Basic Concepts and Techniques of Risk Management

We introduce the basic concepts and techniques of risk management in these lecture notes. We will closely follow the content and notation of Chapter 2 of *Quantitative Risk Management* by *McNeil, Frey and Embrechts*. This chapter should be consulted if further details are required.

1 Risk Factors and Loss Distributions

Let Δ be a fixed period of time such as 1 day or 1 week. This will be the horizon of interest when it comes to measuring risk and calculating loss distributions. Let V_t be the value of a portfolio at time $t\Delta$ so that the portfolio *loss* between times $t\Delta$ and $(t+1)\Delta$ is given by

$$L_{t+1} \ := -\left(V_{t+1} - V_t\right). \tag{1}$$

Note that we treat a loss as a positive quantity so, for example, a negative value of L_{t+1} denotes a profit. The time[1] t value of the portfolio depends of course on the time t value of the securities in the portfolio. More generally, however, we may wish to define a set of d *risk factors*, $\mathbf{Z_t} := (Z_{t,1}, \ldots, Z_{t,d})$ so that V_t is a function of t and $\mathbf{Z_t}$. That is

$$V_t \ = \ f(t, \mathbf{Z_t})$$

for some function $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$. In a stock portfolio, for example, we might take the stock prices or some function of the stock prices as our risk factors. In an options portfolio, however, $\mathbf{Z_t}$ might contain stock factors together with implied volatility and interest rate factors. Now let $\mathbf{X_t} := \mathbf{Z_t} - \mathbf{Z_{t-1}}$ denote the change in the values of the risk factors between times t and $t-1$. Then we have

$$L_{t+1}(\mathbf{X_{t+1}}) \ = \ -\left(f(t+1, \mathbf{Z_t} + \mathbf{X_{t+1}}) - f(t, \mathbf{Z_t})\right) \tag{2}$$

and given the value of $\mathbf{Z_t}$, the distribution of L_{t+1} then depends only on the distribution of X_{t+1}.

1.1 Linear Approximations to the Loss Function

Assuming $f(\cdot,\cdot)$ is differentiable, we can use a first order Taylor expansion to approximate L_{t+1} with

$$\widehat{L}_{t+1}(\mathbf{X_{t+1}}) \ := \ -\left(f_t(t, \mathbf{Z_t})\Delta + \sum_{i=1}^{d} f_{z_i}(t, \mathbf{Z_t}) \ X_{t+1,i}\right) \tag{3}$$

where the f-subscripts denote partial derivatives. The first order approximation is commonly used when X_{t+1} is likely to be small. This is often the case when Δ is small, e.g. $1/365 \equiv 1$ day, and the market is not too volatile. Second and higher order approximations also based on Taylor's Theorem can also be used. It is important to note, however, that if X_{t+1} is likely to be very large then Taylor approximations of any order are likely to work poorly if at all.

[1] When we say time t we typically have in mind time $t\Delta$.

1.2 Conditional and Unconditional Loss Distributions

When we discuss the *distribution* of $\widehat{L}_{t+1}$ it is important to clarify exactly what we mean. In particular, we need to distinguish between the *conditional* and *unconditional* loss distributions. Consider the series $\mathbf{X_t}$ of risk factor changes and assume that they form a stationary[2] time series with stationary distribution $F_{\mathbf{X}}$. We also let $\mathcal{F}_t$ denote all information available in the system at time t, including $\{\mathbf{X_s} : s \leq t\}$ in particular. We then have the following two definitions.

Definition 1 *The* **unconditional** *loss distribution is the distribution of L_{t+1} given the time t composition of the portfolio and assuming the CDF of $\mathbf{X_{t+1}}$ is given by $F_{\mathbf{X}}$.*

Definition 2 *The* **conditional** *loss distribution is the distribution of L_{t+1} given the time t composition of the portfolio and conditional on the information in $\mathcal{F}_t$.*

It is clear that if the $\mathbf{X_t}$'s are *IID* then the conditional and unconditional distributions coincide. For long time horizons, e.g. $\Delta = 6$ months, we might be more inclined to use the unconditional loss distribution. However, for short horizons, e.g. 1 day or 10 days, then the conditional loss distribution is clearly the appropriate distribution. This would be particularly true in times of high market volatility when the unconditional distribution would bear little resemblance to the true conditional distribution.

Example 1 (A Stock Portfolio)
Consider a portfolio of d stocks with $S_{t,i}$ denoting the time t price of the i^{th} stock and λ_i denoting the number of units of this stock in the portfolio. If we take log stock prices as our factors then we obtain

$$X_{t+1,i} = \ln S_{t+1,i} - \ln S_{t,i}$$

and

$$L_{t+1} = -\sum_{i=1}^{d} \lambda_i S_{t,i} \left(e^{X_{t+1,i}} - 1 \right).$$

The linear approximation satisfies

$$\widehat{L}_{t+1} = -\sum_{i=1}^{d} \lambda_i S_{t,i} X_{t+1,i} = -V_t \sum_{i=1}^{d} \omega_{t,i} X_{t+1,i}$$

where $\omega_{t,i} := \lambda_i S_{t,i}/V_t$ is the i^{th} portfolio weight. If $X_{t+1,i}$ has mean vector μ and variance-covariance matrix Σ then we obtain $\mathsf{E}_t\left[\widehat{L}_{t+1}\right] = -V_t \omega' \mu$ and $\mathrm{Var}_t\left(\widehat{L}_{t+1}\right) = V_t^2\, \omega' \Sigma \omega$. Note that μ and Σ could refer to the first two moments of either the conditional or unconditional loss distribution. ∎

Example 2 (An Options Portfolio)
The Black-Scholes formula for the time t price of a European call option with strike K and maturity T on a non-dividend paying stock satisfies

$$C(S_t, t, \sigma) = S_t \Phi(d_1) - e^{-r(T-t)} K \Phi(d_2)$$

$$\text{where } d_1 = \frac{\log\left(\frac{S_t}{K}\right) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \qquad d_2 = d_1 - \sigma\sqrt{T-t}$$

$\Phi(\cdot)$ is the CDF of the standard normal distribution, S_t is the time t price of the underlying security and r is the risk-free interest rate. While the Black-Scholes model assumes a constant volatility, σ, in practice an implied volatility, $\sigma(K, T, t)$, that depends on the strike, maturity and current time, t, is observed in the market.

