REPORT COPULA

Contents

Chapter 1

INTRODUCTION

Chapter 2

RISK MEASURES

2.1 Value at Risk

Assume that the expected relative return r with the time t of the investment I is a random variable determined by the distribution function F: $\Omega \rightarrow [0,1]$,

$$F(x) = \int_{-\infty}^{x} p(r)dr, \ (1)$$

Where the corresponding density of probability is p. In fact, this implies that the anticipated rate of return r(t) will reach a value no more than x%$(x \in R) att + 1$ with probability $P(r(t) \leq x) = F(x)$.

Let $\overline{\Omega} = R$ be the same space of price asset R = r(t)I . The expected loss with regard to the time horizon t of the investment I, the negative difference is then given as L(t) = μI-R(t), where μ is mean. Note that any return lower than expected implies an effective loss, even though it is positive. Positive values of L(t) mean a loss at time t+1. Its distribution of function $\overline{F} = \overline{\Omega} \rightarrow [0,1]$ is given by $\overline{F} = 1 - F\left(\frac{L}{I} - \mu\right)$,or

$$\overline{F}(L) = 1 - \int_{-\infty}^{L} p\left(\frac{L'}{I} - \mu\right) dL', \ (2)$$

where $-p\left(\frac{L'}{I} - \mu\right)$ is the probability density. Thus, the Value at Risk at time t of the investment I is defined as the maximum expected loss L(t) not surpass with probability (1-α):

$$P(L(t) \leq VaR) = 1 - \alpha, \ 0 \leq \alpha \leq 1 (3)$$

According to the Basel III, the measurement t is equal to 10 days and $\alpha = 1\%$. The Value at Risk is referred to as the unexpected loss of investment I. Then $P(L(t) \leq VaR = \overline{F} \leq VaR)$) and the VaR is the (1-α) – quantile random variable L(t):

$$VaR = \overline{F}^{-1}(1-\alpha), \text{ (4)}$$

2.2 Expected shortfall

Expected shortfall is also known as CVaR, which is the expected loss in the tail, is defined as the average of the losses that exceeds VaR.

$$CVaR^{\alpha} = E\left(\frac{L}{L} \geq VaR^{\alpha}\right) = ES$$

2.3 Estimating VaR

2.3.1 Historical Simulation

The goal of this approach is to produce a future distribution of potential historical data-based future scenarios. The data typically consists of regular returns, reaching out for a certain span, for all potential properties. The regular return simulation for a day in the future is then simply achieved by selecting one of the historical returns by chance (uniform distribution!). Firstly, we classify the various instruments in our portfolio in order to apply this method and obtain a sample of their historical log returns over some duration of observation. In our current portfolio, we then use the weights to simulate the discounted loss distribution (L_t). Therefore, there are no parametric model assumptions are based on the HS approach. However, it relies on the stationary existence of the historical data set to ensure that the empirical distribution of discounted losses converges with the true distribution of discounted losses. Secondly, for this method, we assume this historical distribution of log returns over the next holding duration is a good estimate of the distribution. In the other words, we believe that the historical data will maintain for a long time. Finally, The related sum from the historical log returns distribution would take us to the predicted VaR portfolio.

Suppose that we observe the price of asset i at time j P_i^j, j is running in the period from 1 to N. Let r_i^j be the log return. The appropriated weight of asset i over the period j in n assets in our portfolio is w_i. Moreover, the historically simulated loss series $\{L_t^j\}_{(j=1)}^N isgeneratedof L_t = -\frac{X}{R_t} = V_t - V_{\frac{(t+1)}{R_t}}, whereV_t = \sum_{(i=1)}^n w_i P_i^t andV_{(t+1)}^j = \sum_{(i=1)}^n w_i P_i^t e^{(R_{(i,j)})} andR_{(i,j)}$ are the historically simulated log return. In fact, from the discounted loss distribution, the VaR portfolio is assumed. There are a variety of attractions to the HS approach; it is conceptually straightforward and fast to enforce. In comparison, it does not focus on the presumption that log returns are allocated so it requires the results to speak for themselves and decide the distribution. HS needs a significant volume of data to ensure adequate accuracy of the calculation. However, the obtaining of such big, suitable sample data is not always technically feasible; and even if the past is present, it does not repeat itself or include adequate extreme observations for the VaR calculation.

