ *for all* X

Proof.

For any X, let $Y = -\rho_t(X)$.

Then $\rho_t(Y) = \rho_t(-\rho_t(X)) = \rho_t(X)$, and so by intertemporal monotonicity, $\rho_s(X) = \rho_s(Y) = \rho_s(-\rho_t(X))$.

Conversely, for any X, Y, if $\rho_t(X) \geq \rho_t(Y)$, then by monotonicity, $\rho_s(X) = \rho_s(-\rho_t(X)) \geq \rho_s(-\rho_t(Y)) = \rho_s(Y)$. $\square$

Dynamic Consistency

- From our examples above, the entropic risk measure is time consistent, that is, we can define

$$\rho_t(X) = \gamma \ln E[\exp(-X/\gamma)|\mathcal{F}_t]$$

  and this is time consistent.
- Similarly for the worst case risk measure.
- ES on the other hand is not time consistent (there is no way to consistently define ES at time t).
- In general, constructing time-consistent risk measures is difficult.

Optimal Risk Transfer

Optimal Risk Transfer

- We now look at problems associated with hedging. Here we assume that there are two parties, who can trade risky positions between themselves. Each of our parties, Alice and Bob, have a risk measure ρ_A, ρ_B.
- One application of this is when Bob is the market, and ρ_B is the best superhedging price available in the market.
- If the market is complete (so ρ_B is linear, $\rho_B(X) = E_\pi[-X]$ where π is the risk-neutral measure) and ρ_A is convex, then this will simply be the market price.
- If the market is incomplete (so perfect hedging is impossible), then ρ_B is a coherent risk measure, given by the maximum of $E_\pi[-X]$ over all risk-neutral measures π.

Optimal Risk Transfer

- Alice wishes to trade with Bob to minimise her risk, however can only trade in a way that Bob finds acceptable.
- That is, if W is Alice's initial position, we wish to find V to minimise $\rho_A(W - V)$ subject to $\rho_B(V) \geq 0$.
- Now Alice can always include cash in V, and so we can assume that $V = F + c$, where c is the minimum amount such that $\rho_B(F + c) \geq 0$.

Optimal Risk Transfer

- By cash translability, this implies $c = \rho_B(F)$. Hence we simply need to find F to minimise $\rho_A(W - F - \rho_B(F)) = \rho_A(W - F) + \rho_B(F)$
- We now define $\rho_A \square \rho_B(W) = \inf_F \{\rho_A(W - F) + \rho_B(F)\}$, which is called the inf-convolution of ρ_A and ρ_B.
- Clearly, our problem with risk transfer comes down to simply calculating $\rho_A \square \rho_B$.
- The value $\rho_A \square \rho_B(W)$ can be interpreted as the residual risk after all transactions.

Optimal Risk Transfer

Under a moderately restrictive assumption, we can find the following result.

Theorem

Suppose that Alice and Bob have risk measures which are scaled versions of each other, that is,

$$\rho_A(X) = \gamma \rho_B(\gamma^{-1} X).$$

Then if Alice has initial allocation W, the optimal transaction is given by

$$F^* = \frac{1}{1+\gamma} W.$$

This means that when the parties differ only in their treatment of the scale of the risks, they will proportionally partition a risk up between each other.

Optimal Risk Transfer

Proof.

We have

$$\begin{aligned}\rho_A(W - F) + \rho_B(F) &= \gamma\rho_B((1+\gamma)^{-1}W) + \gamma\rho_B((1+\gamma)^{-1}W)\\ &= (1+\gamma)\rho_B((1+\gamma)^{-1}W).\end{aligned}$$

By Borch's theorem, under our assumptions,

$$\rho_A \square \rho_B(W) = (1+\gamma)\rho((1+\gamma)^{-1}W),$$

and so we have achieved the optimum desired. $\square$

Euler Principle and Capital Allocation

Euler Principle and Capital Allocation

- In this final section we will talk about allocation of capital between risky sources, when using a positively homogeneous measure of risk (but not necessarily coherent).
- Consider an investor who can invest in a fixed set of d different investments, with losses represented by $L_1, ..., L_d$. Economically we could think of this in a few situations:
  - Performance measurement: For an institution with d different lines of business. We measure performance of each line by 'return/risk adjusted capital'. What is the correct 'risk adjusted capital' for each line of business?
  - Loan pricing: For a loan book manager with a portfolio of d loans. If the manager must maintain a capital reserve of $\rho(L) = \rho(L_1 + L_2 + ... + L_d)$, how much of that capital should be associated with each loan?
  - General investment: For an individual or an institution, the L_i are the losses corresponding to different investments. What is the riskiness of each investment, *considered as a part of the portfolio*?

