1 Chapter 1: Risk and CAPM model

Risk versus return : Investors require a trade off between risk and **expected return** (note: not between risk and actual return). For example, a riskless bond has a return of 5 % while a risky asset may have a distribution of : 0.05 probability of 50 % return, 0.25 probability of 30 % return, 0.4 probability of 10 % return, 0.25 probability of - 10 % return and 0.05 % pf - 30 % return. This asset has an expected return of 10 % which is higher than the riskless return. A popular risk measure is the standard deviation of the annual return.

Portfolio return: Consider a portfolio of 2 assets with weights w_i and returns R_i. The porfolio expected return is $\mu_p = w_1\mu_1 + w_2\mu 2$. The standard deviation is

$$\sigma_P = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2\rho w_1 w_2 \sigma_1\sigma_2}.$$

There might be combination of $w_1, w_2, w_1 + w_2 = 1$ so that $\mu_P > \mu_1$ and $si_P < \sigma_1$ even if $\rho > 0$. That is diversification improves the risk / return combination of an asset.

Efficient frontier (of risky asset only): It is the combination of (σ, μ) of **risky** assets that forms a concave curve so that there is no other investment that dominates a point on the frontier (in the sense that it has both a higher expected return and lower standard deviation). The concavity of the EFF is because of the diminishing marginal return in risk : each unit of risk added to a portfolio gains a smaller and smaller amount of return.

Remark: In terms of the EFF, the further along a portfolio risk / return is in the direction of the north west corner (low SD, high return) the more desirable. This observation gives an intuitive explanation to the EFF of risky and riskless assets below.

Efficient frontier (of risky and riskless asset) : If we also include riskless asset then the efficient frontier is the tangent line that crosses the point $(0, R_F)$ and the efficient frontier (this tangent line is the furthest line in the north west corner that we can achieve using a combination of a riskless asset and a risky asset). The tangent point is $M = (\sigma_M, R_M)$ is referred to as the market portfolio. The reason is as followed : If we include $w_1 = 1 - w, w_2 = w$ of the riskless asset and the market portfolio then the return is

$$\mu_P = (1 - w)R_F + wE(R_M) = R_F + w(E(R_M) - R_F)$$

and the standard deviation is

$$\sigma_P = w\sigma_M.$$

Thus as w ranges over all possible non-negative values, we see that the combination traces a line that crosses $(0, R_F)$ (when $w = 0$) and (σ_M, R_M) (when $W = 1$). w is positive because $E(R_M) > R_F$ by property of risky asset, so the term $R_F + w(E(R_M) - R_F)$ will be less than R_F if $w < 0$. This means that one always want to long the market portfolio and then borrow or lend at the risk free rate to obtain a portfolio on the EFF.

The market portfolio M : One can argue that it must be the portfolio of **all** risky asset. Indeed if an asset is not in M then no investor would hold it. Its price will drop and its return will increase, which makes it become part of the portfolio M again. A stronger principle holds: the proportion of investment of the assets in the portfolio M must be the same as the proportion of their investments in the economy. This is to ensure a balance between supply and demand. This justifies the name of the market portfolio.

The CAPM model: The above observation shows that the market portfolio should play a key role in the expected return investors require for individual investment. A common procedure is to use historical data and regression to determine a best-fit linear relationship between returns from an investment and the market portfoio. Specifically:

$$R = \alpha + \beta R_M + \epsilon.$$

Thus there are two components to the risk in the investment's return: βR_M, which is referred to as systemic risk and ε which is unrelated to R_M and referred to as nonsystemic risk (e.g. the risk of one's company having a fire etc.). The ε of different investments are assumed to be independent of one another. Thus by holding a large portfolio, the non-systemic risk can be assumed to be diversified away. Thus an investor should not require extra expected return over the risk free rate for bearing nonsystemic risk.

By choosing different value of β, the investor trades off systemic risk and expected return in different ways. When $\beta = 0$, there is no systemic risk and thus $E(R) = R_F$. When $\beta = 1$, we have the same systemic risk as the market portfolio and thus $E(R) = E(R_M)$. In general we have

$$E(R) = R_F + \beta(E(R_M) - R_F).$$

This is the CAPM model and thus again reflects the fact that the investor chooses investment along the EFF. The excess expected return over the risk free rate required on the investment is *beta* times the excess expeced return on the market portfolio. β is referred to as the beta of the investment, which measures the sensitivity of the investment return from the market portfolio. Since

$$\rho\sigma_R\sigma_M = Cov(R, R_M) = Cov(\alpha + \beta R_M + \varepsilon, R_M) = \beta\sigma_M^2,$$

we have $\beta = \frac{\rho\sigma_R}{\sigma_M}$. We can define the beta of any investment portfolio by regressing its return against the returns from the market portfolio. Beta represents the amount of systemic risk of the investment. The higher the beta, the greater the systemic risk being taken and the greater the dependence of returns on the market performance.

Assumptions of CAPM model: The CAPM model implies that all investors want to hold the same portfolio of assets (the assets of the martket portfolio) ! That this is not true in practice is because of the following assumptions, which may or may not hold :

1. The investors only care about expected return and risk: since the return is not necessarily normally distributed, they also care about the skewness and skurtosis (excess skurtosis makes very high and very low return more likely).

2. The returns of different assets are only correlated via the market portfolio component (the ε are independent for different assets), i.e. the overall stock market : asset returns are correlated with each other in other ways : stock prices for companies in the same industry are likely to be correlated.

3. Investors have the same horizon of invesments : Different investors may care about different periods of investments, some very long for 30 years and some very short for 1 week.

4. Investors can borrow and lend at the same risk free rate : this clearly doesn't hold in practice.

5. Tax efects were ignored in CAPM model : optimizing taxation may be part of the goal of an investment decision.

6. Investors make the same estimates of expectations, standard deviations and correlations of returns : in reality, investors do not have homogeneous expectations.

Alpha: Alpha is the extra return of the portfolio over that predicted by CAPM :

$$\alpha = R_P - R_F - \beta(R_M - R_F).$$

Alpha may represent better portfolio management (or just sheer luck). The weighed average alpha of all investers **must be zero.** (That is for some investors who have positive alphas there must be others with negative alphas).

Arbitrage pricing theory: an extension of CAPM where the return depends on several factors (GDP, interest rate, inflation rate etc.) Each factor is a separate source of systemic risk. Unsystemic risk in arbitrage pricing theory is the risk unrelated to all of the factors.

Risk vs return for companies: According to CAPM theory, in considering an investment, a company should calculate its beta and its expected return. If the expected return is greater than that given by CAPM, the investment should be accepted. Otherwise it should be rejected. This suggestion implies nonsystemic risk should not be considered when making investment decision. In practice, companies are concerned with both systemic risk and nonsystemic risk (nonsystemic risk definitions are model dependent anyway) : they buy insurance for buildings, hedge their exposures to exchange rates, interest rates and other market variables. Many investors are also concerned about the overall risks : they prefer solid growth and a limit on the overall amount of risks, both systemic and nonsystemic. Taking a project with very high risk (but sufficiently high expected return) may not be desirable because this increases the bankruptcy probability. The bankruptcy costs are high in a non perfect world and lenders may charge a higher interst rates for such high risk projects. Thus relatively small investments can have the effect of reducing the overall risks because of diversification while a large invesstment can dramatically increase these risks.

2 Chapter 7: Valuation and scenario analysis: the risk neutral and real worlds

Risk neutral evaluation: A risk neutral world can be defined as an **imaginary world** where investors require no compensation for bearing risks. The world we live in is clearly not risk neutral. On the other hand, the valuation (pricing) of derivatives by assuming the risk neutral distribution is valid for all worlds, not just the risk neutral world. The key point to risk neutral valuation of a derivative is we are calculating the derivative price in terms of the **current price** of the underlying asset. In this sense, the future returns of the asset is not relevant to derivative price calculation (yet the risk of the asset, i.e the volatility, is relevant?). The future returns is relevant for investment purposes, but not for pricing. In fact, real world probability **can be**

used to price derivatives, by using an appropriate discounting rate (NOT the risk free rate). This rate is hard to obtain in practice, it depends on the particular option's beta (not the stock's beta) and likely to vary during the option's life : as the stock price changes, the leverage implicit in the option changes and so the discount rate changes. To have an idea of what discount rate should be applied when real world distribution is used, one can use a toy model of a binary option and figure out the discount rate needed to obtain the same price as the one from risk neutral evaluation. Finally, risk neutral evaluation can be viewed as an artificial (yet correct) device to value derivatives.

Risk neutral default probabilities: As an application of risk neutral evaluation, default probabilities of assets in a risk neutral world can be implied from prices of their derivatives, such as bond yield spread or credit swap spread. These probabilities are generally higher than real world default probabilities.

Scenario analysis: to examine what may happen in the future. The objective is NOT evaluation, or pricing. This should be carried out in the real world probabilities, as risk managers are not interested in future outcomes in a hypothetical world where everyone is risk neutral. Example : consider a forward contract with strike K and expiry in T years on an asset S. What is the 2 year 1 % VaR of this contract? Note that 1 % is a real world, not risk netral probability. First, for convenience we denote the generic 1 % probability as q. Let V be a (deterministic) value such that

$$P(S_T > V) = q.$$

Then $N(d-) = q$ where

$$d- = \frac{\log \frac{S_0}{V} + (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}.$$

Thus

$$V(q) = S_0 \exp\left[(\mu - \frac{1}{2}\sigma^2)T - N^{-1}(q)\sigma\sqrt{T}\right].$$

$V(q)$ is called the value of S_T that has a probability of q being exceeded. Thus the 1 % VaR of the forward contract is $V(q) - K$, where again q is the real world probability.

