Lecture 4: Value at Risk, Expected Shortfall & Risk Management

Instructor: Lin Juan

SOE & WISE, Xiamen University

Classification of Financial Risks (Basel Accord II)

- Credit risk
  It arises due to a debtor's failure to satisfy the terms of a borrowing arrangement.
- Market risk: risk due to changes in stock prices, interest rates, exchange rates and commodity prices
  For market risk, the preferred approach is VaR.
- Operational risk
  It is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems.

market risk

- high-quality data are available
- easier to understand
- the idea applicable to other types of risk.

Risk measure

- Let $L_t(\ell)$ be the loss variable of a financial position from time t to time $t + \ell$. Here ℓ denotes a holding period.
- Financial loss is concerned with the distribution of $L_t(\ell)$.
- All risk measures available in the literature are summary statistics of $L_t(\ell)$.

Coherence

Let X and Y denote the future values of two risky portfolios. A risk measure $\eta(\cdot)$ is said to be coherent if it satisfies the following properties:

- *Subadditivity*: $\eta(X+Y) \leq \eta(X) + \eta(Y)$.
- *Monotonicity*: If $X \leq Y$ almost surely, then $\eta(X) \geq \eta(Y)$.
- *Positive homogeneity*: For $c > 0$, $\eta(cX) = c\eta(X)$.
- *Translation invariance*: If A is a deterministic portfolio with guaranteed return a. For $c > 0$, $\eta(X + A) = \eta(X) - c$.

Subadditivity

- *Subadditivity*: $\eta(X+Y) \leq \eta(X) + \eta(Y)$.
- The risk of two portfolios together cannot get any worse than adding the two risks separately: this is the diversification principle. In financial risk management, sub-additivity implies diversification is beneficial.

Monotonicity

- *Monotonicity*: If $X \leq Y$ almost surely, then $\eta(X) \geq \eta(Y)$.
- If portfolio Y always has better values than portfolio X under almost all scenarios then the risk of Y should be less than the risk of X.
- In financial risk management, monotonicity implies a portfolio with greater future returns has less risk.

Positive homogeneity

- *Positive homogeneity*: For $c > 0$, $\eta(cX) = c\eta(X)$.
- Loosely speaking, if you double your portfolio then you double your risk. In financial risk management, positive homogeneity implies the risk of a position is proportional to its size.

Translation invariance

- *Translation invariance*: If A is a deterministic portfolio with guaranteed return a. For $a > 0$, $\eta(X + A) = \eta(X) - a$.
- The portfolio A is just adding cash a to your portfolio X.
- In financial risk management, translation invariance implies that the addition of a sure amount of capital reduces the risk by the same amount.

Definition of the risk

- The American Heritage dictionary, Fourth Edition, defines risk as "the possibility of suffering harm or loss; danger".
- In finance, harm or loss has a specific meaning: decreases in the value of a portfolio.
- Value-at-Risk (VaR) is a better measure of risk of the portfolio than the variance since it focuses on losses.

Value-at-Risk (*VaR*)

- The *VaR* of a portfolio measures the value which an investor would lose with some small probability p, over a selected period of time.
- Because *VaR* represents a hypothetical loss, it is usually a positive number.
- p is a small probability that usually takes value between 1% and 10%; Correspondingly, $1-p$ is a large probability (confidence level). For example, $p = 5\%$ and $1-p = 95\%$.

Definition of *VaR*

- Suppose that at the time index t, we are interested in the risk of a financial position for the next ℓ periods.
- Let $\Delta V_t(\ell) = V_{t+\ell} - V_t$ be the change in value of the underlying assets of the financial position from time t to $t + \ell$.
- Let $L_t(\ell)$ be the associated loss function, where $L_t(\ell) > 0$.
- $\Delta V_t(\ell)$ and $L_t(\ell)$ are measured in dollars for now.

long position and short position

- In finance, a long (or long position) is the buying of a security such as stock, commodity or currency with the expectation that the asset will rise in value.
- a short (or short position) is selling first and then buying later with the expectation that the price of the asset will drop.
- For the long position, the loss occurs if the price goes down and $\Delta V_t(\ell) < 0$. For the short position, the loss occurs if the price goes up and $\Delta V_t(\ell) > 0$.

Short Selling (Short Position)

Short selling is the selling of a stock that the seller doesn't own, that is, the seller takes a negative position in a security.

Example: Let us assume that an investor believed that the stock of ABC company, which currently sells for $100 per share, is likely to be selling for $95 per share (expected value) at the end of the year.

Short Selling

Example Continue... If the investor bought one share of ABC stock (long position) , the cash flow would be $-\$100$ at time zero when the stock is purchased and \$3 from the dividend, plus \$95 from selling the stock at time 1.

The cash flows are

	Time	
	0	1
Purchase Stock	−100	
Dividend		+ 3
Sell Stock		+95
Total Cash Flow	−100	+98

Short Selling

Example Continue... Unless this stock had very unusual correlations with other securities, it is unlikely that an investor with these expectations would like to hold any of it in his own portfolio. In fact, an investor would really like to own negative amounts of it.

How might the investor do so?

Short Selling

Assume a friend, Mike, owned a share of ABC company and Mike had different expectations and wished to continue holding it.

The investor might borrow Mike's stock under the promise that he will be no worse off lending him the stock.

The investor could then sell the stock, receiving \$100. When the company pays the \$3 dividend, the investor must reach into his own pocket and pay Mike \$3. He had a cash flow of $-\$3$. He has to do this because neither he nor Mike now owns the stock and Mike would be no worse off by lending him the stock.

Now at the end of the year, the investor could purchase the stock for \$95 and give it back to Mike.

The cash flows for the investor are,

	Time	
	0	1
Sell Stock	+100	
Pay Dividend		− 3
Buy Stock		−95
Total Cash Flow	+100	−98

Notice in the example that the lender of the stock is no worse off by the process and the borrower has been able to create a security that has the opposite characteristics of buying a share of the ABC company.

Definition of *VaR*

- Long position: the loss occurs if the price goes down and $\Delta V_t(\ell) < 0$. In this case, $L_t(\ell)$ is a negative function of $\Delta V_t(\ell)$, $L_t(\ell) = -\Delta V_t(\ell)$.
- Short position: the loss occurs if the price goes up and $\Delta V_t(\ell) > 0$. In this case, $L_t(\ell)$ is a positive function of $\Delta V_t(\ell)$, $L_t(\ell) = \Delta V_t(\ell)$.

Definition of *VaR*

- We define the *VaR* of a financial position over the time horizon ℓ with tail probability p as,

Definition 1

Let $L_t(\ell)$ be a random variable. Let p be a probability such that $0 < p < 1$,

$$p = \Pr[L_t(\ell) \geq VaR] = 1 - \Pr[L_t(\ell) < VaR]$$

Definition of *VaR*

- Let $L_t(\ell)$ be a random variable. Let p be a (small) probability such that $0 < p < 1$,

$$1 - p = \Pr[L_t(\ell) < VaR]$$

- VaR can be interpreted as,
  One is $(1 - p)\%$ confident that the potential loss encountered by the holder of the financial position will not exceed VaR over the time horizon ℓ [maximum].

Define *VaR* using quantile

- Let $F_\ell(x)$ denote the CDF of the loss function $L_t(\ell)$, where the subscript of t is omitted from F.
- *VaR* is concerned with the upper tail behavior of the loss CDF $F_\ell(x)$ (Definition 1, 2 and 3).

Define VaR using quantile

- Let $1-p$ denote the probability. The quantity

$$VaR_{1-p} = \inf\{x|F_\ell(x) \geq 1-p\}$$

  is called the $(1-p)$th quantile of $F_\ell(x)$, where inf denotes the smallest real number x satisfying $F_\ell(x) \geq 1-p$.
- We add the subscript to emphasize that VaR_{1-p} is the $(1-p)$th quantile of $F_\ell(x)$.

Definition of *VaR*

- Let $L_t(\ell)$ be a random variable. Let p be a probability such that $0 < p < 1$,

$$p = \Pr[L_t(\ell) \geq VaR_{1-p}]$$

- VaR can be interpreted as,
  There is a $p\%$ chance that the position holder would encounter a loss greater than or equal to VaR_{1-p} over the time horizon ℓ [minimum];

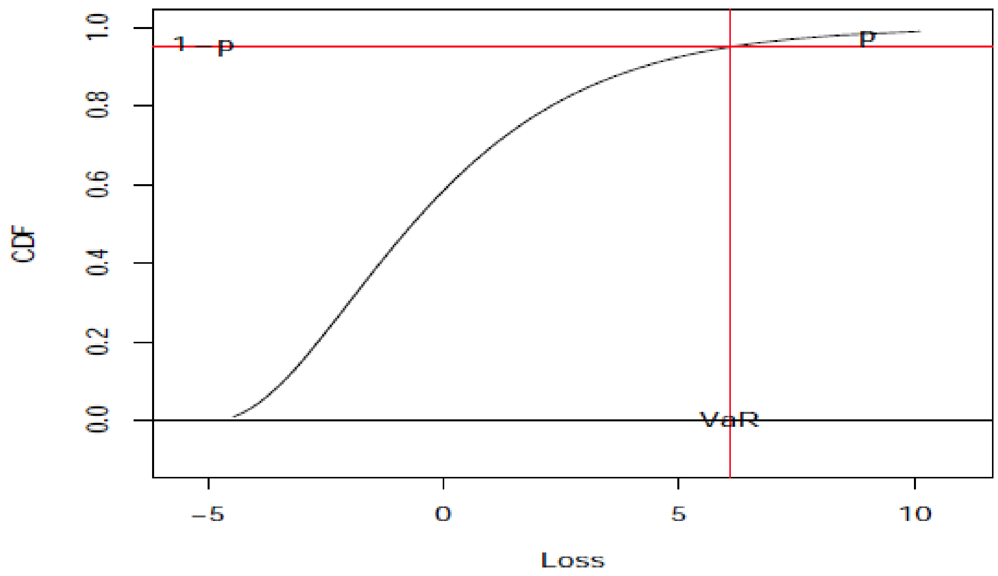


Figure 1: Definition of Value at Risk (VaR) for a continuous loss random variable based on the cumulative distribution function.

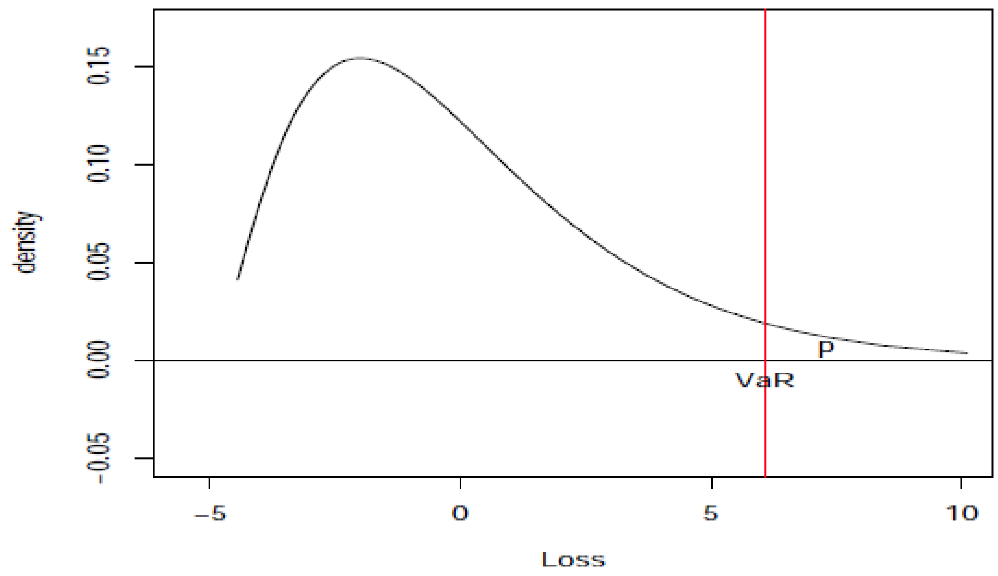


Figure 2: Definition of Value at Risk (VaR) for a continuous loss random variable based on the probability density function.