Consider now a portfolio of European options all on the same underlying security. The portfolio may also contain a position in the underlying security itself. If the portfolio contains d different options with a position of

[2] A time series, $\mathbf{X_t}$, is *strongly stationary* if $(\mathbf{X_{t_1}}, \ldots, \mathbf{X_{t_n}})$ is equal in distribution to $(\mathbf{X_{t_1+k}}, \ldots, \mathbf{X_{t_n+k}})$ for all $n, k, t_1, \ldots, t_n \in \mathbb{Z}^+$. Most risk factors are assumed to be stationary.

λ_i in the i^{th} option, then

$$L_{t+1} = -\lambda_0(S_{t+1} - S_t) - \sum_{i=1}^{d} \lambda_i \left(C(S_{t+1}, t+1, \sigma(K_i, T_i, t+1) - C(S_t, t, \sigma(K_i, T_i, t)) \right) \tag{4}$$

where λ_0 is the position in the underlying security. Note that by put-call parity we can assume that all options are call options. We can also use the linear approximation technique to approximate L_{t+1} in (4). This would result in a *delta-vega-theta* approximation. For derivatives portfolios, the linear approximation technique based on the first-order *Greeks* is often inadequate and second order approximations involving *gamma* and possibly *volga* and *vanna* are often employed.

For risk factors, we can again take the log stock prices but it is not clear how to handle the implied volatilities. There are several possibilities:

1. One possibility is to assume that the $\sigma(K, T, t)$'s simply do not change. This is not very satisfactory but is commonly assumed when historical simulation[3] is used to approximate the loss distribution and historical data on the changes in implied volatilities are not available.

2. Let each $\sigma(K, T, t)$ be a separate factor. In addition to approximately doubling the number of factors, a particular problem with this approach is that the implied volatilities are not free to move around independently. In fact the assumption of no-arbitrage imposes strong restrictions on how the implied volatility surface may move. It is therefore important to choose factors in such a way that those restrictions are easily imposed when we estimate the loss distribution.

3. There are d implied volatilities to consider and in principal this leads to d factors, albeit with restrictions on how these factors can move. We could use dimension reduction techniques such as *principal components analysis* (PCA) to identify just two or three variables that explain most of the movements in the volatility surface.

4. As an alternative to PCA, we could parameterize the volatility surface with just a few parameters and assume that only those parameters can move from one period to the next. The parameterization should be such that the no-arbitrage restrictions are easy to enforce. The obvious candidates would be factors that represent the *level*, *term structure* and *skew* of the implied volatility surface.

▮

Example 3 (A Bond Portfolio)
Consider a portfolio containing quantities of d different default-free zero-coupon[4] bonds where the i^{th} bond has price $P_{t,i}$, maturity T_i and face value equal[5] to 1. Let s_{t,T_i} denote the continuously compounded spot interest rate for maturity T_i so that

$$P_{t,i} = \exp(-s_{t,T_i}(T_i - t)).$$

If there are λ_i units of the i^{th} bond in the portfolio, then the total portfolio value is given by

$$V_t = \sum_{i=1}^{d} \lambda_i \exp(-s_{t,T_i}(T_i - t)).$$

Assume now that we only consider the possibility of parallel changes in the spot rate curve. Then if the spot curve moves by δ the portfolio loss satisfies

$$\begin{aligned} L_{t+1} &= -\sum_{i=1}^{d} \lambda_i \left(\exp(-(s_{t+\Delta,T_i} + \delta)(T_i - t - \Delta)) - \exp(-s_{t,T_i}(T_i - t)) \right) \\ &\approx -\sum_{i=1}^{d} \lambda_i \left(s_{t,T_i}(T_i - t) - (s_{t+\Delta,T_i} + \delta)(T_i - t - \Delta) \right). \end{aligned} \tag{5}$$

[3] See Section 3.
[4] There is no loss of generality here since we can decompose a default-free coupon bond into a series of zero-coupon bonds.
[5] There is also no loss of generality in assuming a face value of 1 since we can compensate by adjusting the quantity of each bond in the portfolio.

We therefore have a single risk factor, δ, that we can use to study the distribution of the portfolio loss. Note that if we assume that Δ is small so that $s_{t+\Delta,T_i} \approx s_{t,T_i}$ then we can further approximate the right-hand-side of (5) to obtain

$$L_{t+1} \;\approx\; \delta \sum_{i=1}^{d} \lambda_i (T_i - t). \tag{6}$$

We can use (how?) (5) or (6) to hedge parallel changes in the spot rate curve. Note that this is equivalent to hedging using the *Fisher-Weil duration* of the portfolio. Had we used a second-order approximation to the exponential function in deriving (5) then we would introduce the standard concept of *convexity*.