Euler Principle and Capital Allocation

- The approach we shall take is
  1. Calculate the total risk capital $\rho(L) = \rho(\sum L_i)$, where ρ is a positively homogeneous risk measure.
  2. Allocate the capital $\rho(L)$ to each investment, that is, find quantities C_i associated with each L_i such that $\rho(L) = \sum C_i$.
- Formally, it is convenient if we consider what would happen if we modified the amounts in our portfolio L.
- That is, we consider portfolios $L(\lambda) = \sum \lambda_i L_i$, where $\lambda_i > 0$ and $\lambda = (\lambda_1, \lambda_2, ..., \lambda_d)$.
- Clearly $L(\mathbf{1}) = L$.

Euler Principle and Capital Allocation

- We can then define:

Definition

Let ρ be a positively homogeneous risk measure, then we can define $r_\rho : \mathbb{R}^d \to \mathbb{R}$, by $r_\rho(\lambda) = \rho(L(\lambda))$.

A map $\pi : \mathbb{R}^d \to \mathbb{R}^d$ is called a *per unit capital allocation* associated with r_ρ if for all λ,

$$r_\rho(\lambda) = \sum_i \lambda_i \pi_i(\lambda).$$

- Here $\pi_i(\lambda)$ has the interpretation of the amount capital allocated to each unit of investment i, when the total portfolio is $L(\lambda)$.

Euler Principle and Capital Allocation

- We can now use the result of Euler, which states that if r_ρ is positively homogeneous and differentiable at λ, then

$$r_\rho(\lambda) = \sum_i \lambda_i \frac{\partial r_\rho}{\partial \lambda_i}(\lambda)$$

which suggests the rule

$$\pi_i(\lambda) = \frac{\partial r_\rho}{\partial \lambda_i}(\lambda).$$

- This is sometimes known as *allocation by the gradient*, or the *Euler principle*.

Euler Principle and Capital Allocation

Example

- Consider a set of outcomes $L_1, ..., L_d$ with continuous loss distributions.
- Suppose we wish to allocate capital, where our risk is measured by the expected shortfall ES_α.
- Then we have

$$\begin{aligned}\pi_i(\lambda) &= \frac{\partial r_{\mathrm{ES}}}{\partial \lambda_i}(\lambda) = \frac{\partial}{\partial \lambda_i} E(L|L \geq \mathrm{VaR}_\alpha)\\ &= \frac{\partial}{\partial \lambda_i} E(\lambda_1 L_1 + \lambda_2 L_2 + ... + \lambda_d L_d | L \geq \mathrm{VaR}_\alpha)\\ &= E(L_i | L \geq \mathrm{VaR}_\alpha)\end{aligned}$$

- This is called the *expected shortfall contribution* of the investment possibility.
- This is a popular allocation principle in practice.

Euler Principle and Capital Allocation

- In many cases, we may not be able to calculate π_i analytically, however, as it is just a derivative of a function, there are various numerical methods which can be used, depending on the situation.
- The question remains why π_i is a 'correct' capital allocation principle, rather than just an easy mathematical tool. One answer comes from performance measurement...

Euler Principle and Capital Allocation

- Suppose r_ρ is differentiable on Λ.
- We say that a per-unit capital allocation principle π is *suitable for performance measurement* if, for all λ,

$$\operatorname{sign}\left(\frac{\partial}{\partial \lambda_i}\left(\frac{-E(L(\lambda))}{r_\rho(\lambda)}\right)\right) = \operatorname{sign}\left(\frac{-E(L_i)}{\pi_i(\lambda)} - \frac{-E(L(\lambda))}{r_\rho(\lambda)}\right)$$

- This means that (expected return/total risk capital) will increase when we increase the amount of any asset with asset-specific (expected return/allocated capital) more than the average.
- A result of Tasche (1999) shows that the only per-unit allocation principle suitable for performance measurement is the Euler principle.

Summary

1. Introduction
2. Loss distributions
3. Risk Measurement
   - Value at Risk
     - Problems with VaR
   - Expected Shortfall
   - Other common risk measures
4. Convex and Coherent Risk Measures
   - Importance sampling
   - Dual Representations
5. Dynamic Consistency
6. Optimal Risk Transfer
7. Euler Principle and Capital Allocation