Definition of percentage *VaR*

Definition 2

Let $L_t(\ell)$ be a random variable. Let p be a probability such that $0 < p < 1$,

$$\begin{aligned} p &= \Pr[L_t(\ell)/V_t \geq \%VaR_{1-p}] \\ &= 1 - \Pr[L_t(\ell)/V_t < \%VaR_{1-p}] \end{aligned}$$

where $\%VaR_{1-p} = VaR_{1-p}/V_t$.

Definition of *VaR* of log returns

- Log returns correspond approximately to percentage changes in value of a financial assets.
- For a long financial position, loss occurs when the returns are negative.
- We use negative returns in data analysis for a long financial position.
- Let r_t denote the log returns on the portfolio. Let $x_t = -r_t$ for the long position and $x_t = r_t$ for the short position.

Definition of *VaR* of log returns

Definition 3

The p percentage Value-at-Risk ($\% VaR_{1-p}$) of a portfolio is defined as the largest return such that the probability that x_t over some period of time ℓ is greater than $\% VaR_{1-p}$ is p,

$$Pr(x_t \geq \% VaR_{1-p}) = p,$$

Define *VaR* using the lower tail

- Assume that one holds the long position;
- VaR can also be defined based on the lower tail behavior of $\Delta V_t(\ell)$ or r_t (Definition 4 and 5).

Define *VaR* using the lower tail

Definition 4

The p Value-at-Risk (VaR) of a portfolio is defined as the largest value such that the probability that the change in the value of the portfolio over some period of time ℓ is small than $-VaR$ is p,

$$Pr(\Delta V_t(\ell) < -VaR) = p,$$

where $\Delta V_t(\ell) = V_{t+\ell} - V_t$ is the change in the value of the portfolio.

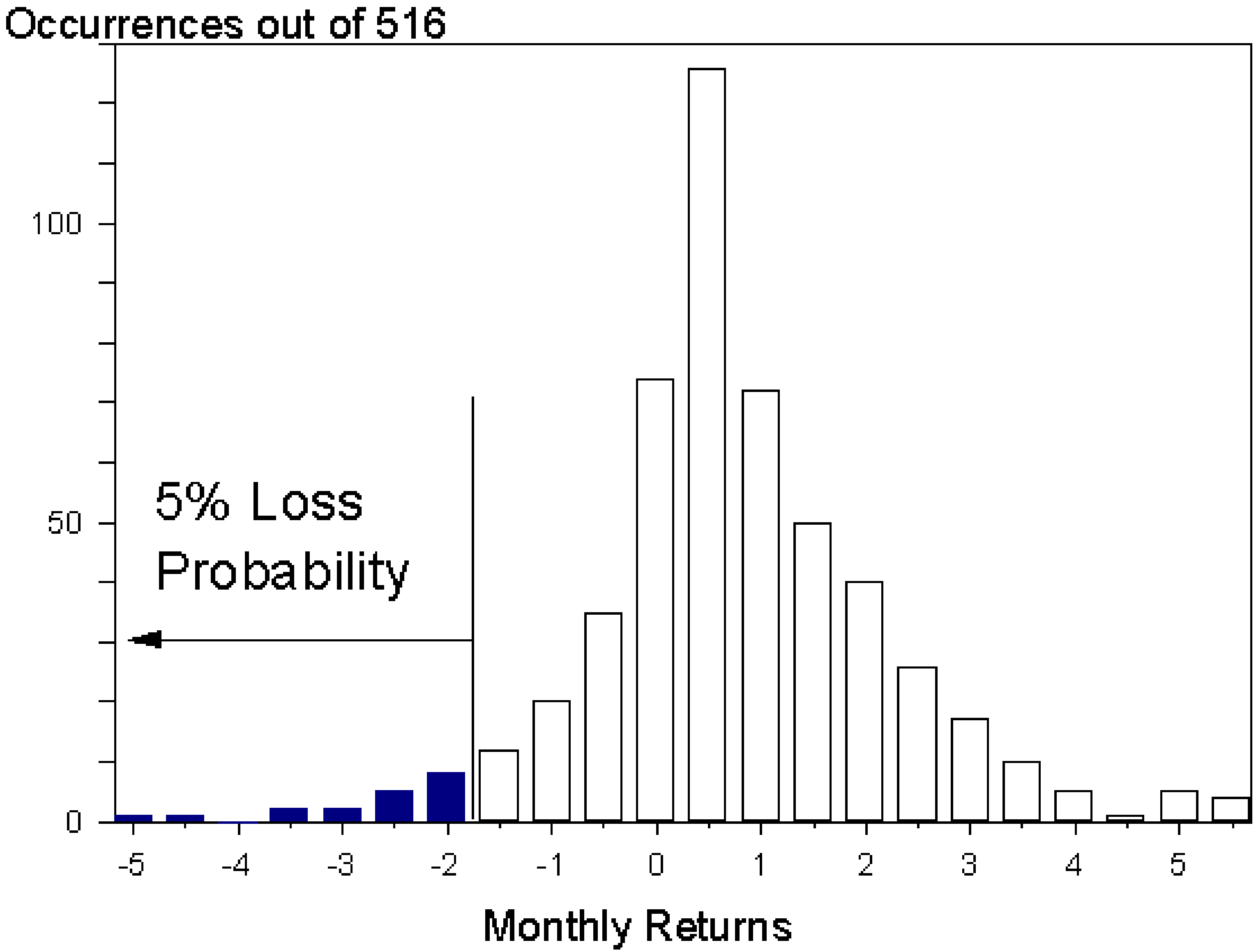
Define *VaR* using the lower tail

Definition 5

The p percentage Value-at-Risk ($\%VaR$) of a portfolio is defined as the largest return such that the probability that the return on the portfolio over some period of time ℓ is less than $-\%VaR$ is p,

$$Pr(r_t < -\%VaR) = p,$$

where r_t is the percentage return on the portfolio, $r_t = (V_{t+\ell} - V_t)/V_t$.


Occurrences out of 516
100
50
0
5% Loss
Probability
-5
-4
-3
-2
-1
0
1
2
3
4
5
Monthly Returns

Advantages of VaR

- Universal: VaR can be measured on portfolios of any type;
- Global: VaR summarize in a single number all the risks of a portfolio;
- Probabilistic: VaR provides a loss and a probability occurrence;
- Expressed in Lost Money: VaR is expressed in the best of possible units of measures: LOST MONEY;
- Some closed-form solutions of VaR are available;
- VaR is coherent for a normally distributed loss.

Some closed-form solutions of VaR

- For simplicity, we write $L_t(\ell)$ as L_t.
- We assume the mean and variance of L_t are constant for now.
- Normal distribution: $L_t \sim N(\mu, \sigma^2)$, then $Z_t = \frac{L_t - \mu}{\sigma}$ is standard normal and the VaR for L_t is
$$VaR_{1-p} = \mu + z_{1-p}\sigma,$$
where z_{1-p} denotes the $(1-p)$th quantile of $N(0,1)$.
- In R, 0.95th quantile of N(0,1): qnorm(0.95)=1.645

Proof: $VaR_{1-p} = \mu + z_{1-p}\sigma$

$$\begin{aligned} p &= Pr(L_t \geq VaR_{1-p}) \\ &= Pr\left(\frac{L_t - \mu}{\sigma} \geq \frac{VaR_{1-p} - \mu}{\sigma}\right) \\ 1 - p &= Pr\left(Z_t < \frac{VaR_{1-p} - \mu}{\sigma}\right) \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{VaR_{1-p} - \mu}{\sigma} &= z_{1-p} \\ VaR_{1-p} &= \mu + z_{1-p}\sigma \end{aligned}$$

Example 6

The dollar change in the value of a portfolio position follows a normal distribution with mean \$1000 and standard deviation of \$500. Compute the 1% VaR for this position.

- $\Delta V_t(\ell) \sim N(1000, 500^2)$.
- Due to the long position, we have $L_t \sim N(-1000, 500^2)$.
- Let z_{1-p} denote the $(1-p)$-quantile of the standard normal distribution. $z_{0.99} = 2.3263$.

Example 6

The dollar change in the value of a portfolio position follows a normal distribution with mean \$1000 and standard deviation of \$500. Compute the 1% VaR for this position.

- The 1% VaR for this portfolio is

$$
\begin{aligned}
VaR_{0.99} &= \mu + z_{0.99}\sigma \\
&= -1000 + 2.3263 * 500 = 163\$.
\end{aligned}
$$

Example 7

- Suppose we construct a portfolio from two risky assets, with percentage returns r_1 and r_2, has a bivariate normal distribution with mean and covariance matrix

$$\mu = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 9 & 3 \\ 3 & 4 \end{bmatrix}$$

- The percentage returns are defined as, $r_{it} = 100 \times \frac{P_{i,t} - P_{i,t-1}}{P_{i,t}}$, where $P_{i,t}$ is the price of asset i at time t of which are bivariate normally distributed with mean and covariance matrix

Example 8

- Note that μ and Σ are the mean and volatility of the percentage returns. Our portfolio has weights $w = [0.5, 0.5]'$. We have invested 25000$ in this portfolio. Compute the 1% Value-at-Risk for this position.

Example 8

- Due to the long position, we have $x_1 = -r_1$ and $x_2 = -r_2$.
- $(x_1, x_2)'$ has a bivariate normal distribution with mean μ and covariance matrix Σ, where

$$\mu = \begin{bmatrix} -3 \\ -2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 9 & 3 \\ 3 & 4 \end{bmatrix}$$

- It follows that $E(x_1) = -3$, $E(x_2) = -2$, $var(x_1) = 9$, $var(x_2) = 4$ and $cov(x_1, x_2) = 3$.

Example 8

- Due to the normality assumption, we know the exact portfolio distribution $x_P = 0.5x_1 + 0.5x_2$, which is also normal.
- The mean of x_p, $E(x_p) = 0.5E(x_1) + 0.5E(x_2) = -2.5$.
- The variance of x_p,

$$\begin{aligned} var(x_p) &= 0.5^2 var(x_1) + 0.5^2 var(x_2) \\ &\quad + 2 \times 0.5 \times 0.5 \times cov(x_1, x_2) \\ &= 4.75 \end{aligned}$$

Example 8

- The 99%-quantile of the standard normal is $z_{0.99} = 2.3263$.
- The 99%-quantile of the normal with mean -2.5 and variance 4.75 is,

$$-2.5 + 2.3263\sqrt{4.75} = 2.5702,$$

Example 8

- The 1% VaR of the long position of 25000\$ is,

$$\frac{2.5702}{100} \times 25000\$ = 642.54\$,$$

- If the log returns are in percentages, we need to divide by 100 when calculating VaR.