When both worlds have to be used: A possible scenario is to calculate the VaR of a portfolio of deratives at some future point. The real world is used to generate scenarios out to the time horizon considered. The risk neutral world is then used to value all outstanding transactions (prices of derivatives) at the future points conditioned on the **real world scenario**. For example, the q-VaR at a time point t of a forward contract with expiry T and strike K is

$$V(q,t) - Ke^{-r(T-t)},$$

where

$$V(q,t) = S_0 \exp\Big[(\mu - \frac{1}{2}\sigma^2)t - N^{-1}(q)\sigma\sqrt{t}\Big].$$

Estimating real world processes: The challenge in scenario analysis is that we usually have much more information about the behavior of market variables in the risk neutral world than in the real world. This is because these behaviors can be implied from the prices of options and other derivatives. There is no similar way of implying the behavior in the real world. From Girsanov theorem, the volatiltiy in the real world is the same as the one in the risk neutral world. Estimating the return is harder (in practice, the amount of historical data to get a reasonably accurate estimate is huge, much greater than required for volatility estimate). One alternative approach is from CAPM model. First the correlation ρ between the stock and a market representative, such as S&P 500 can be estimated. Then the stock $\beta = \rho\frac{\sigma}{\sigma_M}$. The return in the real world is $R_F + \beta E(R_M)$ where $E(R_M)$ is usually estimated to be 5 % or 6 %.

Another possible way is from the market price of risk, to obtain the real world return as $R_F + \lambda\sigma$ where $\lambda = \frac{\rho}{\sigma_M}E(R_M)$.

3 Chapter 8: How traders manage their risks

Delta: linear product is one whose value at any given time is linearly dependent on the value of an underlying market variable (such as forward contract but not options). Linear product has constant delta wrt the underlying market variable. It has the attractive property that hedges protect against large as well as small changes in the value of the underlying asset. The hedge, once set up, never needs to be changed (hedge and forget). The opposite is nonlinear product. Making a nonlinear portfolio delta neutral only protect against small changes in the underlying variable. Also the hedge needs to be changed frequently (dynamics hedging).

Delta hedging is sometimes easier and sometimes harder for exotic options. The Asian option is relatively easier to hedge. As time passes, we observe more of the asset prices that will be used to calculate the final average. Thus our uncertainty about the payoff decreases with the passage of time. The barrier options are relatively difficult to hedge, especially near the barrier. The delta of the option is discontinuous wt the barrier, making conventional hedging very difficult.

4 Interst rate risks

Duration (continuous compounding):

$$D = -\frac{1}{B}\frac{\Delta B}{\Delta y}.$$

Alternatively, let v_i be the **present value** of the cashflows c_i received at times $t_i, i = 1, \cdots, n$. We can define

$$D = \sum_{i=1}^{n} t_i \frac{v_i}{B}.$$

Modified duration (discrete compounding) : If y is expressed with compounding m times per year, the expression

$$D = \sum_{i=1}^{n} t_i \frac{v_i}{B}.$$

must be divided by $1 + \frac{y}{m}$.

Dollar duration: product of a bond's duration and its price :

$$D_{\$} = -\frac{dB}{dy}.$$

Interest rate deltas in practice : one approach is to define delta as the dollar duration. This is the sensitivity of the portfolio to a parallel shift in the zero-coupon yield curve. A particular measure is DV01, which s the impact of a one basis point increase in all rates. It is the dollar duration multiplied by 0.0001; or the duration multiplied by 0.0001 multiplied by the portfolio value.

Another approach is to use partial duration: just one point on the zero curve is shifted and all other points remain the same. In general, it is

$$D_i = -\frac{1}{P}\frac{\Delta P}{\Delta y_i},$$

where i represents the ith point on the yield curve and P is the portfolio value. The sum of all partial duration equals the usual duration.

PCA: The approach above can lead to 10 to 15 different deltas for every zero curve. On the other hand, the variables considered are highly correlated. That is the yield of different maturities tend to move in the same direction. One approach to handling the risk arising from groups of highly correlated market variables is PCA. It takes historical data on daily changes in the market variables and define a set of components that explain the movements. For risk free rates of n different maturities (e.g. $n = 8$) there correspond n vectors of size n of factor loadings (which comprises a factor matrix F). The jth factor loading for the ith rate is the (i, j) component of the factor matrix. The (daily) factor scores is the vector solution to $Ax = \Delta R$ where ΔR is the vector of daily change in rates. To implement PCA, one first calculae the covariance matrix from the observations. The factor loadings are the eigenvectors and the variance of the factor scores are the eigenvalues. The factor scores have the property that they are uncorrelated across the data.

Calculate Delta with PCA: The first few (two) factor loading vectors corresponding to the two biggest eigenvalues account for the majority of the variance in the market variables change. Thus we only need to calculate the change of the portfolio value corresponding to 1 unit change (which is just the vector itself) in the factor loading vectors (the vectors should be in the same unit as the market variables).

5 Chapter 10: Volatility

Volatility is the standard deviation of the continuously compounded (usually) daily return, where for the variable S on day i it is

$$\log \frac{S_i}{S_{i-1}}.$$

This is almost the same as

$$\frac{S_i - S_{i-1}}{S_{i-1}}.$$

Risk managers often focus on the variance rate perday instead of volatiity (since volatility is in some sense per "square root of day").

It is natural to assume that volatiltiy is caused by new information reacing the market. This has not been supported by research, since the variance of asset returns between Friday and Monday has NOT been found to be three times as high as that of a daily return (it is about 20 % higher). The reasonable conclusion is that volatiltiy is caused by trading activity itself.

If we assume daily returns are independent with the same standard deviation than $\sigma_{year} = \sigma_{day}\sqrt{252}$.

Power law: Daily returns on succesive days are not identically distributed (one reason is because volatiltiy is not constant). As a result, heavy tails are observed in the returns over relatively long periods as well as in the returns observed over one day. In this case, the power law provides an alternative to normal assumption. It asserts that for a random variable v it is approximately true that

$$P(v > x) = Kx^{-\alpha},$$

for some constants K, α. This equation has been found to be approximately true for v as diverse as income, city size and website visits per day. Taking the log on both sides give

$$\log P(v > x) = logK - \alpha \log x.$$

This provides both a way to test and to estimate α, K by linear regression.

Modeling daily volatility : Define u_i as the percentage change in the market vaiable between day $i-1$ and i :

$$u_i = \frac{S_i - S_{i-1}}{S_{i-1}}.$$

$\bar{u}$ is assumed to be zero. This is justified by the expected change in the daily returns is very small compared to the standard deviation. For example, suppose that Microsoft has an expected return of 20 % per annum and a daily volatility of 2 %. Over a one day period, the expected return is $\frac{0.2}{252} \approx 0.08\%$, which is very small compared to 2 %. Even over a typical period of 10 days, the expected return is 0.8 % whereas the standard deviation of return is $2 \times \sqrt{10} = 6.3\%$

The MLE estimate (with equal weights) for σ_n^2 is

$$\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}^2.$$

This is under the assumption that daily percentage changes are independent with mean 0 and same standard deviation σ.

If we retain the independent assumption but no longer assume the same standard deviation of daily returns, then the likelihood function is

$$\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi v_i}} e^{-\frac{u_i^2}{2v_i}},$$

where $v_i = \sigma_i^2$. Maximizing this is the same as maximizing

$$\sum_{i=1}^{m} -\log(v_i) - \frac{u_i^2}{v_i}.$$

We solve the optimization problem by positing relations between the v_i and u_i. The optimization problem reduces to optimizing the parameters in the recurrence relation. This results in ARCH(m), EWMA, GARCH(1,1) models.

ARCH(m) :

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u_{n-i}^2,$$

where

$$\gamma + \sum_{i=1}^{m} \alpha_i = 1,$$

and V_L is the long run variance rate. This is similar to the MLE estimate but with different weights.

EWMA :

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1-\lambda) u_{n-1}^2.$$

It can be showed that

$$\sigma_n^2 = (1-\lambda) \sum_{i=1}^{m} \lambda^{i-1} u_{n-i}^2 + \lambda^m \sigma_{n-m}^2.$$

For large m, the term $\lambda^m \sigma_{n-m}^2$ is sufficiently small so this is approximately ARCH(m) model with $\gamma = 0$ and $\alpha_i = (1-\lambda)\lambda^{i-1}$. EWMA approach requires the memory of only the current estimate of thevariance rate and the most recent observation on the value of the market variable. It is designed to track changes in the volatility. The value of λ governs how responsive the estimate is to the most recent daily percentage change update. A low λ gives a big weight to the most recent change while a high λ produces estimates that respond slowly to new information. RiskMetrics data base uses the EWMA model with $\lambda = 0.94$.

Garch(1,1):

$$\begin{aligned} \sigma_n^2 &= \gamma V_L + \alpha u_{n-1}^2 + \beta \sigma^2 n - 1, \\ \gamma + \alpha + \beta &= 1. \end{aligned}$$

EWMA is Garch(1,1) with $\gamma = 0, \alpha = 1-\lambda, \beta = \lambda$. Garch(p,q) uses the most recent p obersevations of u^2 and most recent q estimates of v_i. Garch(1,1) is by far the most popular.

It can be showed that

$$\begin{aligned} \sigma_n^2 &= \sum_{i=1}^{m} \beta^{i-1}\omega + \sum_{i=1}^{m} \alpha\beta^{i-1}u_{n-i}^2 + \alpha\beta^m u_{n-m}^2, \\ \omega &= \gamma V_L. \end{aligned}$$

Thus the weights decline exponentially at rate β. It is simlar to λ in EWMA. β defines the relative importance of the observations on the u_i in determining the variance rate. E.g., if $\beta = 0.9, u_{n-2}^2$ is 90% as important as u_{n-1}^2; u_{n-3}^2 is 81 % as important as u_{n-1}^2 etc.

Garch(1,1) is mean reversion: we have

$$\sigma_n^2 = (1-\alpha-\beta)V_L + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$$

so that

$$\sigma_n^2 - V_L = \alpha(u_{n-1}^2 - V_L) + \beta(\sigma_{n-1}^2 - V_L).$$

Replacing n by $n+t$ gives

$$\sigma_{n+t}^2 - V_L = \alpha(u_{n+t-1}^2 - V_L) + \beta(\sigma_{n+t-1}^2 - V_L).$$

Takig expectation on both sides, noting that $E(u_{n+t-1}^2) = \sigma_{n+t-1}^2$ gives

$$E(\sigma_{n+t}^2) - V_L = (\alpha+\beta)(\sigma_{n+t-1}^2 - V_L).$$

Thus

$$E(\sigma_{n+t}^2) = V_L + (\alpha+\beta)^t(\sigma_n^2 - V_L).$$

Since $\alpha + \beta < 1$ this is mean reversion as $t \to \infty$ $E(\sigma_{n+t}^2)$ approaches V_L.