While hedging against parallel shifts in the spot rate curve often leads to good results, in practice the spot rate curve tends to move in a non-parallel manner and more sophisticated[6] hedging techniques produce superior results. ∎

---

2 Risk Measurement

2.1 Approaches to Risk Measurement

We now outline several approaches to the problem of risk measurement.

Notional Amount Approach

This approach to risk management defines the risk of a portfolio as the sum of the notional amounts of the individual positions in the portfolio. Each notional amount may be weighted by a factor representing the perceived riskiness of the position. While it is simple, it has many weaknesses. It does not reflect the benefits of diversification, does not allow for any *netting* and does not distinguish between long and short positions. Moreover, it is not always clear what the notional amount of a derivative is and so the notional approach can be difficult to apply when such derivatives are present in the portfolio.

Factor Sensitivity Measures

A factor sensitivity measure gives the change in the value of the portfolio for a given change in the factor. Commonly used examples include the Greeks of an option portfolio or the duration and convexity of a bond portfolio. These measures are often used to set position limits on trading desks and portfolios. They are generally not used for *capital adequacy* decisions as it is often difficult to aggregate these measures across different risk factors and markets.

For an example consider an option with time t price C that is written on an underlying security with price process S_t. We assume the time t price, C, is a function of only S_t and the implied volatility, σ_t. Then a simple application of Taylor's Theorem yields

$$\begin{aligned} C(S+\Delta S, \sigma+\Delta\sigma) \;&\approx\; C(S,\sigma) + \Delta S \frac{\partial C}{\partial S} + \frac{1}{2}(\Delta S)^2 \frac{\partial^2 C}{\partial S^2} + \Delta\sigma \frac{\partial C}{\partial \sigma} \\ &=\; C(S,\sigma) \;+\; \Delta S\,\delta \;+\; \frac{1}{2}(\Delta S)^2\,\Gamma \;+\; \Delta\sigma \text{ vega.} \end{aligned}$$

where we have omitted the dependence of the various quantities on t. We therefore obtain

$$\begin{aligned} \text{P\&L} \;&\approx\; \delta\Delta S \;+\; \frac{\Gamma}{2}(\Delta S)^2 \;+\; \text{vega } \Delta\sigma \\ &=\; \text{delta P\&L} \;+\; \text{gamma P\&L} \;+\; \text{vega P\&L}\,. \end{aligned} \tag{7}$$

[6]For example, methods based on *principal components analysis* (PCA) are often used in practice. We will study PCA when we study dimension reduction techniques.

When $\Delta\sigma = 0$, we obtain obtain the well-known **delta-gamma approximation** which is often used, for example, in historical Value-at-Risk (VaR) calculations. Note that we can also write (7)

$$\begin{aligned} \text{P\&L} \quad &\approx \quad \delta S \left(\frac{\Delta S}{S}\right) + \frac{\Gamma S^2}{2}\left(\frac{\Delta S}{S}\right)^2 + \text{vega } \Delta\sigma \\ &= \quad \text{ESP} \times \text{Return} \; + \; \$ \text{ Gamma} \times \text{Return}^2 \; + \; \text{vega } \Delta\sigma \end{aligned} \tag{8}$$

where **ESP** denotes the **equivalent stock position** or **"dollar" delta**. Note that it is easy to extend this calculation to a portfolio of options on the same underlying security. It is also straightforward[7] to extend these ideas to derivatives portfolios written with many different underlying securities.

Depending on the particular asset class, investors / traders / risk managers should always know their exposure to the Greeks, i.e. dollar delta, dollar gamma and vega etc. It is also very important to note that approximations such as (8) are **local** approximations as they are based (via Taylor's Theorem) on "small" moves in the risk factors. These approximations can and indeed do break down in violent markets where changes in the risk factors can be very large.

Scenario or Stress Approach

The scenario approach defines a number of scenarios where in each scenario the various risk factors are assumed to have moved by some fixed amounts. For example, a scenario might assume that all stock prices have fallen by 10% and all implied volatilities have increased by 5 percentage points. Another scenario might assume the same movements but with an additional steepening of the volatility surface. A scenario for a credit portfolio might assume that all credit spreads have increased by some fixed absolute amount, e.g. 100 basis points, or some fixed relative amount, e.g. 10%. The risk of a portfolio could then be defined as the maximum loss over all of the scenarios that were considered. A particular advantage of this approach is that it does not depend on probability distributions that are difficult to estimate. In that sense, **the scenario approach to risk management is very robust and should play a vital role in any financial risk management operation**.

Underlying	SPX Index								
	Underlying and Volatility Stress Table								
Sum of PnL	Vol Stress								
Underlying Stress	-10	-5	-2	-1	0	1	2	5	10
-20	13,938	11,774	10,631	10,277	9,936	9,608	9,293	8,419	7,183
-10	6,109	4,946	4,436	4,291	4,158	4,035	3,922	3,634	3,296
-5	1,831	1,652	1,637	1,643	1,654	1,670	1,689	1,766	1,946
-2	(314)	89	356	447	539	631	723	1,001	1,461
-1	(920)	(338)	15	132	248	363	478	816	1,361
0	(1,463)	(714)	(280)	(139)	0	137	273	668	1,293
1	(1,939)	(1,035)	(527)	(363)	(203)	(45)	110	559	1,259
2	(2,346)	(1,300)	(723)	(539)	(359)	(182)	(9)	489	1,258
5	(3,125)	(1,744)	(1,003)	(769)	(541)	(318)	(102)	518	1,460
10	(2,921)	(1,297)	(423)	(146)	123	385	641	1,372	2,483
20	2,344	3,559	4,272	4,506	4,738	4,967	5,194	5,860	6,919

Figure 1: P&L for an Options Portfolio on SP5X under Stresses to Underlying and Implied Volatility

Figure 1 shows the P&L under various scenarios of an options portfolio with the S&P500 as the underlying security. The vertical axis represents *percentage* shifts in the price of the underlying security whereas the horizontal axis represents *absolute* changes in the implied volatility of each option in the portfolio. For example, we see that if the S&P500 were to fall by 20% and implied volatilities were to all rise by 5 percentage points, then the portfolio would gain 8.419 million dollars (assuming that the numbers in Figure 1 are expressed in units

[7] And we shall do this when we discuss dimension reduction methods in a later set of lecture notes.

of $1,000$ dollars). When constructing scenario tables as in 1 we can use approximations like (8) to check for internal consistency and to help identify possible bugs in the software.