Some closed-form solutions of VaR

- Assume $Y_t = \frac{L_t - \mu}{\sigma}$ is a Student-t distribution with ν degrees of freedom.
- VaR is calculated by,

$$VaR_{1-p} = \mu + t_{1-p,\nu}\sigma,$$

  where $t_{1-p,\nu}$ is the $(1-p)$th quantile of Student-t distribution with ν degrees of freedom.
- In R, 0.95th quantile of Student-t with 5 degrees of freedom: qt(0.95,5)=2.015.

Proof: $VaR_{1-p} = \mu + t_{1-p,\nu}\sigma$

$$
\begin{aligned}
p &= Pr(L_t \geq VaR_{1-p}) \\
&= Pr\left(\frac{L_t - \mu}{\sigma} \geq \frac{VaR_{1-p} - \mu}{\sigma}\right) \\
1 - p &= Pr\left(Y_t < \frac{VaR_{1-p} - \mu}{\sigma}\right)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{VaR_{1-p} - \mu}{\sigma} &= t_{1-p,\nu} \\
VaR_{1-p} &= \mu + t_{1-p,\nu}\sigma.
\end{aligned}
$$

Example 9

Suppose that the yearly return on a stock is Student's t distributed. We have the mean $\mu = 0.03$, standard deviation $\sigma = 0.116$ and degrees of freedom $\nu = 5$, i.e. $r_t \sim t(\mu = 0.03, \sigma = 0.116, \nu = 5)$. Note that μ and σ are defined for the log returns. An investor purchases $20,000$\$ worth of this stock.

Computer the 1-year horizon 5% VaR and 1% VaR for this position.

Example 9

- Due to the long position, $x_t = -r_t$. We have $x_t \sim$ Student-$t(-\mu, \sigma, \nu)$.
- The 95%-quantile of the Student's t distribution is,

$$-\mu + t_{1-p,\nu}\sigma = -0.03 + t_{0.95,5}0.116 = 0.2037$$

- The 5% VaR of this portfolio is

$$VaR_{0.95} = 0.2037 \times 20,000\$ = 4074\$,$$

Example 9

- The 99%-quantile of the Student's t distribution is,

$$-0.03 + t_{0.99,5}0.116 = 0.3603$$

- The 1% VaR of this portfolio is

$$VaR_{0.99} = 0.3603 \times 20,000\$ = 7206\$,$$

Student-t and Standardized Student-t

- Throughout the lecture note, we use *VaR* to denote the value-at-risk and use *var* to denote the variance.
- If X follows the Student-t distribution, the density function $f_\nu(x)$ is,

$$f_\nu(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

- $var(X) = \frac{\nu}{\nu-2}$ for $\nu > 2$.

Student-t and Standardized Student-t

- If X follows the Student-t distribution, $Y = X/\sqrt{\nu/(\nu-2)}$ follows the standardized Student-t distribution, the density function $f_\nu^*(y)$ is,

$$f_\nu^*(y) = f_\nu(\sqrt{\nu/(\nu-2)}y)\left|\frac{dy}{dx}\right|$$

$$f_\nu^*(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{y^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}$$

- $var(Y) = 1$ for $\nu > 2$.

Student-t and Standardized Student-t

If q is the pth quantile of a Student-t distribution with ν degrees of freedom,

$$p = \Pr(X \leq q) = \Pr\left[\frac{X}{\sqrt{\nu/(\nu-2)}} \leq \frac{q}{\sqrt{\nu/(\nu-2)}}\right]$$

$$= \Pr\left[Y \leq \frac{q}{\sqrt{\nu/(\nu-2)}}\right]$$

Then $q/\sqrt{\nu/(\nu-2)}$ is the pth quantile of a standardized Student-t distribution with ν degrees of freedom.

VaR for Student-t distribution

- Assume $Y_t = \frac{L_t - \mu}{\sigma}$ follows a Student-t distribution with ν degrees of freedom, where $\nu > 2$.
- Assume the $(1 - p)$th quantile of Standardized Student-t distribution, $t^*_{1-p,\nu}$, is given.
- In R, 0.95th quantile of Standardized-t with 5 degrees of freedom: qstd(0.95,nu=5)=1.56085.

VaR for Student-t distribution

- The $(1-p)$th quantile of Student-t distribution is $\sqrt{\nu/(\nu-2)}\,t^*_{1-p,\nu}$.
- VaR_{1-p} is calculated by,

$$VaR_{1-p} = \mu + t^*_{1-p,\nu}\sigma\sqrt{\nu/(\nu-2)},$$

Proof: $VaR_{1-p} = \mu + t^*_{1-p,\nu}\sigma\sqrt{\nu/(\nu-2)}$

$$
\begin{aligned}
p &= Pr(L_t \geq VaR_{1-p}) \\
&= Pr\left(\frac{L_t - \mu}{\sigma\sqrt{\nu/(\nu-2)}} \geq \frac{VaR_{1-p} - \mu}{\sigma\sqrt{\nu/(\nu-2)}}\right) \\
1 - p &= Pr\left(Y_t^* < \frac{VaR_{1-p} - \mu}{\sigma\sqrt{\nu/(\nu-2)}}\right)
\end{aligned}
$$

where $Y_t^* = (L_t - \mu)/(\sigma\sqrt{\nu/(\nu-2)})$ follows the Standardized Student-t distribution.

VaR for standardized Student-t dist.

Therefore,

$$\frac{VaR_{1-p} - \mu}{\sigma\sqrt{\nu/(\nu-2)}} = t^*_{1-p,\nu}$$

$$VaR_{1-p} = \mu + t^*_{1-p,\nu}\sigma\sqrt{\nu/(\nu-2)}$$

VaR for standardized Student-t dist.

- Assume $Y_t = \frac{L_t - \mu}{\sigma}$ follows a standardized Student-t distribution with ν degrees of freedom.
- VaR_{1-p} is calculated by,

$$VaR_{1-p} = \mu + t^*_{1-p,\nu}\sigma,$$

where $t^*_{1-p,\nu}$ is the $(1-p)$th quantile of Standardized Student-t distribution with ν degrees of freedom.

Drawbacks of VaR

- VaR is coherent for a normally distributed loss, however, VaR is not coherent in general.
- VaR does not describe the actual tail behavior of the loss random variable. It is not a perfect risk measure.

Coherence of VaR

- VaR is coherent for a normally distributed loss.
- Let *Var* denote the variance. Consider the subadditivity,

$$\begin{aligned}
\sigma^2_{x+y} &\equiv Var(X+Y) \\
&= Var(X) + Var(Y) + 2Cov(X,Y) \\
&= \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y \\
&\leq \sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y \\
&= (\sigma_x + \sigma_y)^2,
\end{aligned}$$

where $\rho = corr(X, Y)$. Therefore, $\sigma_{x+y} \leq \sigma_x + \sigma_y$.

Coherence of VaR

- This implies that $\mu_{X+Y} + z_{1-p}\sigma_{x+y} \leq E(X) + z_{1-p}\sigma_x + E(Y) + z_{1-p}\sigma_y$, where $\mu_{X+Y} \equiv E(X+Y)$, $\mu_X = E(X)$ and $\mu_Y = E(Y)$.
- $\mu_{X+Y} = \mu_X + \mu_Y$.
- This implies that VaR of $X+Y$ is less than or equal to the sum of VaR of X and VaR of Y.

Coherence of VaR

- VaR is not coherent in general.
- A simple counterexample; see also Example 3.13 of Klugman, Panjer and Willmot (2008).

Coherence of VaR

Example 10

Suppose the CDF $F_\ell(x)$ of a continuous loss random variable X satisfies the following probabilities:

$$F_\ell(80) = 0.9215, \;\; F_\ell(90) = 0.95, \;\; F_\ell(100) = 0.97.$$

For $p = 0.05$, the VaR of X is 90, because 90 is the 0.95th quantile of X. We denote this by $VaR^x_{0.95} = 90$.

Example 10

Now, define two loss random variables X_1 and X_2 by

$$X_1 = \begin{cases} X & \text{if } X \leq 100, \\ 0 & \text{if } X > 100 \end{cases}$$

and

$$X_2 = \begin{cases} 0 & \text{if } X \leq 100, \\ X & \text{if } X > 100 \end{cases}$$

These two loss variables are simply truncated versions of X and we have $X = X_1 + X_2$.

Example 10

Since the total probability must be 1, the CDF $F_\ell^1(x)$ of X_1 satisfies,

$$\begin{aligned} F_\ell^1(80) &= \Pr(X \leq 80 \cup X > 100) \\ &= 0.9215 + 0.03 = 0.9515 \\ F_\ell^1(90) &= \Pr(X \leq 90 \cup X > 100) \\ &= 0.95 + 0.03 = 0.98 \\ F_\ell^1(100) &= 1 \end{aligned}$$

The 0.95th quantile of X_1 is smaller than 80, where the superscript 1 is used to denote X_1.

Example 10

On the other hand,
$\Pr(X_2 \leq 0) = P(X \leq 100) = 0.97$. Therefore, the 0.95th quantile of X_2 is less than or equal to 0. We denote this by $VaR^2_{0.95} \leq 0$.

Taking the sum, we have $VaR^1_{0.95} + VaR^2_{0.95} < 80$.

In this particular instance, $X = X_1 + X_2$, yet $VaR^x_{0.95} = 90 > VaR^1_{0.95} + VaR^2_{0.95}$. Therefore, the subadditivity of VaR fails.

Expected Shortfall (ES)

- VaR does not describe the actual tail behavior of the loss random variable. It is not a perfect risk measure.
- Given a tail probability p, VaR is simply the $(1-p)$th quantile of the loss function.
- In practice, the actual loss, if it occurs, can be greater than VaR. In this sense, VaR may underestimate the actual loss.
- To have a better assessment of the potential loss, one can consider the expected value of the loss function if the VaR is exceeded.

Expected Shortfall (ES)

- Expected Shortfall (ES): also known as tail value at risk (TVaR) and conditional VaR (CVaR).
- Simply put, ES is the expected loss of a financial position after a catastrophic event. ES of a loss variable X is defined as

$$ES_{1-p} = E[X|X > VaR_{1-p}].$$

ES_{1-p} is the expected loss of X given that X exceeds its VaR_{1-p}.

Conditional Expectation given an event

Definition 11

The conditional expectation of a discrete random variable X given an event A is denoted as $E[X \mid A]$ and is defined by

$$\mathrm{E}[X \mid A] = \sum_x x \Pr[X = x \mid A]$$

Conditional Expectation given an event

It follows that

$$\mathrm{E}[X \mid A] = \sum_{x} x \Pr[X = x \mid A] = \sum_{x} x \frac{\Pr[X = x \text{ and } A]}{\Pr[A]}$$

Conditional Expectation: Example

Suppose that X and Y are discrete random variables with values in $\{1, 2\}$ s.t.

$$\begin{aligned} \Pr[X = 1, Y = 1] &= \tfrac{1}{2}, \quad \Pr[X = 1, Y = 2] = \frac{1}{10} \\ \Pr[X = 2, Y = 1] &= \tfrac{1}{10}, \quad \Pr[X = 2, Y = 2] = \frac{3}{10} \end{aligned}$$

Calculate $\mathrm{E}[X \mid Y = 1]$.