Volatility term structure, volatility used for option pricing: Define $V(t) := E(\sigma_{n+t}^2)$ and $a = \log \frac{1}{\alpha+\beta}$. The above equation can be written as

$$V(t) = V_L + e^{-at}[V(0) - V_L].$$

The average variance rate perday between today and time T is

$$\frac{1}{T}\int_0^T V(t)dt = V_L + \frac{1-e^{-aT}}{aT}[V(0) - V_L].$$

Again this can be seen as mean reverting as the average variance rate approaches V_L as $T \to \infty$.

If we define $\sigma(T)$ as the volatility per annum that should be used to price a T-day option under the Garch(1,1) model then

$$\begin{aligned}\sigma(T)^2 &= 252\Big\{V_L + \frac{1-e^{-aT}}{aT}[V(0) - V_L]\Big\} \\ &= 252\Big\{V_L + \frac{1-e^{-aT}}{aT}[\frac{\sigma(0)^2}{252} - V_L]\Big\}.\end{aligned}$$

(This simply converts the variance rate per day to variance rate per annum as discussed above. The reason is in option pricing, other rates such as the risk free rate are also quoted per annum).

The relationship between the volatilties of options and their maturitires is referred to as the volatility term structure. It is usually calculated from implied volatility. On the other hand, the above equation provides a different approach to estimate the vol term structure from the Garch(1,1) model. Even though the Garch(1,1) vol term structure is not the same as the one from implied volatility, it is often used to predict the way the actual vol term structure will respond to volatility changes. Specifically, if $\sigma(0)$ changes by $\Delta\sigma(0), \sigma(T)$ changes by approximately

$$\frac{1-e^{-aT}}{aT}\frac{\sigma(0)}{\sigma(T)}\Delta\sigma(0).$$

Many financial institutions use analyses such as this when determining the exposure of their books to vol changes. Rather than consider an across the board increase of 1 % in implied vol when calculating vegas, they relate the size of the vol increase that is considered to the maturity of the option. For example, a 1 % increase in the instantaneous volatility (100 basis point change) with $V(0) = 0.0003$ and $\sigma(0) = sqrt252 \times 0.0003 = 27.50\%$ and $a = \log(1/0.99351)$ gives 0.97 % increase in vol for 10 day option, .92 % increase for 30 day option and .87 % increase for 50 day option.

EWMA and Garch(1,1) for covariance : We can also have a procedure to update covariance estimates similar to the EWMA and Garch(1,1) scheme for variance estimates. Specifically, for EWMA we have

$$cov_n = \lambda cov_{n-1} + (1-\lambda)x_{n-1}y_{n-1}.$$

For Garch(1,1) we have

$$cov_n = \omega + \alpha x_{n-1}y_{n-1} + \beta cov_{n-1}.$$

EWMA again can be viewed as the unequal weights version of the traditional equal weight covariance estimates when the means of X, Y are 0:

$$cov_n = \frac{1}{m}\sum_{i=1}^{m} x_{n-i}y_{n-i},$$

and Garch(1,1) is EWMA with a long term average. In covariance computation, we should check for the non-negative definiteness of the covaiance matrix obtained. Once we have the covariance and variance updating schemes, the correlation can be computed as:

$$\rho_n = \frac{cov_n}{\sqrt{var_{x,n}var_{y,n}}}.$$

6 Chapter 11: Correlations and copulas

Let $X_1, X_2, \cdots, X_n$ have marginal distributions $F_1, F_2, \cdots, F_n$ and joint distribution F. The copula of $(X_1, X_2, \cdots, X_n$ is defined by a joint distribution on the unit cube :

$$C(u_1, u_2, \cdots, u_n) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \cdots, F_n^{-1}(u_n)).$$

The copula captures the correlation structure of $X_1, X_2, \cdots, X_n$ (and the marginals $F_1, F_2, \cdots, F_n$ captures the rest). Indeed, Sklar's theorem states that

$$F(x_1, x_2, \cdots, x_n) = C(F_1(x_1), F_2(x_2), \cdots F_n(x_n)).$$

Since $F_1, F_2, \cdots, F_n$ captures only the marginals, it is intuitive that the copula C captures the rest of the distribution F, namely the correlation structure of F. Either way, Sklar's theorem allows a way to introduce a specified correlation structure to some given set of marginal distributions $F_1, F_2, \cdots, F_n$. The problem is, namely that given a set of marginal distributions $F_1, F_2, \cdots, F_n$ and a covaiance / correlation matrix A, can we construct a joint distribution F such that the marginals of F agree with the F_i and the correlation matrix of F is A ?

The answer is yes, via the copula as the "middle ground." First we construct a multidimensional distribution G that has the correlation structure as A. (In this sense, the distribution is ambiguous up to the individual variances of the marginals as these are not specified). The most convenient multidimensional distribution to use is the multivariate normal distribution. Next, the copula C_G associated with

G is computed. Finally, the desired multidimensonal distribution is defined via the equation

$$F(x_1, x_2, \cdots, x_n) = C_G(F_1(x_1), F_2(x_2), \cdots F_n(x_n)).$$

We next describe how to generate a multivariate Normal distribution with a specified correlation. One traditional way is via Cholesky decomposition. An alternative is via the so called factor model.

One factor model:

$$U_i = a_i F + \sqrt{1 - a_i^2} Z_i,$$

where F, Z_i have independent standard Normal distributions. The correlation between U_i, U_j is $\rho_{ij} = a_i a_j$. A one factor model imposes some structure on the correlation (the correlations within one row i or one colum j are multiples of a_i or a_j). With the one factor model we only need to estimates n parameters while without assuming the model the number of correlations that have to be estimated is $\frac{n(n-1)}{2}$. It also has the advantage that the resulting covariance matrix is always positive defininte. An example of the one factor model is the CAPM model.

Multifactor model :

$$U_i = \sum_{j=1}^{M} aijF_j + \sqrt{1 - \sum_{j=1}^{M} a^2ijZ_i},$$

again where F_j, Z_i are iid standard normal. In this case the correlation between U_i and U_j are

$$\sum_{i=1}^{M} a_{im} a_{jm}.$$

Vasicek's model for worst case default rate of loan portfolios:

Consider a portfolio of loans whose time to default T_i have the same marginals F and pairwise correlation ρ. That is the probability that the loan i has defaulted by time t is given by $T(t)$. The percentage of loan defaults by time t, PD(t) is clearly a random variable. Given a tolerance level X, we want to know the worst case default rate level $WCD(t, X)$ such that

$$P(PD(t) \leq WCD(t, X)) = X.$$

Typically X is chosen to be some high value such as 99 % or 99.9 %. To answer the question, we first need to model the T_i. We do this by using the one factor model :

$$U_i = \sqrt{\rho}F + \sqrt{1-\rho}Z_i,$$

where F, Z_i are iid standard normal. Next we model the T_i by $F^{-1}(N(U_i))$. That is

$$P(T_i \leq t) = P(T^{-1}(N(U_i)) \leq t) = P\big\{U_i \leq N^{-1}(T(t))\big\}.$$

Now

$$P\big\{U_i \leq N^{-1}(T(t))\big\} = P(Z_i \leq \frac{N^{-1}(T(t)) - \sqrt{\rho}F}{\sqrt{1-\rho}}).$$

Since Z_i are independent conditioned on F, we can think of the percentage of loan defaults by time t (conditioned on F) as

$$PD(t) \approx P(Z_i \leq \frac{N^{-1}(T(t)) - \sqrt{\rho}F}{\sqrt{1-\rho}}|F) = N(\frac{N^{-1}(T(t)) - \sqrt{\rho}F}{\sqrt{1-\rho}}).$$

Remark: Note that $PD(t)$ is clearly a random variable dependent on F while the default probability of each loan at time $t, P(T_i \leq t) = T(t)$ is a constant.

We have

$$P(PD(t) \leq WCD(t,X)) = P\big\{N(\frac{N^{-1}(T(t)) - \sqrt{\rho}F}{\sqrt{1-\rho}}) \leq WCD(t,X)\big\}.$$

If we define

$$WCD(t,X) = N(\frac{N^{-1}(T(t)) - \sqrt{\rho}G(X)}{\sqrt{1-\rho}}),$$

where $G(X)$ is some function of X to be specified then

$$P\big\{N(\frac{N^{-1}(T(t)) - \sqrt{\rho}F}{\sqrt{1-\rho}}) \leq WCD(t,X)\big\} = P(F \geq G(X)),$$

since $PD(t|F)$ is clearly a decreasing function in F. We want this quantity $P(F \geq G(X)) = X$. Thus $G(X) = -N^{-1}(X)$. Finally

$$WCD(t,X) = N(\frac{N^{-1}(T(t)) + \sqrt{\rho}N^{-1}(X)}{\sqrt{1-\rho}}).$$

Estimating $T^{-1}(t)$ and ρ : The equations

$$\begin{aligned} P(PD(t) \leq WCD(t,X)) &= X \\ WCD(t,X) &= N(\frac{N^{-1}(T(t)) + \sqrt{\rho}N^{-1}(X)}{\sqrt{1-\rho}}) \end{aligned}$$

leads to

$$WCD(t,X) \quad = \quad N(\frac{N^{-1}(T(t)) + \sqrt{\rho}N^{-1}\{P(PD(t) \le WCD(t,X))\}}{\sqrt{1-\rho}}),$$

for any X. Denote G as the cumulative distribution of $PD(t)$ we actually have

$$PD(t) = N(\frac{N^{-1}(T(t)) + \sqrt{\rho}N^{-1}\{G(PD(t))\}}{\sqrt{1-\rho}}),$$

where we think of $PD(t)$ as an actual realization of $PD(t)$. Rearranging this equation and denoting $PD(t)$ by x we have

$$G(x) = N(\frac{\sqrt{1-\rho}N^{-1}(x) - N^{-1}\{T(t))\}}{\sqrt{\rho}}).$$

This gives us a way to calculate $T(t)$ and ρ as the MLE that maximizes the likelihood function associated with the density $g(x) := G'(x)$.

Concretely, we use historical data (say the annual percentage default $x_i, i = 1, \cdots, n$ from the last n years, that is $t = 1$ year here) to find the MLE associated with the likelihood function $\prod_{i=1}^{n} g(x_i)$. An example of such data is Table 11.4 of annaul percentage default rate for all rated companies, 1970-2013.