While scenario tables are a valuable source of information there are many potential pit-falls associated with using them. These include:

1. *Identifying the relevant risk factors*

   While it is usually pretty clear what the main risk factors for a particular asset class are, it is quite possible that a portfolio has been constructed so that it is approximately neutral to changes in those risk factors. Such a portfolio might then only have (possibly very large) exposures to secondary risk factors. It is important then to include shifts in these secondary factors in any scenario analysis. The upshot is that the the relevant risk factors *depend* on the specific portfolio under consideration rather than just the asset class of the portfolio.

2. *Identifying "reasonable" shifts for these risk factors*

   For example, we may feel that a shift of -10% is plausible for the S&P 500 because we know from experience that such a move, while extreme, is indeed possible in a very volatile market. But how do we determine plausible shifts for less transparent risk factors? The answer typically lies in the use of statistical techniques such as PCA, extreme-value theory, time series methods, common sense(!) etc.

A key role of any risk manager then is to understand what scenarios are plausible and what scenarios are not. For example, in a crisis we would expect any drop in the price of the underlying security to be accompanied by a rise in implied volatilities. We would therefore pay considerably less attention to the numbers in the upper left quadrant of Figure 1.

Measures Based on Loss Distribution

Many risk measures such as *value-at-risk* (VaR) or *conditional value-at-risk* (CVaR) are based on the loss distribution of the portfolio. Working with loss distributions makes sense as the distribution contains all the information you could possibly wish to know about possible losses. A loss distribution implicitly reflects the benefits of netting and diversification. Moreover it is easy to compare the loss distribution of a derivatives portfolio with that of a bond or credit portfolio, at least when the same time horizon is under consideration. However, it must be noted that it may be very difficult to estimate the loss distribution. This may be the case for a number of reasons including a lack of historical data, non-stationarity of risk-factors and poor model choice among others.

2.2 Value-at-Risk

Value-at-Risk (VaR) is the most widely used risk measure in the financial industry. Despite the many weaknesses of VaR, financial institutions are required to use it under the Basel II capital-adequacy framework. In addition, many institutions routinely report their VaR numbers to shareholders, investors or regulatory authorities. VaR is a risk measure based on the loss distribution and our discussion will not depend on whether we are dealing with the conditional or unconditional loss distribution. Nor will it depend on whether we are using the true loss distribution or some approximation to it. We will assume that the horizon, Δ, has been fixed and that the random variable L represents the loss on the portfolio under consideration over the time interval Δ. We will use $F_L(\cdot)$ to denote the cumulative distribution function (CDF) of L. We first define the *quantiles* of a CDF.

Definition 3 *Let $F : \mathbb{R} \to [0, 1]$ be an arbitrary CDF. Then for $\alpha \in (0, 1)$ the α-quantile of F is defined by*

$$q_\alpha(F) \;:=\; \inf\{x \in \mathbb{R} \;:\; F(x) \geq \alpha\}.$$

Note that if F is continuous and strictly increasing, then $q_\alpha(F) = F^{-1}(\alpha)$. For a random variable L with CDF $F_L(\cdot)$, we will often write $q_\alpha(L)$ instead of $q_\alpha(F_L)$. Using the fact that any CDF is by definition right-continuous, we immediately obtain the following result.

Lemma 1 *A point $x_0 \in \mathbb{R}$ is the α-quantile of F_L if and only if (i) $F_L(x_0) \geq \alpha$ and (ii) $F_L(x) < \alpha$ for all $x < x_0$.*

Definition 4 *Let $\alpha \in (0,1)$ be some fixed confidence level. Then the VaR of the portfolio loss at the confidence level, α, is given by $VaR_\alpha := q_\alpha(L)$, the α-quantile of the loss distribution.*

Example 4 (The Normal and t Distributions)
Because the normal and t CDFs are both continuous and strictly increasing, it is straightforward to calculate their VaR_α. If $L \sim \text{N}(\mu, \sigma^2)$ then

$$\text{VaR}_\alpha \;=\; \mu + \sigma\Phi^{-1}(\alpha) \quad \text{where } \Phi \text{ is the standard normal CDF.} \tag{9}$$

By the previous lemma, this follows if we can show that $F_L(\text{VaR}_\alpha) = \alpha$. But this follows immediately from (9).

If $L \sim \text{t}(\nu, \mu, \sigma^2)$ so that $(L - \mu)/\sigma$ has a standard t distribution with $\nu > 2$ degrees-of-freedom, then

$$\text{VaR}_\alpha \;=\; \mu + \sigma t_\nu^{-1}(\alpha) \quad \text{where } t_\nu \text{ is the CDF for the t distribution with } \nu \text{ degrees-of-freedom.}$$

Note that in this case we have $\text{E}[L] = \mu$ and $\text{Var}(L) = \nu\sigma^2/(\nu - 2)$. ∎

VaR has several weaknesses:

1. VaR attempts to describe the entire loss distribution with a single number and so significant information is not captured in VaR. This criticism does of course apply to all scalar risk measures. One way around this is to report VaR_α for several different values of α.