Conditional Expectation: Example

Suppose that X and Y are discrete random variables with values in $\{1, 2\}$ s.t.

$$\Pr[X = 1, Y = 1] = \tfrac{1}{2}, \quad \Pr[X = 1, Y = 2] = \tfrac{1}{10}$$
$$\Pr[X = 2, Y = 1] = \tfrac{1}{10}, \quad \Pr[X = 2, Y = 2] = \tfrac{3}{10}$$

We have $\Pr[Y = 1] = \Pr[X = 1, Y = 1] + \Pr[X = 2, Y = 1] = \frac{1}{2} + \frac{1}{10} = \frac{3}{5}$

Conditional Expectation: Example

Suppose that X and Y are discrete random variables with values in $\{1, 2\}$ s.t.

$$\Pr[X = 1, Y = 1] = \frac{1}{2}, \quad \Pr[X = 1, Y = 2] = \frac{1}{10}$$

$$\Pr[X = 2, Y = 1] = \frac{1}{10}, \quad \Pr[X = 2, Y = 2] = \frac{3}{10}$$

We have

$$\begin{aligned} \mathrm{E}[X \mid Y = 1] &= 1 \frac{\Pr[X = 1, Y = 1]}{\Pr[Y = 1]} + 2 \frac{\Pr[X = 2, Y = 1]}{\Pr[Y = 1]} \\ &= 1 \frac{1/2}{3/5} + 2 \frac{1/10}{3/5} = \frac{5}{6} + 2 \frac{1}{6} = \frac{7}{6} \end{aligned}$$

Conditional Expectation: Discrete random variables

Definition 12

The conditional expectation of a discrete random variable Y given that $X = x$ is defined as

$$\mathrm{E}[Y \mid X = x] = \sum_{y} y \Pr[Y = y \mid X = x]$$

Conditional Expectation: Continuous random variables

Definition 13

The conditional expectation of a continuous random variable Y given that $X = x$ is defined as

$$\mathrm{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y f_{Y|X=x}(y) dy = \frac{E[Y\mathbf{1}(X = x)]}{P(X = x)},$$

where $\mathbf{1}(\cdot)$ is an indicator function.

Expected Shortfall (ES)

ES of a loss variable X is defined as

$$\begin{aligned} ES_{1-p} &= E[X|X > VaR_{1-p}] \\ &= \frac{E[X\mathbf{1}(X > VaR_{1-p})]}{\Pr(X > VaR_{1-p})} \\ &= \frac{\int_{VaR_{1-p}}^{\infty} xf(x)dx}{\Pr(X > VaR_{1-p})}, \end{aligned}$$

Expected Shortfall (ES)

- Assume that X is continuous. Let $u = F(x)$ for $VaR \leq x \leq \infty$. Then, we have $du = f(x)dx$, $F(VaR) = 1 - p$, $F(\infty) = 1$, and $x = F^{-1}(u) = VaR_u$.
  The prior equation becomes
  $$ES_{1-p} = \frac{\int_{1-p}^{1} VaR_u du}{p}.$$
  Thus, ES can be seen to average all VaR_u for $1 - p \leq u \leq 1$.
- This averaging leads to coherence of ES.
- See VaRES.xlsx

For the two loss densities in Figure 3, their ES are different with the dash line corresponding to a higher value.

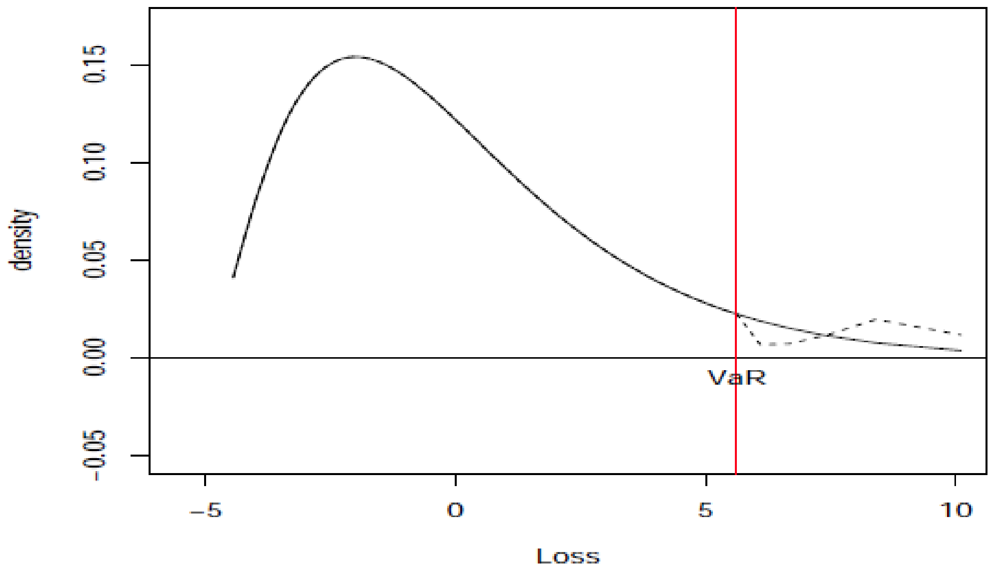


Figure 3: Density functions of two loss random variables that have the same VaR, but different loss implications.

ES for normally distributed losses

Closed-form solutions for ES are also available for some loss distributions.

- Normal distribution: $L_t \sim N(\mu, \sigma^2)$, then $Z_t = \frac{L_t - \mu}{\sigma}$ is standard normal and the ES for L_t is,

$$ES_{1-p} = \mu + \frac{\phi(z_{1-p})\sigma}{p},$$

where $\phi(z)$ is the pdf of $N(0,1)$ and z_{1-p} is the $(1-p)$th quantile function of $\phi(z)$.

Proof $ES_{1-p} = \mu + \phi(z_{1-p})\sigma/p$

$$
\begin{aligned}
ES_{1-p} &= E[L_t | L_t > VaR_{1-p}] \\
&= E[L_t | L_t > \mu + \sigma z_{1-p}] \\
&= E[L_t | \frac{L_t - \mu}{\sigma} > z_{1-p}] \\
&= \mu + \sigma E\left[\frac{L_t - \mu}{\sigma} \middle| \frac{L_t - \mu}{\sigma} > z_{1-p}\right] \\
&= \mu + \sigma E[Z_t | Z_t > z_{1-p}], \quad \text{with } Z_t \sim N(0,1),
\end{aligned}
$$

where we use $E[a|Y] = a$ and $E[aX|Y] = aE[X|Y]$.
Furthermore,

$$
\Pr(Z_t > z_{1-p}) = 1 - \Phi(z_{1-p}) = p
$$

Proof of ES under normal distribution

Note that

$$(-\phi(y))' = -\frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})(-y) = y\phi(y)$$

Hence,

$$\begin{aligned} E[Z_t | Z_t > z_{1-p}] &= \frac{E[Z_t \mathbf{1}(Z_t > z_{1-p})]}{P(Z_t > z_{1-p})} \\ &= \frac{\int_{z_{1-p}}^{\infty} y\phi(y)dy}{p} \\ &= \frac{-\phi(y)\Big|_{z_{1-p}}^{\infty}}{p} \\ &= \frac{\phi(z_{1-p})}{p} \end{aligned}$$

VaR v.s. ES

- Normal distribution: $L_t \sim N(\mu, \sigma^2)$, then
- The VaR is,

$$VaR_{1-p} = \mu + z_{1-p}\sigma,$$

- The expected shortfall is,

$$ES_{1-p} = \mu + \frac{\phi(z_{1-p})\sigma}{p},$$

where $\phi(z)$ is the pdf of $N(0,1)$ and z_{1-p} is the $(1-p)$th quantile function of $\phi(z)$.

Expected Shortfall (ES)

- Assume $Y_t = \frac{L_t - \mu}{\sigma}$ follows a Student-t distribution with ν degrees of freedom.
- The density function for L_t is given by,

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}\sigma}\left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}$$

- Let $t_{1-p,\nu}$ note the $(1-p)$th quantile of Student-t distribution with ν degrees of freedom.

Expected Shortfall (ES)

- ES_{1-p} is calculated as,

$$ES_{1-p} = \mu + \sigma \frac{f_\nu(t_{1-p,\nu})}{p} \left(\frac{\nu + t^2_{1-p,\nu}}{\nu - 1} \right),$$

where $f_\nu(x)$ denotes the pdf of Student-t distribution with ν degrees of freedom, where

$$f_\nu(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

Expected Shortfall (ES)

- Assume $Y_t = \frac{L_t - \mu}{\sigma}$ follows a Student-t distribution with ν degrees of freedom, where $\nu > 2$.
- Let $t^*_{1-p,\nu}$ denote the $(1-p)$th quantile of Standardized Student-t distribution with ν degrees of freedom.
- The $(1-p)$th quantile of Student-t distribution is $\sqrt{\nu/(\nu-2)}\,t^*_{1-p,\nu}$.

Expected Shortfall (ES)

- ES_{1-p} is calculated as,

$$ES_{1-p} = \mu + \sigma\sqrt{\nu/(\nu-2)}\frac{f_\nu^*(t^*_{1-p,\nu})}{p} \left(\frac{(\nu-2)+[t^*_{1-p,\nu}]^2}{\nu-1}\right),$$

where $f_\nu^*(x)$ is the pdf of a Standardized Student-t distribution with ν degrees of freedom.

Expected Shortfall (ES)

- Assume $Y_t = \frac{L_t - \mu}{\sigma}$ follows a Student-t distribution with ν degrees of freedom.
- We could either use,

$$ES_{1-p} = \mu + \sigma \frac{f_\nu(t_{1-p,\nu})}{p} \left(\frac{\nu + t_{1-p,\nu}^2}{\nu - 1} \right),$$

or use

$$ES_{1-p} = \mu + \sigma \sqrt{\nu/(\nu-2)} \frac{f_\nu^*(t_{1-p,\nu}^*)}{p} \left(\frac{(\nu-2) + [t_{1-p,\nu}^*]^2}{\nu - 1} \right).$$

Expected Shortfall (ES)

- Assume $Y_t = \frac{L_t - \mu}{\sigma}$ follows a standardized Student-t distribution with ν degrees of freedom, where $\nu > 2$.
- Let $f_\nu^*(x)$ and $t_{1-p,\nu}^*$ denote the pdf and the $(1-p)$th quantile of Standardized Student-t distribution with ν degrees of freedom.
- ES_{1-p} is calculated as,

$$ES_{1-p} = \mu + \sigma \frac{f_\nu^*(t_{1-p,\nu}^*)}{p} \left(\frac{(\nu - 2) + [t_{1-p,\nu}^*]^2}{\nu - 1} \right),$$

Calculation of VaR

Calculation of VaR involves several factors:

- Tail probability p: $p = 0.01$ for risk management and $p = 0.001$ in stress testing.
- The time horizon ℓ: 1 day or 10 days for market risk and 1 year or 5 years for credit risk.
- The CDF $F_\ell(x)$ or its quantiles of the loss random variable.
- The amount of the financial position or the mark-to-market value of the portfolio.