2.3.2 Variance – Covariance method

This VaR strategy assumes that a regular distribution is accompanied by the frequent price returns for a given position. We measure the standard deviation and covariance of the log returns of certain predefined risk factors and the portfolio's vulnerability to such risk factors based on the distribution of daily returns derived from the daily price series. The regular risk VaR value is simply a function of the standard deviation and the target degree of trust. Suppose that we build a portfolio which has n assets with r_i as the log return of each asset, w_i is relative weight of each asset, thus the portfolio return and variance of this portfolio are R_p and σ_p^2 at time t are defined as:

$$R_p = \sum_{i=1}^n w_i r_i$$

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij}$$

where σ_{ij} is the covariance between two assets i and j at time t. We assume that X is normal distribution $X\ N(\mu, \sigma^2) withX = V_{(t+1)} - V_t$, after using ellipticity property of normal distribution, we have the outcoming that is $X =^d \mu + \sigma$Z with Z N(0,1),

$$VaR_\alpha(X) = -\mu + \sigma\phi^{-1}(1-\alpha)$$

where $\mu = E[R_p], \sigma = \sigma_p$ and φ is known as the distribution function of Z. Where normality is assumed, VaR calculation is relatively simple, as standard mathematical properties of normal distribution can be used to measure VaR at various levels. Furthermore, the presumption of normality enables basic translatability between various stages and holding times. For various time horizons, VaR can be modified by $VaR_{t_2} = \sqrt{\frac{t_2}{t_1} VaR_{t_1}}$, thus the different confidence levels can be defined as $VaR_\alpha = \frac{\phi^{-1}(\alpha)}{\phi^{-1}(\alpha_1)} VaR_{\alpha_1}$. The most prevalent drawback of the strategy, though, is the presumption of normality. Many financial assets are considered to have distributions of fat tailed log return, indicating that extreme outcomes are more common in practice than the usual distribution would imply. As a consequence, the calculation of VaR may be understated.

Chapter 3

METHODOLOGY

3.1 Extreme Value Theory

Many of the challenges of risk assessment are fundamental to the mathematical study of extremes. In order to analyze some of the fundamental principles underlying EVT, under the simplifying assumption, we address the most important results: losses will be considered to be distinct and distributed identically. Many of the observations can be generalized to much more general models. However, in this context, we mention two approaches for extremal events models, which are Generalized Pareto Distribution (GPD) and The Peak Over Threshold Method (POT).

3.1.1 Generalized Pareto Distribution (GPD)

The GPD defines the restricted distribution over a certain threshold for modeling excesses. If X is a random variable distributed by Pareto, its distribution function has the shape of its distribution function

$$G_{\gamma,\beta}(x) = \begin{cases} 1 - \left(1 - \frac{\gamma x}{\beta}\right)^{\frac{-1}{\gamma}} (\gamma \neq 0) \\ 1 - e^{\frac{-x}{\beta}} (\gamma = 0) \end{cases}$$

where γ is the shape parameter and β is the scale parameter, $\beta > 0$ and x$\geq$0 when $\gamma \geq 0$ and 0$\leq$x$\leq \frac{-\beta}{\gamma}$ when $\gamma < 0$. The GPD is extended in the sense that a variety of special cases are used:

- When $\gamma = 0$, our GPD become exponential distribution (a light tailed).
- When $\gamma > 0$, we have a heavy tailed on ordinary Pareto distribution.
- When $\gamma < 0$, the distribution function $G_{\gamma,\beta}$ has type II of a short tailed Pareto distribution.