2. There is significant model risk attached to VaR. If the loss distribution is heavy-tailed, for example, but a normal distribution is assumed, then VaR_α will be severely underestimated as α approaches 1. A fundamental problem with VaR and other risk measures based on the loss distribution is that it can be very difficult to estimate the loss distribution. Even when there is sufficient historical data available there is no guarantee that the historical data-generating process will remain unchanged in the future. For example, in the buildup to the sub-prime housing crisis it was clear that the creation and selling of vast quantities of structured credit products could change the price dynamics of these and related securities. And of course, that is precisely what happened.

3. VaR is not a *sub-additive* risk measure so that it doesn't lend itself to *aggregation*. For example, let $L = L_1 + L_2$ be the total loss associated with two portfolios, each with respective losses, L_1 and L_2. Then

$$q_\alpha(F_L) \;>\; q_\alpha(F_{L_1}) \;+\; q_\alpha(F_{L_2}) \quad \text{is possible.} \tag{10}$$

   In the risk literature this is viewed as being an undesirable property as we would expect some diversification benefits when we combine two portfolios together. Such a benefit would be reflected by the combined portfolio having a *smaller* risk measure than the sum of the two individual risk measures.

An advantage of VaR is that it is generally easier[8] to estimate. This is true when it comes to quantile estimation in general as quantiles are not very sensitive to outliers. This is not true of other risk measures such as CVaR which we discuss below. Despite this fact, it becomes progressively more difficult to estimate VaR_α as α gets closer to 1. Extreme Value Theory (EVT) can be useful in these circumstances, however, and we will return to EVT in later lectures.

The value of Δ that we use in practice generally depends on the application. For credit, operational and insurance risk, Δ is often on the order of 1 year. For financial risks, however, typical values of Δ are on the order of days, with 1 and 10 days being very common.

[8]This assumes of course that we have correctly specified the appropriate probability model so that the second weakness above is not an issue. This assumption is often not justified!

2.3 Expected Shortfall (ES) / Conditional Value-at-Risk (CVaR)

We now define expected shortfall which is also commonly referred to as conditional value-at-risk (CVaR).

Definition 5 *For a portfolio loss, L, satisfying $E[|L|] < \infty$ the expected shortfall at confidence level $\alpha \in (0,1)$ is given by*

$$ES_\alpha := \frac{1}{1-\alpha} \int_\alpha^1 q_u(F_L)\, du.$$

The relationship between ES_α and VaR_α is therefore given by

$$\mathrm{ES}_\alpha := \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_u(L)\, du$$

from which it is clear that $\mathrm{ES}_\alpha(L) \geq \mathrm{VaR}_\alpha(L)$. A more well known representation of $\mathrm{ES}_\alpha(L)$ holds when F_L is continuous.

Lemma 2 *If F_L is a continuous CDF then*

$$ES_\alpha := \frac{E[L;\ L \geq q_\alpha(L)]}{1-\alpha} = E[L \mid L \geq VaR_\alpha]\,. \tag{11}$$

Proof: See Lemma 2.13 in *McNeil, Frey and Embrechts*. The proof is straightforward and relies on the standard result that the random variable $F_L^{-1}(U)$ has distribution F_L when $U \sim \mathrm{Uniform}(0,1)$.

Example 5 (Expected Shortfall for a Normal Distribution)

We can use (11) to compute the expected shortfall of an $\mathrm{N}(\mu, \sigma^2)$ random variable. In particular we can check that

$$\mathrm{ES}_\alpha = \mu + \sigma \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha} \tag{12}$$

where $\phi(\cdot)$ is the PDF of the standard normal distribution. ∎

Example 6 (Expected Shortfall for a t Distribution)

Let $L \sim \mathrm{t}(\nu, \mu, \sigma^2)$ so that $\tilde{L} := (L-\mu)/\sigma$ has a standard t distribution with $\nu > 2$ degrees-of-freedom. Then as in the previous example it is easy to see that $\mathrm{ES}_\alpha(L) = \mu + \sigma \mathrm{ES}_\alpha(\tilde{L})$. It is straightforward using direct integration to check that

$$\mathrm{ES}_\alpha(\tilde{L}) = \frac{g_\nu\left(t_\nu^{-1}(\alpha)\right)}{1-\alpha} \left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu - 1}\right) \tag{13}$$

where $t_\nu(\cdot)$ and $g_\nu(\cdot)$ are the CDF and PDF, respectively, of the standard t distribution with ν degrees-of-freedom. ∎

Exercise 1 *Prove (13).*

Remark 1 *The t distribution has been found to be a much better model of stock (and other asset) returns than the normal model. In empirical studies, values of ν around 5 or 6 are often found to fit best.*

The Shortfall-to-Quantile Ratio

One method of comparing VaR_α and ES_α is to consider their ratio as $\alpha \to 1$. It is not too difficult to see that in the case of the normal distribution, $\mathrm{ES}_\alpha/\mathrm{VaR}_\alpha \to 1$ as $\alpha \to 1$. However, $\mathrm{ES}_\alpha/\mathrm{VaR}_\alpha \to \nu/(\nu-1) > 1$ in the case of the t distribution with $\nu > 1$ degrees-of-freedom. Figure 2 displays the shortfall-to-quantile ratio for the normal distribution and t distribution with 4 degrees of freedom.