Methods for calculating financial risk

- RiskMetrics
- Econometric modeling
- Quantile and quantile regression
- Extreme value theory: traditional & Peaks over Thresholds
- Serially correlated data: extremal index

RiskMetrics

- J. P. Morgan developed the RiskMetrics methodology to VaR calculation.
- Denote the daily log return by r_t and the information set available at time $t-1$ by Ω_{t-1}.
- RiskMetrics assumes that $r_t|\Omega_{t-1} \sim N(\mu_t, \sigma_t^2)$, where $\mu_t = 0$ and

$$\begin{aligned} r_t &= e_t \\ e_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \lambda \sigma_{t-1}^2 + (1-\lambda) r_{t-1}^2, \quad 0 < \lambda < 1, \end{aligned}$$

where e_t is an IGARCH(1,1) process without drift.

$$\begin{aligned}\sigma_t^2 &= \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2 \\ &= \lambda\sigma_{t-1}^2 + (1-\lambda)e_{t-1}^2\end{aligned}$$

It's easy to show that

$$\sigma_{t-1}^2 = \lambda\sigma_{t-2}^2 + (1-\lambda)e_{t-2}^2$$

Replacing σ_{t-1} by its value in function of σ_{t-2} and r_{t-2}, we have

$$\sigma_t^2 = (1-\lambda)e_{t-1}^2 + \lambda(1-\lambda)e_{t-2}^2 + \lambda^2\sigma_{t-2}^2$$

Weights of previous returns

Replacing σ_{t-2} again and recursively continue the process until reach the initial position. Then we obtain the following formula:

$$\sigma_t^2 = (1-\lambda)(e_{t-1}^2 + \lambda e_{t-2}^2 + \lambda^2 e_{t-3}^2 + \ldots)$$

Since the decay factor $\lambda < 1$, the weight of past observations declines geometrically.

RiskMetrics approach assigns greater importance to recent data of returns instead of simple average in moving average method.

Comparison of RiskMetrics and GARCH

Recalling the GARCH(1,1) conditional variance dynamics,

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 e_{t-1}^2$$

The RiskMetrics equation can be seen as a special case of GARCH(1,1) model by setting $\alpha_0 = 0$, $\alpha_1 = 1 - \lambda$, $\beta_1 = \lambda$,

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) e_{t-1}^2$$

The value of λ is often in the interval $(0.9, 1)$ with a typical value of 0.94.

Multiple periods

- A nice property of RiskMetrics model is that the conditional distribution of a multiperiod return is easily available.
- Let $r_t[k]$ denote the k-period horizon log return from time $t+1$ to time $t+k$ (inclusive), where $r_t[k] = r_{t+1} + ... + r_{t+k-1} + r_{t+k}$.

Multiple periods

- According to RiskMetrics model, we have,

$$r_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$$
$$r_t[k]|\Omega_t \sim N(0, k\sigma_{t+1}^2),$$

  where Ω_t denotes the information set available at time t.
- Under the RiskMetrics model, the conditional variance of $r_t[k]$ is proportional to the time horizon k.

Proof: $r_t[k]|\Omega_t \sim N(0, k\sigma^2_{t+1})$.

Let $\sigma^2_t[k] = Var(r_t[k]|\Omega_t)$. Since

$$\begin{aligned} r_t[k] &= r_{t+1} + ... + r_{t+k-1} + r_{t+k} \\ &= e_{t+1} + ... + e_{t+k-1} + e_{t+k} \end{aligned}$$

We have,

$$\begin{aligned} \sigma^2_t[k] &= \sum_{i=1}^{k} Var(e_{t+i}|\Omega_t) \\ &= \sum_{i=1}^{k} E(e^2_{t+i}|\Omega_t) \end{aligned}$$

Proof: $r_t[k]|\Omega_t \sim N(0, k\sigma^2_{t+1})$.

Let $\mathcal{G}$ and $\mathcal{H}$ be two sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{G} \subseteq \mathcal{H}$. Then for the integrable random vector z,

$$E[E(z|\mathcal{H})|\mathcal{G}] = E[E(z|\mathcal{G})|\mathcal{H}] = E(z|\mathcal{G})$$

in particular, $E[E(z|\mathcal{G})] = E(z)$.

Note that for $i \geq 2$,

$$\begin{aligned} E(e^2_{t+i}|\Omega_t) &= E[E(e^2_{t+i}|\Omega_{t+i-1})|\Omega_t] \\ &= E[\sigma^2_{t+i} E(\varepsilon^2_{t+i}|\Omega_{t+i-1})|\Omega_t] \\ &= E[\sigma^2_{t+i}|\Omega_t] \end{aligned}$$

Proof: $r_t[k]|\Omega_t \sim N(0, k\sigma_{t+1}^2)$.

- Using $r_{t-1} = e_{t-1} = \sigma_{t-1}\varepsilon_{t-1}$, we can rewrite the volatility equation of the IGARCH(1,1) model as,

$$\sigma_t^2 = \sigma_{t-1}^2 + (1-\lambda)\sigma_{t-1}^2(\varepsilon_{t-1}^2 - 1) \quad \text{for all } t$$

In particular, we have, for $i = 2, ..., k$,

$$\sigma_{t+i}^2 = \sigma_{t+i-1}^2 + (1-\lambda)\sigma_{t+i-1}^2(\varepsilon_{t+i-1}^2 - 1)$$

Since $E(\varepsilon_{t+i-1}^2 - 1|\Omega_t) = 0$ for $i \geq 2$, the prior equation shows that

$$E(\sigma_{t+i}^2|\Omega_t) = E(\sigma_{t+i-1}^2|\Omega_t) \quad \text{for } i = 2, ..., k.$$

RiskMetrics model

By RiskMetrics model, we have, for $i \geq 1$,

$$\begin{aligned} E(e_{t+i}^2|\Omega_t) &= E[\sigma_{t+i}^2|\Omega_t] \\ &= E(\sigma_{t+1}^2|\Omega_t) \end{aligned}$$

Since $\sigma_{t+1}^2 = \alpha\sigma_t^2 + (1-\alpha)r_t^2$, we have $E(\sigma_{t+1}^2|\Omega_t) = \sigma_{t+1}^2$. Therefore,

$$\begin{aligned} \sigma_t^2[k] &= \sum_{i=1}^{k} E(e_{t+i}^2|\Omega_t) \\ &= k\sigma_{t+1}^2 \end{aligned}$$

RiskMetrics model

- By RiskMetrics model, $r_t[k]|\Omega_t \sim N(0, k\sigma^2_{t+1})$
- *VaR* is a prediction concerning possible loss of a portfolio in a given time horizon. It should be computed using the predictive distribution of future returns of the financial position.
- Because RiskMetrics assumes log returns are normally distributed with mean zero, the loss function is symmetric and VaR are the same for long and short financial positions.

RiskMetrics model

- Let z_{1-p} denote the $(1-p)$th quantile of $N(0,1)$. If the tail probability is set to $p\%$, for the next trading day, $VaR_{1-p} = z_{1-p}\sigma_{t+1}$. This is the upper $p\%$ quantile (or the (1-p)th percentile) of $N(0, \sigma^2_{t+1})$;
- For the next k trading days, $VaR[k]_{1-p} = \sqrt{k} z_{1-p}\sigma_{t+1}$, which is the upper $p\%$ quantile of $N(0, k\sigma^2_{t+1})$.

Example 14

The sample standard deviation of the continuously compounded daily return of the German mark/U.S. dollar exchange rate was about 0.53% in June 1997. Suppose that an investor was long in \$10 million worth of mark/dollar exchange rate contract.

Example 14

- The 5% VaR for a 1-day horizon (qnorm(0.95) $= 1.6449$) is

$$\$10,000,000 \times (1.6449 \times 0.0053) = \$87,179.7$$

- The 5% ES for a 1-day horizon is

$$\$10,000,000 \times (\phi(1.6449) \times 0.0053/0.05)$$
$$= \$109,315.4$$

Example 14

- The 1% VaR for a 1-day horizon (qnorm(0.99) $= 2.3263$) is

$$\$10,000,000 \times (2.3263 \times 0.0053) = \$123,293.9$$

- The 1% ES for a 1-day horizon is

$$\$10,000,000 \times (\phi(2.3263) \times 0.0053/0.01)$$
$$= \$141,272.1$$

Example 14

- The 5% VaR for a 10-day horizon is

  $\$10,000,000 \times (\sqrt{10} \times 1.6449 \times 0.0053) \approx \$275,686.4$

- The 5% ES for a 10-day horizon is

  $$\$10,000,000 \times (\sqrt{10} \times \phi(1.6449) \times 0.0053/0.05)$$
  $\approx \$345,685.8$

Example 14

- The 1% VaR for a 10-day horizon is

$$\$10,000,000\times(\sqrt{10}\times 2.3263\times 0.0053) = \$389,889.5$$

- The 1% ES for a 10-day horizon is

$$\$10,000,000 \times (\sqrt{10} \times \phi(2.3263) \times 0.0053/0.01)$$
$$=\$446,741.6$$

Example 15

Consider the daily IBM log returns of 9190 observations. Assume RiskMetrics is used and the fitted model is as follows,

$$r_t = e_t, \quad e_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = 0.9396\sigma_{t-1}^2 + 0.0604e_{t-1}^2,$$

where $\{\varepsilon_t\}$ is the standard Gaussian white noise series.

Assume $r_{9190} = -0.0128$ and $\sigma_{9190}^2 = 0.0003472$.

What are the 1% VaR and Expected Shortfall of \$10 millions?

Example 15

- The 1-step-ahead volatility forecast is,

$$\begin{aligned}\sigma^2_{9190}(1) &= 0.9396 \times 0.0003472 \\ &\quad + 0.0604 \times (-0.0128)^2 \\ &= 0.000336\end{aligned}$$

- The 99% quantile of the conditional distribution $r_{9191}|\Omega_{9190}$ is $2.326 \times \sqrt{0.000336} = 0.04265.$
- The 1% VaR of a long position of \$10 millions is,

$$\$10000000 \times 0.04265 = \$426500$$

Example 15

- The 1% ES of a long position of \$10 millions is,

$$\$10000000 \times \frac{\phi(2.326)}{0.01} \times \sqrt{0.000336} \approx \$488937.3$$

Drawbacks of RiskMetrics

- Assumed model is rejected by empirical data;
- The square-root of time rule fails if either of the model assumptions is rejected.

VaR for 2 assets

- For 2 assets, we have,

$$VaR_P = \sqrt{VaR_1^2 + VaR_2^2 + 2\rho_{12} VaR_1 VaR_2}$$

- It's similar to $Var(X+Y) = Var(X) + Var(Y) + 2\rho_{12} std(X) std(Y)$.

VaR for multiple positions

- The generalization of VaR to a position consisting of m instruments is straightforward as

$$VaR_P = \sqrt{\sum_{i=1}^{m} VaR_i^2 + 2\sum_{i<j}^{m} \rho_{ij} VaR_i VaR_j}$$

where ρ_{ij} is the cross-correlation coefficient between returns of the ith and jth instruments and VaR_i is the VaR of the ith instrument.

Proof of VaR_P for two assets

- Two assets with log returns r_{1t} and r_{2t}. The portfolio consists of w_1 and w_2 amounts of position invested in asset 1 and asset 2, respectively.
- Under RiskMetrics, we have

$$r_{1t}|\Omega_{t-1} \sim N(0, \sigma_{1t}^2),$$
$$r_{2t}|\Omega_{t-1} \sim N(0, \sigma_{2t}^2),$$

where

$$\sigma_{1t}^2 = \beta\sigma_{1,t-1}^2 + (1-\beta)r_{1,t-1}^2$$
$$\sigma_{2t}^2 = \beta\sigma_{2,t-1}^2 + (1-\beta)r_{2,t-1}^2$$

Proof of VaR_P for two assets

- For tail probability 0.05, we have VaR for two assets are $VaR_1 = 1.645 w_1 \sigma_{1t}$ and $VaR_2 = 1.645 w_2 \sigma_{2t}$, respectively.
- Let r_{P_t} be the log return of the portfolio. Then, we have

$$r_{P_t} \approx w r_{1t} + (1 - w) r_{2t}.$$

  where $w = w_1/(w_1 + w_2)$ and $1 - w = w_2/(w_1 + w_2)$. The prior approximation becomes equality for simple returns.
- $(w_1 + w_2)w = w_1$ and $(w_1 + w_2)(1 - w) = w_2$.