Alternatives to Gaussian copula: The factor model has limitations. That is it has very little tail dependence: an an usually early default does not often happen together with another unusually early deault. Given a $T(t)$ (that may be estimated independently from else where) it may be difficult to find a ρ that fits the data. For example, there is no ρ that is consistent with $T(1) = 1\%$ and the situation where one year in 10 the default rate is greater than 3 %. Other one factor model can provide a better fit to the data.

Recall that in the one factor model, we had

$$U_i = \sqrt{\rho}F + \sqrt{1-\rho}Z_i,$$

where F, Z_i are independent standard normals. An alternative is to choose other distribution for F, Z_i that have heavier tails than the normal distribution (scaled so that they have mean zero and standard deviation 1). The $WCDR(T,X)$ becomes

$$WCD(t,X) = \Phi(\frac{\Psi^{-1}(T(t)) + \sqrt{\rho}\Theta^{-1}(X)}{\sqrt{1-\rho}}),$$

where Φ, Θ, Ψ are the distribution functions of Z_i, F, U_i respectively. Note that Ψ may have to be computed numerically.The ditribution of $PD(t)$ becomes

$$G(x) = \Theta(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Psi^{-1}\{T(t))\}}{\sqrt{\rho}}).$$

7 Chapter 12: VaR and Expected Shortfall

VaR: A function of 2 parameters, time horizon T and confidence level X. It is the loss level during a time period of length T that we are X percent confident will not be exceeded.

Expected short fall (conditional VaR, conditional tail expectation, expected tail loss) : also a function of 2 parameters, time horizon T and confidence level X. It is the expected loss during time T conditioned on the loss being greater than the X-th percentile of the loss distribution.

Formulas under normal distribution assumption: When the loss L is assumed to be normally distributed with mean μ and standard deviatio σ,

$$\begin{aligned} VaR &= \mu + \sigma N^{-1}(X) \\ ES &= \mu + \sigma \frac{e^{-\frac{[N^{-1}(X)]^2}{2}}}{\sqrt{2\pi}(1-X)}. \end{aligned}$$

This is because we require

$$P(L \leq VaR) = 1 - X,$$

i.e.

$$P(Z \leq \frac{VaR - \mu}{\sigma}) = 1 - X.$$

While

$$\begin{aligned} ES &= \frac{\int_{VaR}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}{1-X} \\ &= \frac{\frac{1}{\sqrt{2\pi}} \int_{N^{-1}(X)}^{\infty} (\sigma x + \mu) e^{-\frac{x^2}{2}} dx}{1-X} \\ &= \frac{1}{\sqrt{2\pi}(1-X)} (\sigma e^{-\frac{[N^{-1}(X)]^2}{2}} + \mu(1-X)) \\ &= \mu + \sigma \frac{e^{-\frac{[N^{-1}(X)]^2}{2}}}{\sqrt{2\pi}(1-X)}. \end{aligned}$$

Time horizon: When market risks are being considered, analysts often start by calculating VaR or ES for one day. The usual assumption is

$$\text{T-day VaR} = \text{1-day VaR} \times \sqrt{T}$$
$$\text{T-day ES} = \text{1-day ES} \times \sqrt{T}.$$

These formulas are approximation and they are exactly true when the changes in portfolio value on successive days have independent and identical normal distribution with mean 0. The derivation can be seen from the above calculation of VaR and ES under the normal assumption. If there is autocorrelation, there needs to be modification as followed. Suppose there is first order correlation between consecutive changes of daily portfolio value, that is $corr(\Delta P_i, \Delta P_{i-1}) = \delta$. The model for this is

$$\Delta P_i = \rho \Delta P_{i-1} + \sqrt{1-\rho^2}\varepsilon_i,$$

where ε_i is independent of P_{i-1}. Note that this implies $corr(\Delta P_i, \Delta P_{i-m}) = \delta^m$. Under this model, the standard deviation for the total lost in T days, $\sum_{i=1}^{T} \Delta P_i$ is

$$\sigma\sqrt{T + \sum_{i=1}^{T-1} 2(T-i)\rho^i}.$$

Converting between different confidence levels under normal assumptions:

$$\begin{aligned} VaR(X^*) &= VaR(X)\frac{N^{-1}(X^*)}{N^{-1}(X)} \\ ES(X^*) &= ES(X)\frac{(1-X)e^{-\frac{(Y^*-Y)(Y^*+Y)}{2}}}{1-X^*} \end{aligned}$$

where $Y^* = N^{-1}(X^*), Y = N^{-1}(X)$. If normal assumptions are not satisfied then extreme value theory is appropriate. In general, to estimate VaR directly when the confidence level is very high may be very difficult.

Spectral risk measure: Risk measures can be viewed as giving weights to percentiles of the loss distribution. VaR gives 100 % weight to the Xth percentile. ES gives equal weight to all percentiles greater than the Xth percentile and zero weight to other percentiles below X. A general spectral measure assigns weights to the qth percentile and it is coherent if the weights is a non decreasing function of q. (This is the reason why VaR is not coherent and ES is). An example is the exponential spectral risk measure, where the weight assigned to the qth percentile is $e^{-\frac{1-q}{\gamma}}$, where γ is a constant.

VaR vs ES:

1. As coherent risk measures: a coherent risk measure satisfies four porperties: monotonicity, translation invariance, homogeneity, subadditivity. VaR is not a coherent risk measure : it is not subadditive, see example below while ES is.

2. For risk management: VaR may encourage some undesirable risk taking behavior (it can lead to a loss distribution with a large spike in the tail) while ES does not do so (since it averages out the tail loss).

3. Back-testing: VaR is easier to back-test with historical simulation, while ES is harder. Thus ES is usualy used to determine regulatory capital while VaR estimates are used for back-testing.

Marginal, incremental and component measures: Consider a portfolio that is composed of a number of subportfolios. Suppose that the amount invested in the ith subportfolio is x_i. The marginal VaR for the ith portfolio is $\frac{\partial VaR}{\partial x_i}$. The component VaR for the ith subportfolio is $\frac{\partial VaR}{\partial x_i}x_i$. The incremental VaR for the ith subportfolio is the difference in VaR with or without the ith subportfolio. Component VaR is a reasonable approximation to incremental VaR. Under the assumption that the ith marginal VaR does not change as x_i decreases to 0, we see that the incremental VaR is $\frac{\partial VaR}{\partial x_i}(x_i - 0)$ equalling the component VaR.

Allocating risk measures to subportfolios: Euler's theorem states that if $V(x_1, x_2, \cdots, x_n)$ is linearly homogeneous:

$$V(\lambda x_1, \lambda x_2, \cdots, \lambda x_n) = \lambda V(x_1, \cdots, x_n),$$

then

$$V = \sum_{i=1}^{M} \frac{\partial V}{\partial x_i} x_i.$$

This property is satisfies by most risk measures, including VaR and ES. Thus this equation provides a natural allocation of risk measures to the subportfolio: the VaR (ES) of the ith portfolio is the ith component VaR.

Aggregating VaRs and ESs: The opposite of allocating risk measures to subportfolios is aggregating the risk measures of subportfolios to a total VaR. The formula is

$$\mu_{total} = \sqrt{\sum_{ij} \rho_{ij}\mu_i\mu_j},$$

where ρ_{ij} is the correlation between the losses from components i and j. μ represents either VaR or ES.

Back-testing: Suppose that we have developed a procedure for calculating a one day X % VaR. Back-testing involves looking at home often the loss in a day would have exceeded the one-day X % VaR. Days when this occur are referred to as exceptions. Suppose that the number of exceptions is m and the total number of days is m. We perform a hypothesis test on the (unknown) true level of exception p versus $p_0 = 1-X$. The null hypothesis is

$$H_0 : p = 1 - X$$

and the alternative can either be

$$H_1 : p > 1 - X$$

(the VaR level is too low, the true level of exception is higher than 1 - X) or

$$H_1 : \frac{m}{n} < 1 - X$$

(the VaR level is too high, the true level of exception is less than 1 - X). The test statistics is

$$\hat{\theta} = \sum_{i=m}^{n} \binom{n}{i} p_0^i (1 - p_0)^{n-i}$$

for the first case (VaR is too low) and

$$\hat{\theta} = \sum_{i=0}^{m} \binom{n}{i} p_0^i (1 - p_0)^{n-i}$$

for the second case (VaR is too high). The null hypothesis H_0 is rejected if $\theta < \alpha$, where α is some chosen confidence level.

Hypothetical versus actual change of the portfolio: We can compare VaR with the hypothetical change in the portfolio value calculated on the assumption that the composition of the portfolio remains unchanged during the day. The other alternative is to compare VaR with the actual change in the value of the portfolio during the day. The assumption of VaR is that the portfolio remains unchanged during the day so the hypothetical comparison is more theoretically correct. In practice, both kinds of change are relevant in risk management. The actual changes are adjusted for items unrelated to market risk, such as fee income and profits carried out at prices different from mid-market.

There are also statistical tests to test for the two sided hypothesis

$$H_1 : p \neq 1 - X$$

and the bunching effect (the effect that losses on successive days are bunched together). See Hull at the end of chapter 12 for more details.

8 Chapter 13: Historical simulation for risk measures and extreme value theory

The methodology: Historical simulation involves using past data as a guide to what would happen in the future. For example, suppose we want to compute one day 99% VaR of a portfolio using 501 days of data (these are the standard time horizon and confidence level used for market risk VaR). 501 days of data leads to 500 scnarios, where scenario i includes the change of the market variables from day i-1 to day i, as well as the change in portfolio value in response to these changes. These data points define the empirical distribution of the portfolio daily loss. The (historically estimated) VaR is the 99 percentile of this empirical distribution (the fifth worst outcome). On the other hand, we also obtain an empirical distribution of the market variale tomorrow:

$$v_{n+1}(\omega_i) = v_n \frac{v_i}{v_{i-1}},$$

where ω_i denotes the ith scenario under consideration. Note that the portfolio value is a function of the market variables:

$$\pi_{n+1} = G(v^1_{n+1}, v^2_{n+1}, \cdots, v^M_{n+1}).$$

Thus the portfolio value under scenario i is exactly the function G valuated at $(v^1_{n+1}(\omega_i), v^2_{n+1}(\omega_i), \cdots, v^M_{n+1}(\omega_i))$. Finally, the (historically estimated) expected short fall ES is the average of the 5 worst loss recorded from the most recent 501 days of data.