p	VaR.normal	VaR.t4	ES.normal	ES.t4	Ratio.normal	Ratio.t4
0.900000	162.10	137.13	221.99	223.55	1.37	1.63
0.950000	208.06	190.68	260.91	286.47	1.25	1.50
0.975000	247.92	248.33	295.71	357.19	1.19	1.44
0.990000	294.26	335.14	337.13	466.94	1.15	1.39
0.995000	325.82	411.80	365.81	565.71	1.12	1.37
0.999000	390.89	641.59	425.91	866.36	1.09	1.35
0.999900	470.42	1165.77	500.71	1560.43	1.06	1.34
0.999990	539.47	2086.89	566.52	2785.93	1.05	1.33
0.999999	601.27	3718.84	625.92	4960.36	1.04	1.33

Figure 2: Shortfall-to-Quantile Ratio for Normal and t Distributions

3 Standard Techniques for Risk Measurement

We now discuss some of the principal techniques for estimating the loss distribution.

3.1 Historical Simulation

Instead of using some probabilistic model to estimate the distribution of $L_{t+1}(\mathbf{X_{t+1}})$, we could estimate the distribution using a historical simulation. In particular, if we know the values of $\mathbf{X_{t-i+1}}$ for $i = 1, \ldots, n$, then we can use this data to create a set of historical losses:

$$\{\tilde{L}_i \ := \ L_{t+1}(\mathbf{X_{t-i+1}}) \ \text{ for } \ i = 1, \ldots, n\}.$$

$\tilde{L}_i$ is the loss on the portfolio that would occur if the changes in the risk factors on date $t-i+1$ were to recur. In order to calculate the value of a given risk measure we simply assume that the distribution of $L_{t+1}(\mathbf{X_{t+1}})$ is discrete and takes on each of the values $\tilde{L}_i$ with probability $1/n$ for $i = 1, \ldots, n$. That is, we use the *empirical distribution* of the $\mathbf{X_t}$'s to determine the loss distribution.

If we wish to compute VaR, then we can do so using the appropriate quantiles of the $\tilde{L}_i$'s. For example suppose $\tilde{L}_{j,n}$ is the j^{th} reverse order statistic of the $\tilde{L}_i$'s so that

$$\tilde{L}_{n,n} \leq \cdots \leq \tilde{L}_{1,n}.$$

Then a possible estimator of $\text{VaR}_\alpha(L_{t+1})$ is $\tilde{L}_{[n(1-\alpha)],n}$ where $[n(1-\alpha)]$ is the largest integer not exceeding $n(1-\alpha)$. The associated ES could be estimated by averaging $\tilde{L}_{[n(1-\alpha)],n}, \ldots, \tilde{L}_{1,n}$.

The historical simulation approach is generally difficult to apply for derivative portfolios as it is often the case that historical data on at least some of the associated risk factors is not available. If the risk factor data is available, however, then the method is easy to apply as it does not require any statistical estimation of multivariate distributions. It is also worth emphasizing that the method estimates the *unconditional* loss distribution and not the conditional loss distribution. As a result, it is likely to be very inaccurate in times of market stress. Unfortunately these are precisely the times when you are most concerned with obtaining accurate risk measures.

3.2 Monte-Carlo Simulation

The Monte-Carlo approach is similar to the historical simulation approach except now we use some parametric distribution for the change in risk factors to generate sample portfolio losses. The distribution (conditional or unconditional) of the risk factors is estimated and, assuming we know how to simulate from this distribution, we can generate a number of samples, m say, of portfolio losses. We are free to make m as large as possible, subject to constraints on computational time. Variance reduction methods are often employed to obtain improved estimates of the required risk measures. While Monte-Carlo is an excellent tool, it is only as good as the model used to generate the data: if the distribution of $\mathbf{X_{t+1}}$ that is used to generate the samples is poor, then the Monte-Carlo samples will be of little value.

3.3 Variance-Covariance Approximations

In the variance-covariance approach we assume that $\mathbf{X_{t+1}}$ has a multivariate normal distribution so that

$$\mathbf{X_{t+1}} \;\sim\; \mathsf{MVN}\left(\mu, \mathbf{\Sigma}\right).$$

We also assume that the linear approximation in (3) is sufficiently accurate. Writing $\widehat{L}_{t+1}(\mathbf{X_{t+1}}) = -(c_t + \mathbf{b_t}^\top X_{t+1})$ for a constant scalar, c_t, and constant vector, $\mathbf{b_t}$, we therefore obtain that

$$\widehat{L}_{t+1}(\mathbf{X_{t+1}}) \;\sim\; \mathsf{N}\left(-c_t - \mathbf{b_t}^\top \mu,\; \mathbf{b_t}^\top \mathbf{\Sigma} \mathbf{b_t}\right).$$

We can now use our knowledge of the normal distribution to calculate any risk measures of interest. Note that this technique can be either conditional or unconditional, depending on how μ and $\mathbf{\Sigma}$ are estimated.

The strength of this approach is that it provides a straightforward analytically tractable method of determining the loss distribution. It has several weaknesses, however. Risk factor distributions are often[9] *fat-* or *heavy-tailed* but the normal distribution is light-tailed. As a result, we are likely to underestimate the frequency of extreme movements which in turn can lead to seriously underestimating the risk in the portfolio. This problem is generally easy to overcome as there are other multivariate distributions that are also *closed under linear operations.* In particular, if we assume that $\mathbf{X_{t+1}}$ has a multivariate t distribution so that

$$\mathbf{X_{t+1}} \sim \; t\left(\nu, \mu, \mathbf{\Sigma}\right)$$

then we obtain

$$\widehat{L}_{t+1}(\mathbf{X_{t+1}}) \;\sim\; t\left(\nu, -c_t - \mathbf{b_t}^\top \mu,\; \mathbf{b_t}^\top \mathbf{\Sigma} \mathbf{b_t}\right).$$

where ν is the degrees-of-freedom of the t-distribution. A more serious problem with this approach is the assumption that the linear approximation will work well. This is generally not true for portfolios of derivative securities or other portfolios where the portfolio value is a non-linear function of the risk-factors. It is also problematic when the time horizon Δ is large. We can also overcome this problem to some extent by using quadratic approximations to L_{t+1} instead of linear approximations and then using Monte-Carlo to estimate the risk measures of interest.