Proof of VaR_P for two assets

- Let $\rho_{12} = Cov(r_{1t}, r_{2t})/[Var(r_{1t})Var(r_{2t})]^{0.5}$.
- Under RiskMetrics, we have

$$r_{P_t}|\Omega_{t-1} \approx N(0, \sigma^2_{P_t}),$$

where

$$\begin{aligned} \sigma^2_{P_t} &= Var(r_{P_t}|\Omega_{t-1}) \\ &= w^2\sigma^2_{1t} + (1-w)^2\sigma^2_{2t} + 2w(1-w)\rho_{12}\sigma_{1t}\sigma_{2t} \end{aligned}$$

Proof of VaR_P for two assets

- For tail probability of 0.05, we have the VaR for the portfolio is $1.645(w_1 + w_2)\sigma_{P_t}$.
- The square of VaR for the portfolio with tail probability 0.05 is

$$
\begin{aligned}
(VaR_P)^2 &= (1.645)^2(w_1 + w_2)^2\sigma^2_{P_t} \\
&= (1.645)^2(w_1 + w_2)^2[w^2\sigma^2_{1t} \\
&\qquad + (1 - w)^2\sigma^2_{2t} + 2w(1 - w)\rho_{12}\sigma_{1t}\sigma_{2t}] \\
&= (1.645)^2[w_1^2\sigma^2_{1t} + w_2^2\sigma^2_{2t} \\
&\qquad + 2w_1 w_2\rho_{12}\sigma_{1t}\sigma_{2t}] \\
&= VaR_1^2 + VaR_2^2 + 2\rho_{12}VaR_1 VaR_2
\end{aligned}
$$

VaR for multiple positions

- The formula continues to hold for expected shortfall provided that the mean returns of the two assets are zero. [Which is assumed under RiskMetrics.]
- If the means are not zero, then some adjustments are needed for the portfolio.

Econometric Approach to *VaR* Calculation

- Let r_t be the log return of an asset. A general time series model for r_t can be written as,

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + e_t - \sum_{j=1}^{q} \theta_j e_{t-j},$$

$$e_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{u} \alpha_i e_{t-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t-j}^2$$

Econometric Approach to *VaR* Calculation

- The 1-step-ahead forecasts of the conditional mean and conditional variance of r_t are,

$$\hat{r}_t(1) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t+1-i} - \sum_{j=1}^{q} \theta_j e_{t+1-j},$$

$$\hat{\sigma}_t^2(1) = \alpha_0 + \sum_{i=1}^{u} \alpha_i e_{t+1-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t+1-j}^2$$

Econometric Approach to *VaR* Calculation

- If one assumes ε_t is Gaussian, then the conditional distribution of r_{t+1} given the information available at time t is $N[\hat{r}_t(1), \hat{\sigma}_t^2(1)]$.
- Then the $(1 - p)$th quantile is $\hat{r}_t(1) + z_{1-p}\hat{\sigma}_t(1)$, where z_{1-p} is the $(1 - p)$th quantile of a standard normal distribution.

Econometric Approach to *VaR* Calculation

- If one assumes ε_t is standardized Student-t distribution with ν degrees of freedom.
- Then $(1-p)$th quantile is $\hat{r}_t(1) + t^*_{1-p,\nu}\hat{\sigma}_t(1)$, where $t^*_{1-p,\nu}$ is the $(1-p)$th quantile of a standardized Student-t distribution with ν degrees of freedom.

Econometric Approach to *VaR* Calculation

- Assume ε_t is standardized Student-t distribution with ν degrees of freedom.
- The $(1-p)$th quantile of a Student-t distribution with ν degrees of freedom, $t_{1-p,\nu}$ is known.
- Then $(1-p)$th quantile used to calculate the 1-period horizon VaR at time index t is

$$\hat{r}_t(1) + \frac{t_{1-p,\nu}}{\sqrt{\nu/(\nu-2)}}\hat{\sigma}_t(1)$$

Example 16

- Consider the daily IBM log returns from July 3, 1962 to December 31, 1998, for 9190 observations.
- We use two volatility models to calculate VaR of 1-day horizon at $t = 9190$ for a long position of \$10 million.
- Because the position is long, we use $x_t = -r_t$, where r_t is the log return calculated from IBM market price.

Example 16

- CASE 1: Assume ε_t is standard normal. The fitted model is

$$
\begin{aligned}
x_t &= -0.00066 - 0.0247x_{t-2} + e_t \\
e_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= 0.00000389 + 0.0799e_{t-1}^2 + 0.9073\sigma_{t-1}^2
\end{aligned}
$$

- From the data, we have $x_{9189} = 0.00201$, $x_{9190} = 0.0128$, $e_{9190}^2 = 0.0001661$ and $\sigma_{9190}^2 = 0.00033455$.

Example 16

- The AR(2)-GARCH(1,1) model produces 1-step-ahead forecasts of the conditional mean and conditional variance of r_t as,

$$\begin{aligned}
\hat{x}_{9190}(1) &= E[x_{9191}|\Omega_{9190}] \\
&= -0.00066 - 0.0247x_{9189} \\
&= -0.00071 \\
\hat{\sigma}^2_{9190}(1) &= E[\sigma^2_{9191}|\Omega_{9190}] \\
&= 0.00000389 + 0.0799e^2_{9190} + 0.9073\sigma^2_{9190} \\
&= 0.0003211
\end{aligned}$$

Example 16

- The 95%-quantile of the standard normal is $z_{0.95} = 1.6449$.
- The 95%-quantile of portfolio return distribution is,

$$-0.00071 + 1.6449 \times \sqrt{0.0003211} = 0.02877$$

- The VaR for a long position of \$10 million with shortfall probability of 5% is

$$0.02877 \times 10000000\$ = 287700\$,$$

- This result shows that with probability 95%, the potential loss of holding that position next day will not exceed 287700\$ assuming that the AR(2)-GARCH(1,1) model holds.

Example 16

- The 99%-quantile of the standard normal is $z_{0.99} = 2.3262$.
- The 99%-quantile of portfolio return distribution is,

$$-0.00071 + 2.3262 \times \sqrt{0.0003211} = 0.0409738$$

- The VaR for a long position of \$10 million with shortfall probability of 1% is this portfolio is

$$0.0409738 \times 10000000\$ = 409738\$,$$

- This result shows that with probability 99%, the potential loss of holding that position next day will not exceed 409738\$ assuming that the AR(2)-GARCH(1,1) model holds.

Example 16

- CASE 2: Assume ε_t is standardized-t distribution with 5 degrees of freedom. The fitted model is

$$
\begin{aligned}
x_t &= -0.0003 - 0.0335 x_{t-2} + e_t \\
e_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= 0.000003 + 0.0559 e_{t-1}^2 + 0.9350 \sigma_{t-1}^2
\end{aligned}
$$

- Given $x_{9189} = 0.00201$, $x_{9190} = 0.0128$, $e_{9190}^2 = 0.0001661$ and $\sigma_{9190}^2 = 0.000349$.
- We calculate VaR of 1-day horizon at $t = 9190$ for a long position of \$10 million.

Example 16

- The AR(2)-GARCH(1,1) model produces 1-step-ahead forecast as,

$$\begin{aligned}
\hat{x}_{9190}(1) &= E[x_{9191}|\Omega_{9190}] = -0.0003 - 0.0335x_{9189} \\
&= -0.000367 \\
\hat{\sigma}^2_{9190}(1) &= 0.0003386
\end{aligned}$$

Example 16

- The 95%-quantile of a Student-t distribution with 5 degrees of freedom is 2.015.
- The 95%-quantile of a Standardized Student-t distribution is $2.015/\sqrt{5/3} = 1.5608$.
- The 95%-quantile of portfolio return distribution is,

$$\begin{aligned} t_{0.95}(5) &= -0.000367 + 1.5608 \times \sqrt{0.0003386} \\ &= 0.028354 \end{aligned}$$

Example 16

- Then, the VaR with shortfall probability of 5% of this portfolio is

$$0.028354 \times 10000000 = 283540\$,$$

- This result shows that with probability 95%, the potential loss of holding that position next day will not exceed 283540$ assuming that the AR(2)-GARCH(1,1) model holds.

Example 16

- The 99%-quantile of the Student-t distribution with 5 degrees of freedom is 3.3649.
- The 99%-quantile of a Standardized Student-t distribution is $3.3649/\sqrt{5/3} = 2.606$.
- The 99%-quantile of portfolio return distribution is,

$$\begin{aligned} t_{0.99}(5) &= -0.000367 + 2.606 \times \sqrt{0.0003386} \\ &= 0.0475943 \end{aligned}$$

Example 16

- Then, the VaR with shortfall probability of 1% of this portfolio is

$$0.0475943 \times 10000000\$ = 475943\$,$$

- This result shows that with probability 99%, the potential loss of holding that position next day will not exceed 475943$ assuming that the AR(2)-GARCH(1,1) model holds.

Multiple periods

- The return r_t follows the time series model,

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + e_t - \sum_{j=1}^{q} \theta_j e_{t-j},$$

$$e_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{u} \alpha_i e_{t-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t-j}^2$$

Multiple periods

- The variable of interest is the k-period log return at the forecast origin h, i.e. $r_h[k] = r_{h+1} + ... + r_{h+k}$.
- The conditional mean $E(r_h[k]|\Omega_h)$ can be obtained as follows,

$$\hat{r}_h[k] = \hat{r}_h(1) + ... + \hat{r}_h(k),$$

where $\hat{r}_h(\ell)$ is the ℓ-step-ahead forecast of the return at the forecast origin h.

Multiple periods

- Rewrite the ARMA model as,

$$\Phi(L) r_t = \phi_0 + \Theta(L) e_t,$$

where $\Phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$ and
$\Theta(L) = 1 - \sum_{i=1}^{q} \theta_i L^i$.

- Define

$$\frac{\Theta(L)}{\Phi(L)} = 1 + \psi_1 L + \psi_2 L^2 + ... \equiv \Psi(L),$$

Multiple periods

- Using the MA reprsentation

$$r_t = \mu + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \ldots,$$

where ψ_i is the ℓ-step-ahead forecast of the return at the forecast origin h.

- The ℓ-step-ahead forecast error at the forecast origin h is,

$$\begin{aligned} e_h(\ell) &= r_{h+\ell} - \hat{r}_h(\ell) \\ &= e_{h+\ell} + \psi_1 e_{h+\ell-1} + \ldots + \psi_{\ell-1} e_{h+1}, \end{aligned}$$

Multiple periods

- The forecast error of the expected k-period return $\hat{r}_h[k]$ is,

$$
\begin{aligned}
e_h[k] &= e_h(1) + e_h(2) + \dots + e_h(k) \\
&= e_{h+1} + (e_{h+2} + \psi_1 e_{h+1}) + \dots + \sum_{i=0}^{k-1} \psi_i e_{h+k-i} \\
&= e_{h+k} + (1 + \psi_1) e_{h+k-1} + \dots + \left(\sum_{i=0}^{k-1} \psi_i \right) e_{h+1},
\end{aligned}
$$

where $\psi_0 = 1$.