Each day the daily VaR would be updated using this procedure. In practice, the portfolio is likely to change from day to day. The daily VaR is computed on the assumption that the portfolio remains unchanged over the next business day. The market variables under consideration for VaR calculation include exchange rates, commodity prices and interest rates. In case of interest rates, term structure of zero-coupon interest rates in a number of different currencies may be needed to value the portfolio. There might be as many as 10 market variables for each zero curve (Treasury, LIBOR / swap etc) to which the financial institution is exposed.

Stressed VaR and Stressed ES: To calculate these measures, the institution needs to search for 251 days where its VaR and ES would be greatest. These would serve as the empirical distribution of the stressed loss. The one day 99 % stressed VaR is the 99th percentile of this distribution, which is the mid point between the second and third worst loss. The one day stressed ES would be $0.4L_1 + 0.4L_2 + 0.2L_3$ where L_i is the ith worst loss.

VaR confidence interval : since VaR is an estimate of the q-percentile of the loss distribution, we can consstruct a confidence level for this q-percentile, centered around the VaR. First, the standard error for estimating the q-percentile x_q of a distribution with density function $f(x)$ using n observations is

$$\hat{\sigma} = \frac{1}{f(\hat{x}_q)}\sqrt{\frac{(1-q)q}{n}},$$

where $\hat{x}_q$ is the estimate of x_q ($\hat{x}_q$ is the q-VaR in our case). In other words, $\hat{\sigma}$ is the estimate of the standard deviation of $\hat{x}_q$. $f(x)$ can be estimated by fitting the empirical data to an appropriate distribution whose properties are known. For example, under the assumption that the daily losses are identically normally distributed and indenpendent, $f(x)$ is the normal density with mean equalling the empirical mean (which should be approximately 0) and standard deviation equalling the empirical standard deviation. $\hat{x}_q$ is approximately Normal with mean x_q and the standard deviation equalling $\hat{\sigma}$. Thus a $1-\alpha$ CI for x_q is

$$\hat{x}_q \pm z_\alpha \hat{\sigma}.$$

The identical normality (stationarity) and independence assumption are not realistic in practice. Instead of normal assumption, some heavier tailed distributions are more appropriate (student t, Pareto). Instead of stationarity, one can update the estimates to take into account the volatilities of the market variables. The details are given in the next section.

Alternatively, the bootstrap method can also be used to estimate a confidence interval for VaR. The idea is to sample with replacement from the current data to create new similar data sets. Each of these new data sets gives a new VaR estimate. A 95 % confidence interval for VaR is the range between the 2.5 percentile point and the 97.5 percentile point of the VaRs calculated from these data sets. Note that if we have a data set of 500 data points then we need to sample 500,000 times to create 1000 new data sets with 500 data points each.

Extensions of historical simulation:

Weighting of observations: The historical simulation approach described above gives equal weights to all scenarios. That is, the empirical loss distribution gives equal weights to all scenarios. Similar to the volatility update approach, it makes sense to give more weights to the more recent scenarios and less weights to the more distant ones. That is we assign a weighting scheme similar to the EWMA weighting scheme to the scenarios:

$$w_i = \frac{\lambda^{n-i}(1-\lambda)}{1-\lambda^n}.$$

The weights are so that $\sum_i w_i = 1$. The empirical loss distribution is constructed such that the loss on the ith day is given weight w_i. The 1 - α percentile of the (unequally weighted) empirical loss distribution can be computed by adding the worst losses until their corresponding weights exceed α. The parameter λ can be estimated by tring different values to see which one back-tests best.

Adjusting for volatilities changes: Recall that the portfolio is a function of the market variables:

$$\pi_{n+1} = G(v^1_{n+1}, v^2_{n+1}, \cdots, v^M_{n+1}).$$

The market variable under scenario i is updated according to the formula

$$v_{n+1}(\omega_i) = v_n \frac{v_i}{v_{i-1}}.$$

This formula does not take into account the relative difference between the current estimate of volatility σ_{n+1} (the volatility between today and tomorrow) and σ_i (the volatility between day i - 1 and day i). If the volatility σ_{n+1} is twice as big as the volatlity σ_i it makes sense to expect the change in the market variable from day n to day n+1 is twice as big as the change in the market variable from day i-1 to day i (and similarly when it is half as big). Thus to reflect for the volatility effect, the market variable update under scenario i is

$$v_{n+1}(\omega_i) = v_n \frac{v_i + (v_i - v_{i-1})\frac{\sigma_{n+1}}{\sigma_n}}{v_{i-1}}.$$

The effect of volatility adjustments is to create more variability in the gains and losses in the 500 scenarios. The market variables can be exchange rates and stock indices. Hull and White has showed that this approach is superior to the traditional historical simulation and to the EXMA scenario weighting scheme.

A more direct approach is to adjust for the loss of the portfolio on day n+1 using the estimated standard deviations of the daily losses in the scenarios. Specifically, using the EWMA scheme we can keep track of the standard deviations of the 500 portfolio daily losses. An adjusted loss for the ith scenario is then calculated by multiplying the loss given by the standard approach by the ratio of the estimated standard deviation for the last (500th) scenario to the estimated standard deviation for the ith scenario:

$$L_{adjusted,n+1}(\omega_i) = L_n(\omega_i)\frac{\sigma_{n+1}}{\sigma_i}.$$

Computational shortcut: To reduce computation time when running the historical simulation, financial institutions sometimes use a delta-gamma approximation. Suppose an instrument whose price P depends on a single market variable S. Then ΔP is approximately

$$\Delta P \approx \delta \Delta S + \frac{1}{2}\gamma(\Delta S)^2,$$

where δ, γ are the delta and gamma of P with respect to S. These values are always known because they are calculated each day when the instrument is marked to market. When an instrument depends on several market variables, the approximate change is

$$\Delta P \approx \sum_i \delta_i \Delta S_i + \sum_{i,j} \frac{1}{2}\gamma_{ij}(\Delta S_i)(\Delta S_j),$$

where $\delta_i = \frac{\partial P}{\partial S_i}$ and $\gamma_{ij} = \frac{\partial^2 P}{\partial S_i \partial S_j}$.

Extreme value theory: Recall the power law : Daily returns on succesive days are not identically distributed (one reason is because volatiltiy is not constant). As a result, heavy tails are observed in the returns over relatively long periods as well as in the returns observed over one day. In this case, the power law provides an alternative to normal assumption. It asserts that for a random variable v it is approximately true that

$$P(v > x) = Kx^{-\alpha},$$

for some constants K, α. A justification for the power law is the extreme value theory (EVT). It is a theory for estimating the tails of a distribution. EVT can be used to improve VaR estimates, especially when the estimate is at a very high confidence level. It is a way of smoothing and extrapolating the tails of an empirical distribution.

Let F be the distribution function of a random variable X. Denote

$$F_u(y) = P(u < X < u + y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}.$$

EVT states that for a large class of distributions, as u gets large $F_u(y)$ converges the generalized Pareto distribution:

$$F_u(y) \to G_{\xi,\beta}(y) = 1 - \left[1 + \xi \frac{y}{\beta}\right]^{-\frac{1}{\xi}}.$$

The generalized Pareto distribution is controlled by two parameters: the shape parameter ξ that controls the heaviness of the tail and the scale parameter β. When X has Normal distribution, $\xi = 0$ and

$$G_{\xi,\beta}(y) = 1 - e^{-\frac{y}{\beta}}.$$

The parameters ξ, β can be estimated by the MLE method. First, we choose a level u (such as the 95th percentile of the empirical distribution). We then pick the observations x_i that are above u. The likelihood function to maxiize is:

$$\prod_{i=1}^{n_u} \frac{1}{\beta} \left(1 + \frac{\xi(x_i - u)}{\beta}\right)^{-\frac{1}{\xi}-1}.$$

(This is the product of the density function of $G_{\xi,\beta}$ evaluated at x_i.)

The general tail distribution $1 - F(x)$ is estimated as

$$P(X > x) = P(X > u + (x-u)|X > u)P(X > u) = [1 - G_{\xi,\beta}(x-u)][1 - F(u)].$$

The estimate for $1 - F(u)$ is $\frac{n_u}{n}$. Thus the final estimate for $1 - F(x) = P(X > x)$ is

$$P(X > x) = [1 - G_{\xi,\beta}(x-u)]\frac{n_u}{n} = \frac{n_u}{n}\left[1 + \xi\frac{x-u}{\beta}\right]^{-\frac{1}{\xi}}.$$

When $u = \frac{\beta}{\xi}$ the above equation becomes

$$P(X > x) = \frac{n_u}{n}\left[\frac{\xi x}{\beta}\right]^{-\frac{1}{\xi}}.$$

This is the power law $Kx^{-\alpha}$ with $K = \frac{n_u}{n}\left[\frac{\xi}{\beta}\right]^{-\frac{1}{\xi}}$.

Finally, the q-VaR is the qth percentile of the distribution that satisfies

$$q = \frac{n_u}{n}\left[\frac{\xi x}{\beta}\right]^{-\frac{1}{\xi}}.$$

Thus

$$VaR = u + \frac{\beta}{\xi}\left\{\left[\frac{n}{n_u}(1-q)\right]^{-\xi} - 1\right\}.$$

The expected shortfall is

$$ES = \frac{VaR + \beta - \xi u}{1 - \xi}.$$

It should be noted that the estimates of ξ, β clearly depends on u. However, the estimates of $F(x)$ remains roughly the same for various values of u that are sufficiently high. In general, we want u to be sufficiently high so that we are truly investigating the tail of the distribution, but also sufficiently low so that the number of data included in the MLE estimate is not too low.

9 Chapter 14: Model building approach for risk measures

We consider how to build various models to assess VaR and ES.