If x is fixed, the parametric form is continuous in γ; therefore, $lim_{\gamma\to 0} G_{\gamma,\beta}(x) = G_{0,\beta}(x)$ and by adding a location parameter $\mu \in R$ can spread as:

$$G_{\gamma,\mu,\beta}(x) = G_{\gamma,\mu,\beta}(x - \mu)$$

When $\beta > 0$, in this case we take β= 1 and μ= 0. The GPD is described as standard GPD, which has a form:

$$G_{\gamma,\beta}(x) = \begin{cases} \frac{1}{\beta}\left(1 - \frac{\gamma x}{\beta}\right)^{\frac{-\gamma-1}{\gamma}} & (\gamma \neq 0) \\ \frac{1}{\beta} e^{\frac{-x}{\beta}} (\gamma = 0) \end{cases}$$

If γ increases, the tail will be fatten and the peaks are sharpening. If β increases, the central part of density will be more flat.

3.1.2 Distribution of Exceedances – The Peak Over Threshold Method.

The Generalized Pareto distribution can be constructively defined in term of exceedances. We can sample random values independently of that distribution, beginning with a probability distribution whose right tail drops to 0, such as the normal one. When a threshold value is set, all the values below the threshold are thrown out, and the threshold is subtracted from the values not thrown out, the effect is known as exceedances. The allocation of surpluses is roughly a Generalized Pareto. In the left tail of a distribution, we can set a threshold and disregard all values above that threshold. For the estimate to be rational, the threshold must be far enough beyond the tail of the initial distribution.

For the series $\{X_i\}_{i=1}^n$ of dependent and identical random variables distribution of financial losses in common but unidentified distribution of function F and $x_F = sup\{x \in$

$\mathbb{R} : F(x) < 1\} \leq \infty$ is upper endpoint. Let u be a threshold and denote by $N_u = Card\{i : X_i > u, i = 1, 2, \ldots, n\}$ the number of exceedances of u by $\{X_1, X_2, \ldots, X_n\}$.

- Distribution of function F get heavy tail when the tail probability F(x) decays slowly as x decreases:

$$\lim_{x \to -\infty} \frac{F(x)}{e^{-\lambda(-x)}} = \infty \text{ for every } \lambda > 0$$

Get a heavy right tail when:

$$\lim_{x \to \infty} \frac{\overline{F}(x)}{e^{-\lambda x}} = \infty \text{ for every } \lambda > 0$$

- Distribution of function F get a regularly varying right tail $\overline{F} = 1 - F$ when a number ρ exists such that

$$\lim_{x \to \infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{\rho} \text{ for every } x > 0$$

Get a regularly varying left tail when:

$$\lim_{x \to -\infty} \frac{F(tx)}{F(t)} = x^{\rho} \text{ for every } x > 0$$

We can set $\rho = -\alpha$ for $\alpha \geq 0$, since $\overline{F}(x)$ is decreasing in x. Then we have:

$$\lim_{t \to \infty} P(X > tx \mid X > t) = x^{-\alpha} \text{ for every } x > 1$$

We suppose that function F of the series $\{X_i\}_{i=1}^n$ has regularly varying right tail. From the Theorem 7 [4], state that for the large u approaching x_F, the excesses $y = X_k - u$ with F_u is distribution function which are approximated by GPD:

$$\lim_{u \to x_F} \sup_{0 \leq y \leq x_F - u} \mid F_u(y) - G_{\gamma, \beta(u)}(y) \mid = 0$$

for some $\beta(u)$ is positive. The Peak Over Threshold Method use this result to create estimates of tail probability and quantiles. When $\beta > 0, \gamma > 0$ the GPD is known as:

$$G_{\gamma,\beta}(x) = 1 - \left(1 + \frac{\gamma x}{\beta}\right)^{\frac{-1}{\gamma}} \text{ for } x \geq 0$$

Assume that X is a random variable respect with distribution of function F, which has a regularly varying right tail to $\lim_{u \to \infty} \frac{\overline{F}(\lambda u)}{\overline{F}(u)} = \lambda^{-\alpha} \forall \lambda > 0$ and some $\alpha > 0$.

$$\lim_{u\to\infty} P\left(\frac{X-u}{\frac{u}{\alpha}} > x \mid X > u\right) = \lim_{u\to\infty} \frac{P\left(X>u\left(1+\frac{x}{\alpha}\right)\right)}{P(X>u)} = \left(1-\frac{x}{\alpha}\right)^{-\alpha} = \overline{G}_{\frac{1}{\alpha}1}(x)$$