Multiple periods

- The volatility forecast of the k-period return at the forecast origin h is $Var(e_h[k]|\Omega_h)$,

$$
\begin{aligned}
& Var(e_h[k]|\Omega_h) \\
& = Var(e_{h+k}|\Omega_h) + (1+\psi_1)^2 Var(e_{h+k-1}|\Omega_h) + \ldots \\
& \qquad + \left(\sum_{i=0}^{k-1} \psi_i\right)^2 Var(e_{h+1}|\Omega_h), \\
& = \sigma_h^2(k) + (1+\psi_1)^2 \sigma_h^2(k-1) + \ldots + \left(\sum_{i=0}^{k-1} \psi_i\right)^2 \sigma_h^2(1),
\end{aligned}
$$

where $\psi_0 = 1$.

Special case

- The return r_t follows the special time series model,

$$r_t = \mu + e_t, \qquad e_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

  In this case $\psi_i = 0$ for all $i > 0$.
- $\hat{r}_h[k] = E(r_h[k] | \Omega_h) = k\mu$.
- The associated error is,

$$\begin{aligned} e_h[k] &= r_h[k] - \hat{r}_h[k] \\ &= e_{h+k} + e_{h+k-1} + \ldots + e_{h+1} \end{aligned}$$

Special case

- The return r_t follows the special time series model,

$$r_t = \mu + e_t, \qquad e_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

  In this case $\psi_i = 0$ for all $i > 0$.
- The point forecast of the k-period return at the forecast origin h is $\hat{r}_h[k] = k\mu$.
- The associated forecast error is,

$$e_h[k] = e_{h+k} + e_{h+k-1} + \ldots + e_{h+1}$$

Special case

- The volatility forecast for the k-period return at the forecast origin h is,

$$Var(e_h[k]|\Omega_h) = \sum_{\ell=1}^{k} \sigma_h^2(\ell)$$

Special case

- Using the forecasting method of GARCH(1,1) models, we have

$$\sigma_h^2(1) = \alpha_0 + \alpha_1 e_h^2 + \beta_1 \sigma_h^2,$$
$$\sigma_h^2(\ell) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(\ell - 1)$$

- We obtain that for the case of $\psi_i = 0$ for $i > 0$,

$$Var(e_h[k] \| \Omega_h) = \frac{\alpha_0}{1-\phi}\left(k - \frac{1-\phi^k}{1-\phi}\right) + \frac{1-\phi^k}{1-\phi}\sigma_h^2(1)$$

where $\phi = \alpha_1 + \beta_1 < 1$.

Special case

- If ε_t is Gaussian, then the conditional distribution of $r_h[k]$ given Ω_h is normal with mean $k\mu$ and variance $Var(e_h[k]|\Omega_h)$.

What are extreme values?

- Extreme value is either a very small or a very large value in a probability distribution.
- In the financial risk management parlance, an extreme value is one that has a low probability of occurrence but potentially disastrous (catastrophic) effects if it does happen.
- The occurrence of extreme events is very rare but can prove very costly in financial terms.

What are extreme values?

- The main challenge posed by extreme values is that there are only a few observations from which a credible, reliable analytical model can be built.
- In fact, there are some extreme values that have never occurred, but that does not necessarily imply there's no chance of occurrence in the future.

What are extreme values?

- Researchers tackle the challenge by assuming a certain distribution.
- Choosing a distribution arbitrarily is ill-suited to handle extremes because the distribution will tend to accommodate the more central observations because there are so many of them, rather than the extreme observations, which are much sparser.

Extreme Value Theory (EVT)

- Extreme value theory (EVT) is a branch of applied statistics developed to address study and predict the probabilities of extreme outcomes.
- It differs from "central tendency" statistics where we seek to dissect probabilities of relatively more common events, making use of the central limit theorem.
- Extreme value theory is not governed by the central limit theorem because it deals with the tail region of the relevant distribution.

Convergence of sums

- Let $\{X_i\}_{i \in N}$ be $i.i.d.$ with mean μ and variance σ^2.
- Assume $E(X_1^2) < \infty$, that is, second moment exists.
- Let $S_n = \sum_{i=1}^n X_i$ and $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.
- As $n \to \infty$, $\bar{X}_n \overset{a.s.}{\to} \mu$ by the Strong law of large Numbers (SLLN) so $(\bar{X}_n - \mu)/\sigma \overset{a.s.}{\to} 0$.

Convergence of sums

By the central limit theorem,

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1)$$

$$\lim_{n\to\infty} P\left(\frac{S_n - d_n}{c_n} \leq x\right) = \Phi(x),$$

where the sequences $c_n = \sqrt{n}\sigma$ and $d_n = n\mu$ give normalization and where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2}dz$.

Modeling Maxima and Worst Cases

- Let's consider the analysis of daily negative returns of S&P 500 index.
- What is the probability that next year's annual maximum negative return exceeds all previous negative returns? In other words, what is the probability that next year's maximum negative return is a new record?

Modeling Maxima and Worst Cases

- What is the 20-year return level of the negative returns? That is, what is the negative return which, on average, should only be exceeded in one year every twenty years?
- To answer these questions, the distribution of extreme negative returns on the S&P 500 index is required.

Convergence of maxima

- The risk management is concerned with maximal losses (worst-case losses).
- Let $X_1, \ldots, X_n$ be i.i.d. random variables (can be relaxed to a strictly stationary time series) representing risks or losses with an unknown CDF $F(x) = \Pr\{X_i \leq x\}$ and F is continuous.
- Then the block maximum is given by

$$M_n = \max\{X_1, \ldots, X_n\}$$

M_n is the worst-case loss in a sample of n losses.

Convergence of maxima

- The block maximum is given by

$$M_n = \max\{X_1, ..., X_n\}$$

- One can show that, for $n \to \infty$, $M_n \stackrel{a.s.}{\to} x_F$ (similar as in the SLLN) where $x_F := \sup\{x \in R : F(x) < 1\} = F^{\leftarrow}(1) \leq \infty$ denotes the right endpoint of F (similar to the SLLN).

Convergence of maxima

- The block maximum is given by

$$M_n = \max\{X_1, ..., X_n\}$$

- From the $i.i.d.$ assumption, the CDF of M_n is,

$$\Pr(M_n \le x) = \Pr(X_1 \le x, ..., X_n \le x)$$
$$= \prod_{i=1}^{n} F(x) = [F(x)]^n$$

- The empirical distribution function is often a very poor estimator of $F^n(x)$.
- $[F(x)]^n \to 0$ if $x < u$ and $[F(x)]^n \to 1$ if $x \ge u$ as $n \to \infty$.

Fisher-Tippet Theorem

- Define the standardized maximum value

$$Z_n = \frac{M_n - \mu_n}{\sigma_n}$$

$$\{\mu_n\} : \text{location series}$$

$$\{\sigma_n\} > 0 : \text{series of scaling factors}$$

Fisher-Tippet Theorem

- Fisher-Tippet Theorem: If the standardized maximum Z_n converges to some nondegenerate distribution function, it must be a generalized extreme value (GEV) distribution of the form

$$H_\xi(z) = \begin{cases} \exp\{-(1+\xi z)^{-1/\xi}\} & \text{if } \xi \neq 0 \\ \exp\{-\exp(-z)\} & \text{if } \xi = 0. \end{cases}$$

  where $-\infty < z < \infty$ for $\xi = 0$, and $z < -1/\xi$ for $\xi < 0$ and $z > -1/\xi$ for $\xi > 0$.
- We can combine $z < -1/\xi$ for $\xi < 0$ and $z > -1/\xi$ for $\xi > 0$ and write as $1 + \xi z > 0$.

Fisher-Tippet Theorem

- We can write
$$\Pr\left(\frac{M_n - \mu_n}{\sigma_n} \leq x\right) \xrightarrow{d} H_\xi(x)$$
- The CDF F of the underlying data is in the domain of attraction of H_ξ.
- The Fisher-Tippet Theorem is the analog of the Central Limit Theorem for extreme values.
- ξ is a shape parameter and determines the moments and tail behavior of H_ξ. The large ξ, the heavier tailed H_ξ.
- The parameter $\alpha = 1/\xi$ is called the tail index if $\xi > 0$.

GEV Types

- Result: The tail behavior of the distribution F of the underlying data determines the shape parameter ξ of the GEV distribution
- The GEV distribution encompasses the three types of limiting distribution.

GEV Types: Type I

- If the tail of F declines exponentially, then H_ξ is of the Gumbel type and $\xi = 0$.
- The CDF is of the form,

$$H_\xi(z) = \exp\{-\exp(-z)\}, \quad -\infty < z < \infty$$

- Distributions in the domain of attraction of the Gumbel type are thin tailed distributions such as the normal, log-normal, exponential, and gamma. For these distributions, all moments usually exist.

GEV Types: Type II

- If the tail of F declines by a power function, i.e.

$$1 - F(x) = c \cdot x^{-1/\xi} = x^{-\alpha}$$

  for some constant c, then H_ξ is of the Fréchet type and $\xi > 0$.
- The CDF is,

$$H_\xi(z) = \begin{cases} \exp\{-(1+\xi z)^{-1/\xi}\} & \text{if } z > -1/\xi \\ 0, & \text{otherwise} \end{cases}$$

GEV Types: Type II

- Distributions in the domain of attraction of the Fréchet type include fat tailed distributions like the Pareto, Cauchy, Student-t, alpha-stable with characteristic exponent in (0, 2), as well as various mixture models. Not all moments are finite for these distributions: $E[X^k] = \infty$ for $k \geq \alpha = 1/\xi$. This type is most relevant for financial data.

GEV Types: Type III

- If the tail of F is finite then H_ξ is of the Weibull type and $\xi < 0$.
- The CDF is,

$$H_\xi(z) = \begin{cases} \exp\{-(1+\xi z)^{-1/\xi}\} & \text{if } z < -1/\xi \\ 1, & \text{otherwise} \end{cases}$$

- Distributions in the domain of attraction of the Weibull type include distributions with bounded support such as the uniform and beta distributions. All moments exist for these distributions