3.4 Evaluating the Techniques for Risk Measurement

An important task of any risk manger is to constantly evaluate the risk measures that are being reported. For example, if the daily 95% VaR is reported then we should see the daily losses exceeding the reported VaR approximately 95% of the time. Indeed, suppose the reported VaR numbers are calculated correctly and let Y_i be the indicator function denoting whether or not the portfolio loss in period i exceeds VaR_i, the corresponding VaR for that period. That is

$$Y_i \;=\; \begin{cases} 1, & L_i \geq \mathsf{VaR}_i; \\ 0, & \text{otherwise}. \end{cases}$$

If we assume the Y_i's are *IID*, then $\sum_{i=1}^n Y_i$ should be $\mathsf{Bin}(n, .05)$. We can use standard statistical tests to see if this is indeed the case. Similar tests can be constructed for ES and other risk measures.

More generally, it is important to check that the true reported portfolio losses are what you would expect given the composition of the portfolio and the realized change in the risk factors over the previous period. For example, if you know the delta and vega of your portfolio at time t_{i-1} and you observe the change in the underlying and volatility factors between t_{i-1} and t_i, then the actual realized gains or losses should be consistent with the delta and vega numbers. If they are not, then further investigation is required. It may be

[9] We will define *light-* and *heavy-tailed* distributions formally when we discuss multivariate distributions.

that (i) the moves required second order terms such as gamma, vanna or volga or (ii) other risk factors influencing the portfolio value also changed or (iii) the risk factor moves were very large and cannot be captured by any Taylor series approximation – see the related discussion below (8). Of course one final possibility is that there's a bug in the risk management system!

---

4 Other Considerations

In this section we briefly describe some other issues that frequently arise in practice. We will return to some of these issues at various points throughout the course.

4.1 Risk-Neutral and Data-Generating Measures

It is worth emphasizing that in risk management, we generally want to generate loss samples using the true probability measure, P. This is in contrast to security pricing where we use a risk-neutral or equivalent martingale measure, Q. Sometimes, however, we need to use both probability measures. Imagine for example, a portfolio of complex derivatives that we can only price using Monte-Carlo. In order to estimate the loss distribution, we would then (i) generate samples of the change in risk factors using P and (ii) for each such sample, compute the resulting value of the portfolio by running a Monte-Carlo using Q.

4.2 Data Risk

When using historical data to either estimate probability distributions or run a historical simulation analysis, for example, it is important to realize that the data may be biased in several ways. For example, survivorship bias is a common problem: the stocks that exist today are more likely to have done well than the stocks that no longer exist. So when we run a historical simulation to evaluate the risk of a long stock portfolio, say, we are implicitly omitting the worst returns from our analysis and therefore introducing a bias into our estimates of the various risk measures. We can get around this problem if we can somehow include the returns of stocks that no longer exist in our analysis. We also need to be mindful of biases that arise due to *data mining* or *data snooping* whereby people trawl through data-sets looking for spurious data patterns to justify some investment strategy. The *newsletter scam* is a classic example of what can go wrong! In some cases, *selection bias* may be present inadvertently. Regardless, one should always be alert to the possibility of these biases.

4.3 Multi-Period Risk Measures and Scaling

It is often necessary to report risk measures with a particular investment horizon in mind. For example, many hedge funds are required by investors to report their 10-day 95% VaR and the question arises as to how this should be calculated. They could of course try to estimate the 10-day VaR directly. However, as hedge funds typically calculate a 1-day 95% VaR anyway, it would be convenient if they could somehow scale the 1-day VaR in order to estimate the 10-day VaR. Before considering this problem, it is worth emphasizing one of the implicit assumptions that we make when calculating our risk measures:

The Constant Portfolio Composition Assumption: The risk measures are calculated by assuming that the portfolio composition does not change over the horizon $[t\Delta, (t+1)\Delta]$.

Indeed regulatory requirements typically require that VaR be calculated under this assumption. While this is usually ok for small values of Δ, e.g. 1 day, it makes little sense for larger values of Δ, e.g. 10 days, during which the portfolio composition is very likely to change. In fact in some circumstance the portfolio composition must change. Consider an options portfolio where some of the options are due to expire before the interval, Δ, elapses. If the options are not *cash settled*, then any in-the-money options will be exercised resulting automatically in new positions in the underlying securities. Regardless of how the options settle, the returns on the portfolio around expiration will definitely not be *IID*. It can therefore be very problematic scaling a 1-day

VaR into a 10-day VaR in these circumstances. More generally, it may not even make sense to consider a 10-day VaR when you know the portfolio composition is likely to change dramatically.

When portfolio returns are IID and factor changes are normally distributed then a *square-root* scaling rule can be justified. In this case, for example, we could take the 10-day VaR equal to $\sqrt{10}$ times the 1-day VaR. This follows from Example 4 with $\mu = 0$ and the assumption of continuously compounded returns. But this kind of scaling can lead to very misleading results when portfolio returns are not IID normal, even when the portfolio composition does not change.