X over the threshold u has the excess distribution function:

$$F_u(x) = P(X-u \le x \mid X > u) \text{ for x ¿ 0}$$

This can be considered as:

$$\overline{F}_u(x) = P(X > u+x \mid X > u) = \frac{(P(X>u+x, X>u))}{P(X>u)} = \frac{\overline{F}(u+x)}{\overline{F}(u)} = \frac{\overline{F}(u\left(1+\frac{x}{u}\right))}{\overline{F}(u)}$$

Since $\overline{F}$ is regularly varying with index $-\alpha < 0$, it holds that $\frac{\overline{F}(\lambda u)}{\overline{F}(u)}$ uniformly in $\lambda \ge 1$ as u $\to \infty$

$$\lim_{u\to\infty} \sup_{\lambda\ge 1} \mid \overline{F}(u) - \lambda^{-\alpha} \mid = 0$$

when $\beta(u)\ \frac{u}{\alpha}, \gamma = \frac{1}{\alpha}$ as $x \to \infty$. Because we are finding $\overline{F}(u+x)$ and F^{-1} (p) for large u and p, the we get:

$$\overline{F}(u+x) = \overline{F}(u)\overline{F}_u(x)$$

In the case u is near the tail, then the empirical approximation $\overline{F}(u) \approx \overline{F}_n(u) = \frac{N_u}{n}$ is exact, then we have:

$$\overline{F}_u(x) \approx \overline{G}_{\gamma,\beta(u)}(x) \approx \overline{G}_{\hat{\gamma},\hat{\beta}(x)}(x) = \left(1+\frac{\hat{\gamma}x}{\hat{\beta}}\right)^{\frac{-1}{\hat{\gamma}}}$$

where parameter γ and β estimates $\hat{\gamma}$ and $\hat{\beta}$. After estimating $\overline{F}_u(x)$ and $\overline{F}(u)$ for the tail of function F. The estimator of $\overline{F}(u+x)$ can be given by:

$$\overline{F}(u+x) = \frac{N_u}{n}\left(1+\frac{\hat{\gamma}x}{\hat{\beta}}\right)^{\frac{-1}{\hat{\gamma}}}$$

Applying the definition of quantile, we get the result:

$$\begin{aligned}
\hat{\overline{F}^{-1}}(p) &= min\{x : \hat{\overline{F}}(x) \le 1-p\} \\
&= min\{x+u : \hat{\overline{F}}(x+u) \le 1-p\} \\
&= u + min\{x : \frac{N_u}{n}\left(1+\frac{\hat{\gamma}x}{\hat{\beta}}\right)^{\frac{-1}{\hat{\gamma}}} \le 1-p\} \\
&= u + \frac{\hat{\gamma}}{\hat{\beta}}\left(\left(\frac{n}{N_u}(1-p)\right)^{-\hat{\gamma}} - 1\right)
\end{aligned}$$

3.2 AR-GARCH Model

3.2.1 Autoregressive Model

An autoregressive model is one where a variable's current value, y, depends only on the values that were taken by the variable in previous times, plus an error term. It is possible to articulate an autoregressive model of order p, denoted as AR(p):

$$\gamma_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + u_t$$

where u_t is a white noise disturbance term. In order to demonstrate the properties of an autoregressive model, manipulation of speech would be necessary. It is possible to write this expression more compactly using sigma notation:

$$\gamma_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t$$

or using the lag operator [*], as:

$$\phi(L)\gamma_t = \mu + u_t$$

where $\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)$.

3.2.2 GARCH Model

In this context, a variety of significant characteristics common to many financial data can also not be described by linear structural (and time series) models, including:

- Leptokurtosis – that is, the tendency for financial asset returns to have distributions that exhibit fat tails and excess peakedness at the mean.
- Volatility clustering or volatility pooling – the tendency for volatility in financial markets to appear in bunches. Thus large returns (of either sign) are expected to follow large returns, and small returns (of either sign) to follow small returns. A plausible explanation for this phenomenon, which seems to be an almost universal feature of asset return series in finance, is that the information arrivals which drive price changes themselves occur in bunches rather than being evenly spaced over time.