One asset case: In the case of one asset portfolio, we have

$$\Delta P = \alpha \frac{\Delta S}{S},$$

where α is the dollar amount invested in S and $\frac{\Delta S}{S}$ is the daily return (percentage change) of asset S. To find the VaR and ES of P, we use the formula under normal distribution assumption. assume the daily loss $L = -\Delta P$to be normally distributed with mean 0 and standard deviatio σ, the X level VaR is

$$\begin{aligned} VaR &= \sigma N^{-1}(X) \\ ES &= \sigma \frac{e^{-\frac{[N^{-1}(X)]^2}{2}}}{\sqrt{2\pi}(1-X)}. \end{aligned}$$

Remark: If the volatility of S is σ_S then $\sigma = \alpha \sigma_S$ in the above formula. The fact that ΔP has zero mean comes from the observation that the mean of daily return is insignificant compared with the daily volatility.

Undet the iid assumption of daily losses, the N day VaR is $\sqrt{N} VaR_{daily}$ and the N day ES is $\sqrt{N} ES_{daily}$.

Two asset case: In the two asset case under Normality assumption we use a similar set up as the one asset case, with

$$\Delta P = \alpha_1 \frac{\Delta S_1}{S_1} + \alpha_2 \frac{\Delta S_2}{S_2}.$$

The formula for σ of ΔP is

$$\sigma^2 = \sum_{i,j=1}^{2} \rho_{ij}\alpha_i\alpha_j\sigma_i\sigma_j$$

where ρ_{ij} is the correlation of the returns of the assets i, j.

Note: It may happen that $VaR_{\alpha_1 S_1+\alpha_2 S_2} \leq VaR_{\alpha_1 S_1} + VaR_{\alpha_2 S_2}$ (as VaR is not necessarily subadditive, while ES is). If this is the case then the difference of the two represents the benefit of diversification.

N asset case: In the N asset case under Normality assumption we have

$$\begin{aligned}\Delta P &= \sum_i \alpha_i \frac{\Delta S_i}{S_i} \\ \sigma^2 &= \sum_{i,j} \alpha_i\alpha_j\rho_{ij}\sigma_i\sigma_j\end{aligned}$$

and we apply the general formula for VaR and ES under normality assumption.

Calculation of asset covariances from historical data:

The N asset model approach still requires the input of σ_i and ρ_{ij}. These can be estimated using the methodology covered in the volatility chapter. Specifically, recall that we have for EWMA scheme:

$$cov_{ij,n} = \lambda cov_{ij,n-1} + (1-\lambda)u_{i,n-1}u_{j,n-1}.$$

For Garch(1,1) we have

$$cov_{ij,n} = \omega + \alpha u_{i,n-1}u_{j,n-1} + \beta cov_{ij,n-1},$$

where u_i denotes the return of asset i. The correlation can be computed as:

$$\rho_{ij,n} = \frac{cov_{i,n}}{\sqrt{var_{i,n}var_{j,n}}}.$$

Portfolio exposure to interest rates: The above discussion applies to the portfolio's exposure to the fluctuation of daily returns of assets. The model connects the portfolio change with the asset daily returns:

$$\Delta P = \sum_i \alpha_i \frac{\Delta S_i}{S_i}.$$

Thus, the model is suited for portfolio of stocks and indices. When the portfolio consists of bonds, it is more appropriate to consider the portfolio's exposure to change in the interest rates rather than the bonds' daily returns. That is

$$\Delta P = -DP\Delta y,$$

where D is the modified duration of the portfolio and Δy is the parallel shift in one day. Having presented the equation, we immediately abandon it since a shift in the yield curve doesn't usually happen in a parallel manner. More specifically, we need to consider the portfolio's exposure to change in interest rates of different maturities. The procedure, known as cashflow mapping, is to choose as market variables the price of zero-coupon bonds with standard maturities: 1 month, 3 months, 6 months, 1 year, 2 years, 5 years, 7 years, 10 years and 30 years. The cashflows from instruments in the portfolio are then mapped into cash flows occuring on the standard maturity dates for the purpose of calculating VaR.

For the purpose of demonstration, suppose we want to evaluate the exposure of a cashflow of amount L_T to be received in T years, where $T_0 < T < T_1$ are two standard maturities. The first step is to interpolate the yields y_0, y_1 to get the yield y_T :

$$y_T = \frac{T - T_0}{T_1 - T_0} y_0 + \frac{T_1 - T}{T_1 - T_0} y_1.$$

We then discount with rate y_T to obtain the present value L_0 of L_T

$$L_0 = e^{-y_T T} L_T.$$

The second step is to interpolate the volatilities σ_0, σ_1 to get the volatility σ_T :

$$\sigma_T = \frac{T - T_0}{T_1 - T_0} \sigma_0 + \frac{T_1 - T}{T_1 - T_0} \sigma_1.$$

We then calculate the weights α and $1-\alpha$ allocations to the two bonds with maturities T_0, T_1 respectively by matching the second moment:

$$\sigma_T^2 = \alpha^2 \sigma_0^2 + (1 - \alpha)^2 \sigma_1^2 + 2\alpha(1 - \alpha)\rho_{01}\sigma_0\sigma_1.$$

The amount of money allocated to the bond with maturity T_0 is then αL_0 and to the bond with maturity T_1 is then $(1 - alpha)L_0$. It can be shown that $\alpha, 1 - \alpha$ are always positive in the cashflow mapping procedure.

Having performed the cashflow mapping, we can then revert to the N asset approach discussed in the previous section to calculate the VaR of the portfolio. Namely we calculate

$$\begin{aligned} \Delta P &= \sum_i \alpha_i \frac{\Delta B_i}{B_i} \\ \sigma^2 &= \sum_{i,j} \alpha_i \alpha_j \rho_{ij} \sigma_i \sigma_j, \end{aligned}$$

where B_i are the bonds of standard maturities that we have mapped into with dollar amount α_i respectively.

Principal component analysis: An alternative approach to cashflow mapping is to measure the portfolio's exposure to the factors obtained the interest rates principle component decomposition. Recall that the factor scores f_i are such that

$$\Delta y = A\mathbf{f},$$

where Δy is the vector of daily rates change, A is the factor matrix and f is the vector of factor scores. Also recall the equation of the portfolio exposure to the rate change

$$\Delta_i P = -D_i P \Delta y_i,$$

where y_i is a rate of standdard maturity and $\Delta_i P$ is the change of the portfolio wrt to rate y_i. The total change of the portfolio is

$$\Delta P = \sum_i \Delta_i P = -P \sum_i D_i \Delta y_i = -P\mathbf{1}^T [DA]\mathbf{f},$$

where by DA we mean the matrix obtained by multiplying each ith column of A with D_i and $\mathbf{1}$ is the vector of all ones. The point of this equation is there are exposure factors α_i so that

$$\Delta P \approx \sum_i = 1^m \alpha_i f_i,$$

where m is a small number like 2 or 3. Once we are at this point, we can again reuse the above formulas for calculating VaR and ES, with

$$\sigma^2 = \sum_i \alpha_i^2 \sigma_i^2,$$

where σ_i is the SD of the factor scores and we remember the fact that the factor scores are uncorrelated.

Application of linear models: Linear model can be used to find VaR and ES of portfolios with no derivatives consisting of positions in stokcs, bonds, foreign exchanges and commoditities. This is because the change in the value of the portfolio is linearly dependent on the percentage changes in the prices of the assets comprising the portfolio, using the techniques mentioned above. Examples of derivatives that can be handled by linear models are forward contracts and swaps, because they are just straightforward exchange of assets. Thus the portfolio value change in these derivatives again follow a linear relationship with the percentage change of the underlying exchanged assets.

Weakness of linear model: When the portfolio includes options, the linear model is an approximation. One calculate the delta of the portfolio with respect to the ith market variable:

$$\delta_i = \frac{\Delta P}{\Delta S_i},$$

and it follows approximately that

$$\Delta P \approx \sum_i \delta_i \frac{\Delta S_i}{S_i}.$$

The point here is when the portfolio includes options, δ_i changes with time (whereas before the δ_i are exactly α_i which is constant as long as the portfolio composition does not change). Viewed from a different angle, with the presence of portfolio, the gamma of the portfolio is not zero. Assuming the normal distribution of underlying asset price (which is a good approximation for log normal distribution for short time periods) if the option gamma is positive, the distribution of option price is positively skewed (heavier right tail). If the gamma is negative, the distribution of option price is negatively skewed (heavier left tail) . If we assume a symmetric (normal) distribution while the actual distribution has heavier left tail, the VaR will be too low. On the other hand, if we assume a symmetric (normal) distribution while the actual distribution has heavier right tail, the VaR will be too high.

Quadratic model : To incorporate the gamma of the portfolio, the quadratic model is used. For one asset portfolio, the formula for the change in portfolio value is

$$\begin{aligned} \Delta P &= \delta S \frac{\Delta S}{S} + \frac{1}{2}\gamma S^2 \left(\frac{\Delta S}{S}\right)^2 \\ &= \delta S \Delta x + \frac{1}{2}\gamma S^2 (\Delta x)^2, \end{aligned}$$

where we denoted $\Delta x := \frac{\Delta S}{S}$. For a multiasset portfolio, the formula is

$$\Delta P = \sum_i \delta_i S_i \Delta x_i + \frac{1}{2} \sum_{ij} \gamma_{ij} S_i S_j \Delta x_i \Delta x_j,$$

where $\gamma_{ij} = \frac{\partial^2 P}{\partial S_i \partial S_j}$ and $\Delta x_i = \frac{\Delta S_i}{S_i}$.

VaR for Quadratic model: Even under the normality assumptions of Δx_i, it is not straightforward to calculate the percentile of ΔP under the quadratic model. To approximate the percentile, we use the Cornish-Fisher Expansion, which states that the qth percentile of a distribution with mean μ, standard deviation σ and skewness ξ is approximaely

$$x_q \approx \mu + w_q \sigma,$$

where

$$\begin{aligned} w_q &= z_q + \frac{1}{6}(z_q^2 - 1)\xi \\ z_q &= N^{-1}(q). \end{aligned}$$

The skkewness of a distribution X is defined as

$$\xi := E\left(\frac{X-\mu}{\sigma}\right)^3.$$

The quadratic model allows for the calculation of the first 3 moments of ΔP and the Cornish-Fisher Expansion gives the approximation of the q-VaR based on these moments.