4.4 Model Risk

Model risk can arise when calculating risk measures or estimating loss distributions. For example, if we are considering a portfolio of European options then there should be no problem in using the Black-Scholes formula to price these options as long as we use appropriate volatility risk factors with appropriate distributions. But suppose instead that the portfolio contains barrier options and that we use the corresponding Black-Scholes model for pricing these options. Then we are likely to run into several difficulties as the model is notoriously bad. It is possible, for example, that reported risk measures such as delta or vega will have the wrong sign! Moreover, the range of possible barrier option prices is limited by the assumption of constant volatility. A more sophisticated model would allow for a greater range of prices, and therefore losses. A more obvious example of model risk is using a light-tailed distribution to model risk factors when in fact a heavy-tailed distribution should be used.

4.5 Data Aggregation

A particularly important topic is the issue of data aggregation. In particular, how do we aggregate risk measures from several trading desks or several units within a firm into one single aggregate measure of risk? This issue arises when a firm is attempting to understand its overall risk profile and determining its capital adequacy requirements as mandated by regulations. One solution to this problem is to use *sub-additive* risk measures that can be added together to give a more conservative measure of aggregate risk. We will return to this topic when we discuss *coherent measures of risk*.

Another solution is to simply ignore the risk measures from the individual units or desks. Instead we could try to directly calculate a firm-wide risk number. This might be achieved by specifying scenarios that encompass the full range of risk-factors to which the firm is exposed or using the variance-covariance approach with many risk factors. The resulting mean vector, μ, and variance-covariance matrix, Σ, may be very high-dimensional, however. For a small horizon, Δ, it is generally ok to set $\mu = 0$ but it will often be very difficult to estimate Σ due to its high dimensionality. In that case *factor models* and other techniques can often be used to construct more reliable estimates of Σ.

4.6 Liquidity Risk

Liquidity risk is a very important source of risk particularly if we need to unwind a portfolio. We may know the "fair" market price of the portfolio in a given scenario but will we be able to transact or unwind at that price should the scenario occur? If we expect bid-ask spreads to be tight in that scenario then it is indeed reasonable to assume that we can unwind at approximately the "fair" price. But if the scenario is an extreme scenario then it's possible the bid-ask spreads will have widened substantially so that transacting at the fair price will be impossible. This situation can be exacerbated if many firms hold similar portfolios and are all "rushing for the exits" at the same time. In that event enormous losses can be incurred in the act of unwinding the portfolio. This is the "crowded trade" phenomenon and it was a source of many of the extreme losses that occurred during the 2008 crisis.

Liquidity risk is difficult to model but does need to be accounted for if many firms also hold a similar position and we expect that we may need to unwind our position should the scenario occur. One way to account for this is to model the bid-ask spread in the various scenarios and use these to compute an "unwind" P&L in addition to the market P&L (which does not assume unwinding is necessary and will take the fair price to be the mid-point of the bid-ask spread.)

4.7 P&L Attribution

Risk management is typically about understanding the risks in a portfolio in anticipation of *future* changes in the underlying risk factors. In contrast, P&L attribution is a *backward*-looking process where the goal is to understand what risk factor (changes) have contributed to the P&L that has been realized between times t and $t+1$. There are at least a couple of approaches to solving this problem. Consider, for example, a European option on a stock or index:

1. One approach is based on local approximations via Taylor's Theorem. Let $\text{Return}_{t,t+1}$ and $\Delta\sigma_{t,t+1}$ denote the changes in risk-factors (in this example the underlying return and change in implied volatility) between times t and $t+1$. We can then use (8) to obtain a *predicted* P&L, $\text{P\&L}_{t,t+1}^{pred}$, as a function of the Greeks at time t and the changes in risk-factors:

$$\text{P\&L}_{t,t+1}^{pred} \quad \approx \quad \text{ESP}_t \times \text{Return}_{t,t+1} \; + \; \$ \; \text{Gamma}_t \times \text{Return}_{t,t+1}^2 \; + \; \text{vega}_t \Delta\sigma_{t,t+1}. \tag{14}$$

   Note that $\text{P\&L}_{t,t+1}^{pred}$ is the P&L we would predict for the period $[t, t+1]$ standing at time t if we knew what the changes in risk factors were going to be over that period. Once time $t+1$ arrives we can observe the *realized* P&L, $\text{P\&L}_{t,t+1}^{real}$. We can then write this as

$$\text{P\&L}_{t,t+1}^{real} = \text{P\&L}_{t,t+1}^{pred} + \epsilon_{t,t+1} \tag{15}$$

   where $\epsilon_{t,t+1}$ is defined to be whatever value makes (15) true. Ideally $\epsilon_{t,t+1}$ would be very small so that the predicted P&L given by (14) accounts for almost all of the realized P&L. Note that we can then use (14) to *attribute* the P&L to delta (ESP), gamma and vega respectively. If $\epsilon_{t,t+1}$ is large, however, then the attribution has performed poorly. In that case we would need to include other risk factors, e.g. theta for the (deterministic) passage of time or a dividend factor to account for changes in expected dividends, or else include other second-order sensitivities such as volga, vanna etc. Even then, however, because approximations like (14) are only valid for "small" changes in the risk factors, it is quite possible that this approach will fail in extremely volatile markets such as those encountered at the height of the financial crisis in 2008.

2. An alternative is to use a more global approach that is not based on the (mathematical) derivatives, i.e. the Greeks, that are required by a Taylor series approximation. Such an approach is coarser than the local approach but it will work in extremely volatile markets as long as all important risk factors are included in the analysis.