- Leverage effects – the tendency for volatility to rise more following a large price fall than following a price rise of the same magnitude.[*]

The 'ARCH' model (ARCH stands for' autoregressive conditionally heteroscedastic ') is a particular non-linear model in common use in finance and expressed as:

$$\sigma_t^2 = w + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2$$

where $w > 0, \alpha_j \geq 0$ for j=1,2,...,q so as to the conditional variance keep being positive. However, to fit the data, we need a large q. Therefore, Bollerslev (1986) and Taylor independently developed the GARCH model (1986). The GARCH(p,q) model enables the conditional variance to draw on prior lags of its own, such that the conditional variance equation is in the simplest case, which model is given by:

$$\sigma_t^2 = w + \sum_{i=1}^{p} \beta_i \sigma_{t-j}^2 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2$$

where $w > 0, \alpha_j \geq 0, \beta_i \geq 0$ for j=1,2,...,q and i=1,2,...,p.

Nelson has suggested the exponential GARCH (EGARCH) concept (1991). There are different ways of expressing the equation of the conditional variance, but one possible specification is given by:

$$\ln \sigma_t^2 = w + \sum_{i=1}^{p} \beta_i \ln \sigma_{t-i}^2 + \sum_{j=1}^{q} \frac{\gamma_j \epsilon_{t-j}}{\sigma_{t-j}} + \sum_{j=1}^{q} \frac{\alpha_j \mid \epsilon_{t-j} \mid}{\sigma_{t-j}} - E\left(\frac{\mid \epsilon_{t-j} \mid}{\sigma_{t-j}} \right)$$

3.2.3 GJR Model

With an additional concept applied to account for potential asymmetries, the GJR model is a basic extension of GARCH. Now the conditional variance is given by:

$$\sigma_t^2 = w + \sum_{i=1}^{p} \beta_i \sigma_{t-j}^2 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j}^2 I_{\epsilon_{t-j<0}}$$

where $w > 0, \alpha_j \geq 0, \beta_i \geq 0$ for j=1,2,...,q and i=1,2,...,p and $\gamma_j + \alpha_j \geq 0$.$I_{\{.\}}$ denoted the function of the indicator that returns one if the threshold amount is met and zero otherwise.

3.3 Copula Theory

3.3.1 Definition and Property

We denote the uniform distribution on an interval (0,1) by U(0,1), i.e., the probability of a random variable U satisfying $P(U \leq u) = u$ for $u \in (0, 1)$

Proposition: Let F be a distribution function on $\mathbb{R}$. Then

1. $u \leq F(x)$ if and only if $F^{-1}(u) \leq x$.
2. If F is continuous, then $F(F^{-1}(u)) = u$
3. (Quantile Transform) If $U \backsim U(0, 1)$ then $P(F^{-1}(U)) \leq x = F(x)$
4. (Probability Transform) If X has a distribution function F, then $F(X) \backsim U(0, 1)$ if and only if F is continuous.

Definition: A d-dimensional copula is the distribution function C with $C : [0, 1]^d \to [0, 1]$ of a random vector U whose components U_k are uniformly distributed:

$$C(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d)$$

Let $(X_1, .., X_d)$ be the random vector which has the distribution function $F(x_1, \ldots, x_d) = P(X_1 \leq x_1, .., X_d \leq x_d)$, and we assume that a continuous function for every k is $F_k(x) = P(X_k \leq x)$. The components of the vector U: $(U_1, .., U_d) = (F_1(X_1), .., F_d(X_d))$ are uniformly distributed by the statement (iv) in proposition. the distribution function C of U is a copula which be called copula function of X. From statement (i) in the proposition, we have:

$$\begin{aligned} C(F_1(x_1), .., F_d(x_d)) &= P(U_1 \leq F_1(x_1), \ldots, U_d \leq F_d(x_d)) \\ &= P(F^{-1}(U_1)) \leq x_1, \ldots, P(F^{-1}(U_d)) \leq x_d \\ &= F(x_1, \ldots, x_d) \end{aligned}$$

The above equation is the result from the Sklar's theorem[8]. **Copula Density function:**

$$c(u_1, \ldots, u_d) = \frac{\partial C(u_1, \ldots u_d)}{\partial u_1 \ldots u_d}$$

The copula is associated with the density function of F, which is denoted by f, so called canonical copula:

$$f(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) \prod_{j=1}^{d} f_j(x_j)$$

where f_j are the densities of the marginal $f_j = \frac{dF_j(x_j)}{dx_j}$.