Monte Carlo simulation and modelling: We can directly compute the VaR using Monte Carlo simulation only assuming the distribution of Δx_i. Once the Δx_i are generated, the initial and final values of the portfolio can be calculated from the initial an final values of the market variables S_i and thus ΔP can be deduced. Repeating this process for large number of times can give an approximation for the qth percentile of $-\Delta P$. On the other hand, if the portfolio is complex, direct calculation of ΔP may be undesirable. In this case, the quadratic model can be used as an approximating shortcut to ΔP without having to evaluate the complete initial and final values of the portfolio. In Monte Carlo simulation, any marginals distribution of Δx_i can be assumed, preferably ones with heavier tails than the normal distribution (such as student t). Furthermore, any correlation structure among x_i can also be imposed using the copula approach : We sample $(u_1, u_2, \cdots, u_n)$ from a multivariate Normal distribution with the given correlation structure. The desired value of Δx_i is given by

$$\Delta x_i := F^{-1}(N(u_i))$$

where F is the given marginal distribution of Δx_i.

Model building vs historical simulation : The advantages of model building is that results can be produced very quickly and can easily be used in conjunction with volatility and correlation updating procedures such as EWMA and Garch(1,1). The main disadvantage of the model building approach is the normality assumption on the percentage change of the market variables. This does not hold true in practice. The model building approach is most often used for investment portfolios (it is closely related to the mean-variancem method for portfolio analysis). It is less commonly used for calculating the VaR for the trading operations of a financial institution. This is because financial institutions usually maintain their portfolio deltas close to zer. Neither the linear nor the quadratic model work well when deltas are low and the portfolios are nonlinear. When delta is zero the first order approximation gives $\Delta P = 0$ which is of course incorrect (check out also the DerivaGem calculation). The point is when the first derivative is zero we need higher order approximation beyond the second order to get a decent precision for ΔP. On the other hand, going beyond the second order is definitely going beyond the quadratic model and calculating the third partials may prove tedious. When $\Delta = 0$ the direct Monte Carlo simulation approach may be more desirable.

The historical simulation approach has the advantage that the historical data determines the joint distribution of the market variables. It's also easier to handle interest rates in historical simulation because on each trial, a complete zero curve or both today and tomorrow can be calculated (from historical data). The main disadvantage is that it is computationally much slower than the model buiding approach.

10 Chapter 22: Scenario analysis and Stress testing

The approach of VaR / ES calculation is a historical simulation approach, which is backward looking (assuming the future is similar to the past). Events in the future could happen that is quite different from the period covered by the data. Stress testing is an attempt to overcome this weakness in VaR / ES measure. Stress testing involves estimating how the portfolio would perform under scenarios involving extreme (but plausible) market moves.

Approaches of stress testing:

1. Stressing individual variables: Use scenarios where there is a large move in one

variable and others are unchanged. For example: a) A 100 basis point parellel shift (up or down) in a yield curve b) Increasing or decreasing all implied vol used for an asset by 50 % of current values c) Increasing or decreasing equity index by 10 % d) Increasing or decreasing exchange rate of a major currency by 6 % e) Increasing or decreasing exchange rate of a minor currency by 20 %. The point is that the changes are so large that their impact is not likely to be estimable by delta and gamma or other Greek letters.

2. Stressing several variables: Since variables tend to move together, a common practice is to use extreme movements in market variables that have occured in the past. For example, for US equity, one might set the percentage changes in all market variables equal to those on Oct 19, 1987 (when S&P moved by 22.3 SDs) or Jan 8, 1988 (when S&P moved by 6.8 SDs). For UK interest rates, one might set the percentage changes in all market variables equal to those on April 10, 1992 (when 10 year bond yield moves by 8.7 SDs). Another approach is to take a period when there were moderately adverse market movements and create a scenario where all variables move by three or five times as much. The problem with this approach is increasing market variables by a particular multiples does not increase correlation, which does not reflect what happens in a stressed condition (correlation increases there). Some scenarios are one day shock to market variables. Some, particularly those involving credit and liquidity, involve shocks that take place over several days, weeks or months. Finally it is important to include volatilities of the market variables in scenario analysis as extreme movements in market variables such as interest rates and exchange rates are typicaly accompanied by large increase in volatilities of these variables and other variables as well.

3. Scenarios generated by management: In many ways, the most useful scenarios in stress testing are those generated by senior management or an economic group within the institution. They are in a good position to use their understanding of the markets, world politics, the economic environment and current global uncertainties to develop plausible scenarios that would lead to large losses. In this case, the scenarios are likely to be incomplete as the movements of only a few (core) market variables are specified. One approach is to regress other (peripheral) variables on the core variables being stressed to obtain forecasts on them conditional on the changes being made on the core variables. These forecasts can be incorporated into the stress test. This is known as conditional stress testing. It can be improved by specifying a correlation between the core and peripheral variables different from the normal market conditions.

4. Making scenarios complete: Scenarios should include not only the immediate effect on the portfolio but also the "knock-on" effect resulting from many different institutions being affected by the shock and responding in the same way (such as flight to quality when there is a severe liquidity problem and / or big increase in credit spreads).

Reverse stress testing : involves the use of computational procedures to search for scenarios that lead to a failure of the financial instituion. Typically one identify 5 to 10 key market variables and search for the combination of changes that result in the largest loss. In this process, one needs to pay attention to the plausibility of the combination of changes (a low volatility of the asset and a big change in its price may not be plausible in practice.) Another possiblity is to impose some structure on the problem, such as PCA technique and a search can be conducted to determine the changes in the principal components that result in large losses. The last approach is to determine a multiplier that can be applied to the changes in the market variables in a past scenarios that can result in large loss (e.g. an institution can withstand a scenario similar to 2008 but a 50 % increase in the changes may cause it problems).

Integrating stress testing and VaR calculations: Stress testing can be integrated into VaR by assigning a probability to each stress scenarios that is considered. For example, a probability p can be assigned to the n_s stress scenarios and a probability $1 - p$ can be assigned to the n_v VaR scenarios (from historical simulations). The total loss distribution which includes both the stress and VaR scenarios can then be constructed and the 99 percentile loss level can be reported. The probability to the n_s stress scenariosis a subjective choice. Therefore some guidelines for assigning them are as followed: a) 0.05 % : extremely unlikely b) 0.2 % very unlikely, but the scenario should be given about the same weight as the 500 scenarios used in the historical simulation c) 0.5 % Unlikely, but the scenario should be given more weight than the 500 scenarios used in the historical simulation. The probabilities can be assigned using a Baysian network. For example the probability of a scenario consisting of three events E_1, E_2, E_3 is $P(E_1)P(E_2|E_1)P(E_3|E_1, E_2)$.

Subjective versus objective probabilities: An objective probability for an event is a probability calculated by observing the frequency with which the event happens in repeated trials. A subjective probability for an event is derived from an individual's personal judgment about the change of it occuring. The probabilities in historical simulation (hence VaR and ES calculation) are objective and the probabilities in stress testing are subjective. This may lead analysts to prefer using historical data rather than performing scenario analysis (the possiblity of blaming the data versus taking responsibility if something goes wrong). On the other hand, if analysis is only done based on objective probabilities, risk management is inherently backward looking and fails to capitalize on the judgment and expertise of senior managers.

11 Chapter 23: Operational Risk

Defining operational risk: 1. As a residual risk after market risk and credit risk have been taken into account. This definition is too broad, it includes risk associated with entering new markets, developing new products, economic factors etc. 2. As risk arising from operations, such as risk of mistakes in processing transactions, making payments etc. This definition is too narrow: does not include major risks such as rogue trader.

Internal risks: those over which the company has control (which people to hire, what computer system to use, what controls are in place). Operational risks can be defined as all internal risks. It then includes risks arising from inadequate control such as rogue trader risk and employee fraud. Regulators favor including more than internal risks in the definition of operational risk, such as natural disasters, political and regulatory risks.

Interaction between operational risk and credit / market risks: Ex: When mistakes are made in loan documentation, losses result if and only if the counterparty defaults. When a trader makes mistakes, losses result if and only if the market moves against his position.

Basel categorization of operational risks: 1. Internal fraud 2. External fraud 3. Employment practices and workplace safety 4. Clients, products and business practices 5. Damage to physical assets 6. Business disruption and sytem failures 7. Execution, delivery and process management.

Determination of regulatory capital: 1. Basic indicator approach: operational risk capital is 15 % of annual gross income over last three years. 2. Standardized approach : Bank's activities are divided into 8 business lines : corporate finance, trading and sales, retail banking, commercial banking, payment and settlement, agency services, asset management and retail brokerage. The average gross income over last three years of each line is multiplied by a beta factor of that line and summed up to determine the toal capital. 3. AMA (Advanced measurement approach) Operational risk capital requirement is calculated by bank internally using qualitative and quantitative criteria.

Operational risk VaR using standardized approach: there are 56 combinations of the seven Basel risk types and the eight business lines. Banks must estimate one year 99.9 % VaRs for each combination and aggregate them using the approach in chapter 12 to determine a single one year 99.9 % operational VaR measure.

Implementation of AMA: Required by the Basel Committee to involve 4 elements: internal data, externald data, scenario analysis, business environment and internal control factors. These are used to estimate loss frequency distribution and loss severity distribution. Loss frequency is often modelled using a Poisson(λ) process and loss severity is often modelled using the log normal distribution (that is the loss process is a compound Poisson process).

1. Internal data: Usually there is not enough internal historical data to estimate loss frequency and severity. There are two types of losses: high frequency low severity, such as credit card fraud losses and low frequency high severity, such as rogue trading. The second type of loss is usually more focused on, as it creates the tail of the loss distribution and also the first type of loss is usually taken into account by the pricing of the products. A particular percentile of the loss distribution can be estimated as the corresponding percentile of the first type of loss plus the average of the second type loss.

2. External data: External data can be used to gauge the loss of one's own institution using the loss of other institutions. A scale adjustment should be made in this case. For example, an adjustment is

$$\text{Estimated loss for bank A} \;=\; \text{Observed loss for bank B} \;\times \left(\frac{\text{Bank A revenue}}{\text{Bank B revenue}}\right)^{\alpha},$$

where $\alpha = 0.23$. There are data consortia that shre data between banks and data vendors who collect publicly available data in a systematic way.

3. Scenario analysis: The aim of scenario analysis is to cover the full range of possible low frequency high severity losses. The loss severity and frequency are estimated by a risk management committee. In terms of frequency, a number of categories can be defined, such as $\lambda = \frac{1}{n}$ for scenarios happening once every n years on average where $n = 5, 10, 50, 100, 1000$. On the other hand, there is no model for determining losses and if data is not available, the parameters of the loss severity distribution have to be estimated by the committee. One approach is to estimate an average loss and a high loss that the committee is 99 % certain will not be exceeded. A log normal distribution can then be fitted to the estimate. Another is to use the

power law: $P(v > x) = Kx^{-\alpha}$ to calculate the extreme tail of the loss distribution.

4. Business environment and internal control factors: These factors should be taken into account for the estimate of loss severity and loss frequency.

Insurance: Banks can take out insurance for operational risks. There are two problems with insurance: moral hazard and adverse selection. Moral hazard is the risk for the insurance company that the existence of the insurance contract would cause the bank to behave differently than it otherwise would. For example, having insurance against robberies might cause the bank to relax its security measure. Adverse selection is the risk that as the insurance company attracting bad business. For example, bans without good internal controls are more likely to enter rogue trading insurance; bank without good external controls are morelikely to enter external fraud insurance.

12 Chapter 24: Liquidity risk

Liquidity trading risks: Not all assets are readily convertible into cash. For example, 100 million dollar position in a non-investment grade bond may be difficult to sell at close to the market price in one day. The price at which a particular asset can be sold depends on: a) The mid-market price of the asset (half of sum of bid and ask) b) How much of the asset to be sold c) How quickly it is to be sold d) The economic environment.

Market maker quote: A particular quote from a market maker is good for trades up to a certain size. Above that size, the market maker is likely to increase the bid-offer spread. This is because as the size of the trade increases, the difficulty of hedging the exposure created by the trade also increases.

Bid-offer pattern as a function of quantity trasacted: Offer price tends to increase and bid price tends to decrease as a function of quantity transacted (See Hull Figure 24.1). However, the opposite pattern can also be observed. For example, an individual investing money with a bank might get a better quote as the transaction size increases.

Predatory trading: The practice of trading in anticipation of another's company making a big transaction to profit off the price movement. For example, suppose company A needs to unwind a large position in an asset in the near future. Company B, if aware of this intention, can shor the same stock in anticipation of the price decline. This would make it more difficult for company A to unwind the position at competitive prices.

Liquidity black hole: A situation where liquidity dries up because everyone wants to sell and no one wants to buy, or vice versa.

Measuring liquidity: In dollar amount :

$$p = \text{ Offer price } - \text{ Bid price } .$$

As proportion of asset price :

$$s = \frac{\text{Offer price } - \text{ Bid price}}{\text{Mid market price}},$$

where

$$\text{Mid market price} = \frac{\text{Offer price } + \text{ Bid price}}{2}.$$

The volume of trading per day is also an important measure of liquidity. Lastly, the average of

$$\frac{\text{Absolute value of daily return}}{Dailydollarvolume}$$

over all days in the period consider can also be a measure of the liquidity of the period. This measure is widely used by researchers as it has the property that the asset's expected return increases as its liquidity decreases.

Cost of liquidating a position :

$$c = s\frac{\alpha}{2},$$

where α is the mid-market dollar value of the position. Thus $c = pn$ where n is the number of shares in the position (assuming the position only consists of share of the same asset). Cost of liquidating a multi-position portfolio:

$$c = \sum_{i=1}^{n} s_i \frac{\alpha_i}{2},$$

where n is the number of the positions. Note that this equation also shows that while diversity reduces the market risk, it does not necessarily reduce liquidity risk. Typically s_i increases with the size of the position i. Thus holding many small positions rather than a few large positions tends to entail less liquidity risk. Setting limits to the size of any one position can be one way to reduce liquidity risk.

Cost of liquidation in stressed market condition:

$$c_{stresed} = \sum_{i=1}^{n} \frac{(\mu_i + \lambda\sigma_i)\alpha_i}{2},$$

where μ_i, σ_i are the mean and SD of s_i. λ gives the required confidence level for the spread. For example, assuming the normality of the spreads at 99 % confidence level $\lambda = 2.33$. Without the normality assumption, λ can be estimated using the empirical distribution, e.g. $\lambda = 3.6$ for some heavier tail distribution. The stressed liquidation cost formula also assumes perfect correlation in the bid-ask spreads of the instruments. This is not unreasonable as when liquidity is tight, bid-ask spreads tend to widen for all instruments. We also remark that the distribution os s_i also depends on the unwinding horizon : s_i tends to decrease as a function of the time required for liquidiation.

Liquidity-Adjusted VaR: To take into account the liquidity risk, it is the regular VaR plus the cost of unwinding positions :

$$\begin{aligned} \text{Liquidity-Adjusted VaR} &= \text{VaR} + \sum_{i=1}^{n} \frac{s_i \alpha_i}{2} \\ \text{Liquidity-Adjusted VaR (stresed)} &= \text{VaR} + \sum_{i=1}^{n} \frac{(\mu_i + \lambda\sigma_i)\alpha_i}{2}. \end{aligned}$$

Either definition can be used depending on the risk-management need.

Unwinding a position optimally: An optimization problem where one unwinds a position in n days. The variables to optimize over are q_i : the units traded on day i. Define $p(q_i)$ as the bid-ask spread when the trader trades q_i units of asset. Let x_i be the position of the trader at the end of day i : $x_i = x_{i-1} - q_i$. Here $x_0 = V$, the total number of assets to be liquidated. The optimizing problem is then

$$\min_{q_i} \lambda \sqrt{\sum_{i=1}^{n} \sigma^2 x_i^2} + \sum_{i=1}^{n} q_i \frac{p(q_i)}{2},$$

where σ is the SD of the mid market price change, which is assumed to be normal and λ is the confidence level in the VaR estimate. This formulation can be viewed as minimizing both the variance of the price change applicable to the unwind:

$$\sum_{i=1}^{n} \sigma^2 x_i^2$$

plus the cost of unwinding :

$$\sum_{i=1}^{n} q_i \frac{p(q_i)}{2}$$

(each trade is assumed to cost half of the bid offer spread). The q_i can be determined via a numerical algorithm.

Leveraging: Referring to the phenomenon where banks have lots of liquidity, so they make credit easily available to customers. The credit spreads decrease and availability of credit makes prices increase (by increasing their demands). Assets are often pledged as collaterals and as their prices increase, the values of collaterals increase and borrowing can increase further. This leads to further asset purchases and the cycle repeats. Delveraging refers to the opposite phenomenon.

The impact of regulatio and the importance of diversity: Regulation has a positive effect. On the other hand, it tends to make different institutions responding in the same way to external events. This reduces the independence and increases the correlation of the institutions. This in turn may lead to a liquidiy black hole. Diversity is the opposite of correlated actions and one way of creating diveristy is to recognize that different types of institutions have different types of risks and should be regulated differently. Hedge funds tend to add diversity to the market. On the other hand, they also tend to be highly leveraged. Thus when liquidity tightens, all hedge funds have to unwind their positions which may accentuate the liquidity problem.

13 Chapter 25: Model risk

Model risk: is a type of operational risk. Two main types: pricing and hedging. Pricing risk is the risk of giving the wrong price at the time the product is bought or sold. Hedging risk is the risk of computing the wrong Greeks for hedging the position. Models thus are used for pricing and hedging and these are where the risks arise.

The mark to market process : Financial institutions want to be sure they price nonstandard products consistently with the rest of the market.

Linear model examples: Kidder Peabody's buying the spot and selling the forward and rolling the contract over. The funding cost was not accounted for by the system and the position is registered as a huge profit. Faulty model assumption: assuming that the forward rates will be realized in a swap contract. This works if the rate is given at time t_i and payment received at time t_{i+1}. If the payment is received at time t_i then a convexity adjustment has to be made and the assumption that the forward rate is realized is not justified.

Physics vs finance: Many financial models also arise as physical models. The difference is if used in physical context, the parameters of the models do not change.

In the financial context, the parameters change daily. Thus daily calibration has to be performed for financial models and not physical models.

Using models to price standard products: Standard products are those that are traded actively. Thus their prices can be read off from the market. The models are used as an interpolation for a standard product whose parameters are not exactly the same as those currently traded (strike, maturity etc.) In terms of option pricing, the Black-Scholes model so far is the most popular. Issues to consider for Black-Scholes :volatility smile, volatility surface. Another use for models in standard products is for hedging purposes. In this sense, there is model related hedging risk even for standard products.

Using models to price non-standard products: Non-standard products are those that are not traded actively (or not at all). This makes the model risk greater for non-standard products as there are both pricing and hedging risks. With standard products, there is usually only hedging risk as the prices are known. An example is pricing a BBB-tranche of an ABS, where the assumption that it behaves like a BBB-bond is not a good one.

Quantifying model risk: A financial institution should use several different models for pricing a non-standard product whenever possible. The price given to the client should be the max price given by the models and profit (as calculated by the difference between the max and min prices of the models) should not be registered immediately as such. The important question to answer is : "What range of model prices is possible for the models that price actively traded products correctly?" This analysis, if carried out succesfully, will give a worst and best possible prices for non-standard instrument. Research into this type of model risk is still limited. One approach is weighted Monte Carlo, which involves applying weights to the sampled paths. Constraints are placed on the weights so that standard instruments are priced correctly. An optimization procedure is used to find the weights that produce max and min values for the non-standard products with the constraints.

Dangers in model building: a. Overfitting: When all prices are matched, the model might exhibit some other unreasonable properties. For example, the Black-Scholes model can be extended to match the vol surface exactly (Dupire, Derman and Kani). On the other hand, the joint distribution of the assets at different time points may be less reasonable than the simpler models. b. Over-parametrization: Extending the Black-Scholes to include stochastic volatiltiy or jumps introduces extra parameters to be estimated. These extra parameters may be more stable than those in simpler models and not require daily adjustments. This may be true until a regime shift happens. The more complicated model may not then have the flexibility to adapt to changing market conditions. Also traders like simple models with one unobservable parameter. The more complex models are like black boxes which make it hard for them to develop intuition about.

Detecting model problems: An institution should keep track of : a) The type of trading it is doing with other institutions b) How competitive it is in bidding for different types of structured transactions c) The profits from trading different products. The institution should be concerned if it gets too much of a certain type of business, or making huge profits from relatively simple trading strategies. In short, if they find that their prices are out of line with the market, they must make an effort to their mark-to-market procedures to bring them back into line.