3.3.2 Elliptical Copulas and Archimedean Copulas

In this study, we research two main copulas

Elliptical Copulas

Copulas with elliptically contoured (or elliptical) distributions are known as elliptical copulas. The multivariate normal and Student-t distributions are the most widely used elliptical distributions. The biggest advantage of an elliptical copula is that varying degrees of similarity between the marginals can be defined. However, elliptical copulas lack closed form expressions and are limited to radial symmetry. For elliptical copulas the relationship between the linear correlation coefficient r and Kendall's tau t is given by:

$$\rho(X, Y) = \sin\left(\tfrac{\pi}{2}\tau\right)$$

Archimedean Copulas

a d-variate Archimedean copula as the function defined as:

$$C(u_1, \ldots, u_d) = \phi^{-1}(\phi(u_1) + ... + \phi(u_d))$$

where $\phi(u)$ is called copula's generator such that $\phi(u_d) \in C^2$ with $\phi(1) = 0, \phi'^{(u)} < 0$ when ϕ is strictly deceasing and $\phi''^{(u)} > 0$ when ϕ is convex for all $0 \leq u \leq 1$, and ϕ^{-1} must be completely monotonic $[0,\infty)$[8].

3.3.3 Dependence Measure

Kendall's tau and Spearman's rho are two of the most well-known scale invariant measures of interaction, both of which measure a type of dependency known as concordance

[10]. Informally, two random variables are concordant if large values of one are correlated with large values of the other, and small values of one are associated with small values of the other.

Concordance

Let (x_i, y_i) and (x_j, y_j) be two observations from a vector (X,Y) of continuous random variables. We say that (x_i, y_i) and (x_j, y_j) are concordant if $x_i < x_j$ and $y_i < y_j$ or $x_i > x_j$ and $y_i > y_j$. Similarly, we say that We say that (x_i, y_i) and (x_j, y_j) are discordant if $x_i < x_j$ and $y_i > y_j$, or if $x_i > x_j$ and $y_i < y_j$.

Kendall's Tau

Let $(X_1, Y_1)$and (X_2, Y_2) be i.i.d. random vectors, each with joint distribution function H. The Kendall's Tau is knowns as:

$$\tau = \tau_{X,Y} = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0)$$

Let Q be concordance function. In the Theorem 5.1.1 page 159 [10], it shows that Q depends on $(X_1, Y_2)$and (X_2, Y_2) through their copulas

$$Q = Q(C_1, C_2) = 4 \iint_{I^2} C_2(u, v) dC_1(u, v) - 1$$

where C_1 is a copula of (X_1, Y_1) and C_2 is a copula of (X_2, Y_2) so that $H_1(x, y) = C_1(F(x), G(y))$ and $H_2(x, y) = C_2(F(x), G(y))$.

Spearman's Rho

Let $(X_1, Y_1), (X_2, Y_2)$ and (X_3, Y_3) be three independent random vectors with common joint distribution function H and copula X. The Spearman's Rho is defined as:

$$\rho_{X,Y} = 3(P((X_1 - X_2)(Y_1 - Y_3) > 0) - P((X_1 - X_2)(Y_1 - Y_3) < 0))$$

It means that he probability of concordance minus the probability of discordance for the two vectors$(X_1, Y_1), (X_2, Y_3)$ and a couple of vectors with the same margins but one vector has distribution function H, while the components of the other are independent as well as (X_3, Y_3) could be equal. From theorem 5.1.6 page 167 [10], when continuous random variable X and Y whose copula is C:

$$\rho_{X,Y} = \rho_C = 12 \iint_{I^2} C(u, v) du dv - 3$$

Copula Estimation