GEV Types

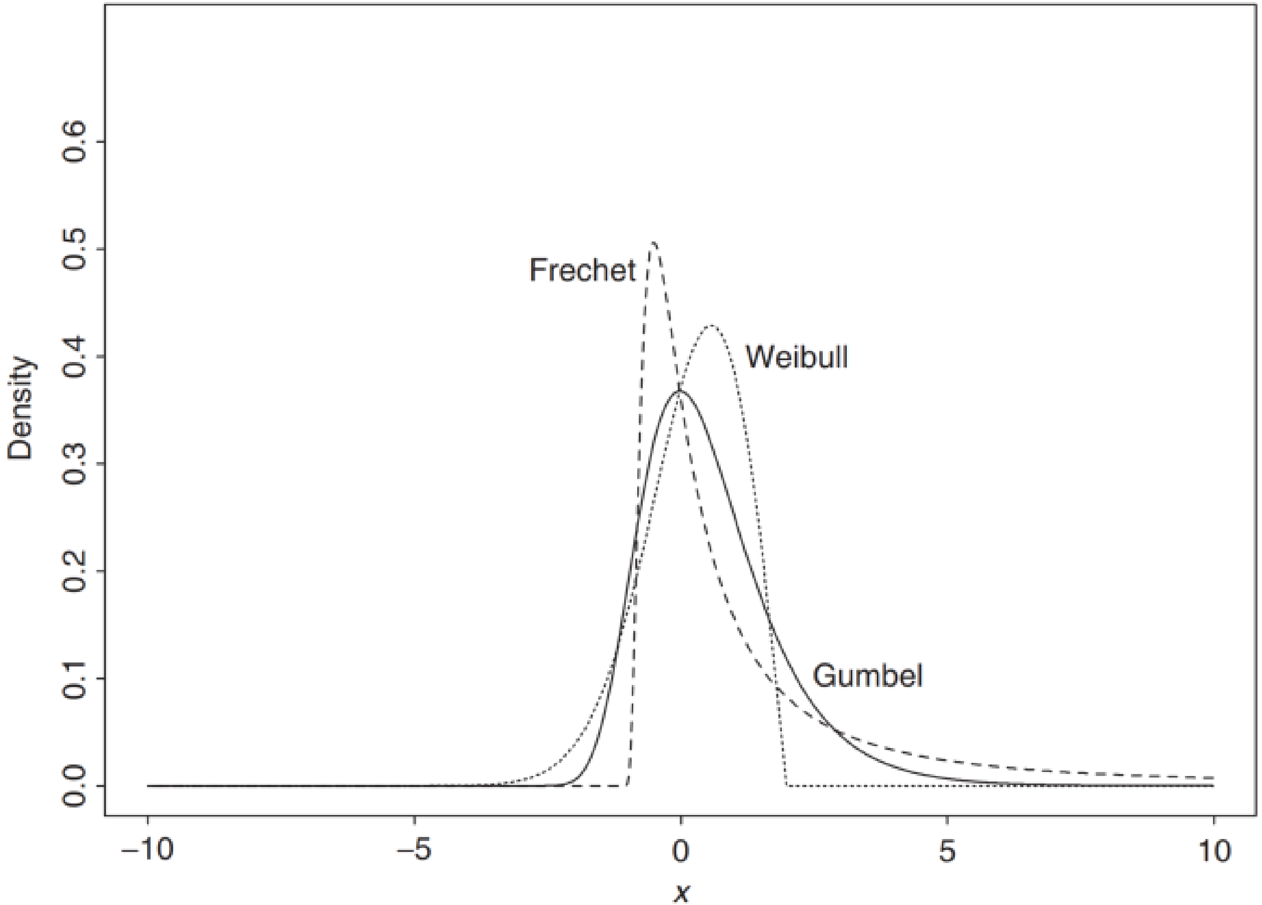


Figure 7.2 Probability density functions of extreme value distributions for maximum. Solid line is for Gumbel distribution, dotted line is for Weibull distribution with $\xi = -0.5$, and dashed line is for Fréchet distribution with $\xi = 0.9$.

Generalized extreme value (GEV) density function

- The generalized extreme value (GEV) density function is of the form

$$h_\xi(z) = \begin{cases} (1+\xi z)^{-1/\xi-1}\exp\{-(1+\xi z)^{-1/\xi}\} & \text{if } \xi \neq 0 \\ \exp\{-z-\exp(-z)\} & \text{if } \xi = 0. \end{cases}$$

  where $-\infty < z < \infty$ for $\xi = 0$, and $z < -1/\xi$ for $\xi < 0$ and $z > -1/\xi$ for $\xi > 0$.
- We can combine $z < -1/\xi$ for $\xi < 0$ and $z > -1/\xi$ for $\xi > 0$ and write as $1 + \xi z > 0$.

Unstandardized Distributions

- For location and scale parameters μ and $\sigma > 0$ we have

$$H_\xi(z) = H_\xi\left(\frac{x-\mu}{\sigma}\right) = H_{\xi,\mu,\sigma}(x)$$

- For large enough n,

$$\Pr\{Z_n < z\} = \Pr\left\{\frac{M_n - \mu_n}{\sigma_n} < z\right\} \sim H_\xi(z)$$

Unstandardized Distributions

- Letting $x = \sigma_n z + \mu_n$ then

$$\Pr\{M_n < x\} = H_\xi \left\{ \frac{x - \mu_n}{\sigma_n} \right\} \sim H_{\xi,\mu_n,\sigma_n}(x)$$

  where μ_n,σ_n and ξ are location, scale and shape parameters, respectively.
- This result is used in practice to make inferences about the maximum loss M_n.

The block maxima method (BMM)

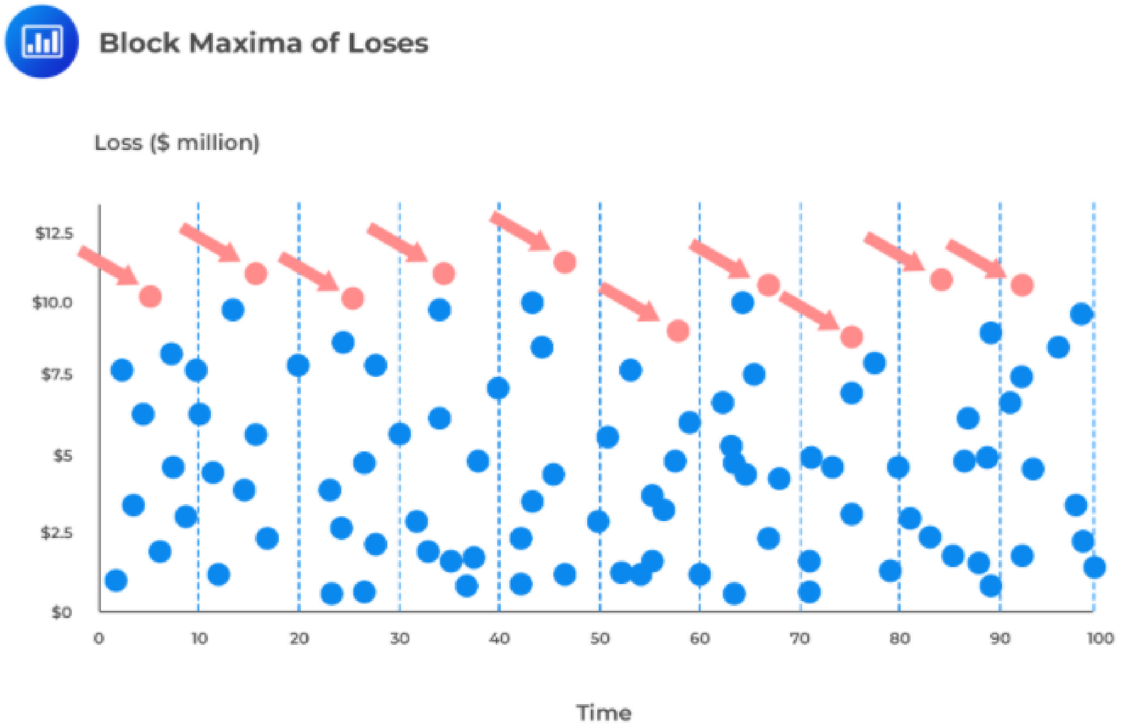
- Let $X_1, ..., X_T$ be $i.i.d.$ losses with unknown CDF F and let M_T denote the sample maximum. Divide the sample into g non-overlapping blocks of essentially equal size $n = T/g$

$$[X_1, ..., X_n | X_{n+1}, ..., X_{2n} | ... | X_{(g-1)n+1}, ..., X_{gn}]$$

$M_n^{(j)}$ is the maximum value of X_i in block $j = 1, ..., g$.

The block maxima method (BMM)

In risk management, we can segregate losses recorded over equal time intervals, say, 10 day periods, and record the largest loss in each interval.



The block maxima method (BMM)

- If we record the largest loss in each interval, we will end up with block maxima $(M_n^{(1)}, \ldots, M_n^{(g)})$.
- If n is sufficiently large, $Z_n = (M_n - \mu_n)/\sigma_n$ should follow an extreme value distribution.
- The collection of block maxima $(M_n^{(1)}, \ldots, M_n^{(g)})$ can be regarded as a sample of g observations from that extreme value distribution.
- The collection of block maxima $(M_n^{(1)}, \ldots, M_n^{(g)})$ is the data we use to estimate the unknown parameters of extreme value distribution.

Maximum Likelihood Estimation

- Construct likelihood for ξ, σ_n and μ_n from

$$\{M_n^{(1)}, \dots, M_n^{(g)}]$$

- Assumed that the block size n is sufficiently large so that the Fisher-Tippet Theorem holds.

Generalized extreme value (GEV) distribution

- The GEV distribution function is given by,

$$H_{\xi,\mu_n,\sigma_n}(x) = \begin{cases} \exp\left[-(1+\xi \times \frac{x-\mu_n}{\sigma_n})^{-1/\xi}\right] & \text{if } \xi \neq 0, \\ \exp\left[-\exp(-\frac{x-\mu_n}{\sigma_n})\right] & \text{if } \xi = 0. \end{cases}$$

where $-\infty < \frac{x-\mu_n}{\sigma_n} < \infty$ for $\xi = 0$, and $\frac{x-\mu_n}{\sigma_n} < -1/\xi$ for $\xi < 0$ and $\frac{x-\mu_n}{\sigma_n} > -1/\xi$ for $\xi > 0$.

Generalized extreme value (GEV) density

- The GEV density function is given by,

$$h_{\xi,\mu_n,\sigma_n}(x) = \begin{cases} \frac{1}{\sigma_n}(1+\frac{\xi(x-\mu_n)}{\sigma_n})^{-1/\xi-1}\exp\{-(1+\frac{\xi(x-\mu_n)}{\sigma_n})^{-1/\xi}\} & \text{if } \xi \neq 0 \\ \frac{1}{\sigma_n}\exp(-\frac{x-\mu_n}{\sigma_n} - \exp(-\frac{x-\mu_n}{\sigma_n})) & \text{if } \xi = 0. \end{cases}$$

where $-\infty < \frac{x-\mu_n}{\sigma_n} < \infty$ for $\xi = 0$, and $\frac{x-\mu_n}{\sigma_n} < -1/\xi$ for $\xi < 0$ and $\frac{x-\mu_n}{\sigma_n} > -1/\xi$ for $\xi > 0$.

Maximum Likelihood Estimation

- The log-likelihood function for $\xi \neq 0$ is,

$$l(\mu_n, \sigma_n, \xi) = -g \ln(\sigma_n) - (1 + 1/\xi) \sum_{i=1}^{g} \ln \left[1 + \xi \left(\frac{M_n^{(i)} - \mu_n}{\sigma_n} \right) \right] - \sum_{i=1}^{g} \left[1 + \xi \left(\frac{M_n^{(i)} - \mu_n}{\sigma_n} \right) \right]^{-1/\xi}$$

and is maximized imposing the constraint

$$1 + \xi \left(\frac{M_n^{(i)} - \mu_n}{\sigma_n} \right) > 0$$

Maximum Likelihood Estimation

- For $\xi > -0.5$ the MLEs for μ_n, σ_n and ξ are consistent and asymptotically normally distributed with asymptotic variance given by the inverse of the observed information matrix
- The bias of the MLE is reduced by increasing the block size n, and the variance of the MLE is reduced by increasing the number of blocks g.
- It's a bias-variance tradeoff!

Example: S&P 500 negative returns

The maximum likelihood estimates of ξ, μ and σ based on annual block maxima are

$$\hat{\xi} = 0.334, \widehat{SE}(\hat{\xi}) = 0.208$$
$$\hat{\mu} = 1.975, \widehat{SE}(\hat{\mu}) = 0.151$$
$$\hat{\sigma} = 0.672, \widehat{SE}(\hat{\sigma}) = 0.131$$

Given $\max\left(M_{260}^{(1)}, \ldots, M_{260}^{(28)}\right) = 6.68.$

Example: S&P 500 negative returns

What is the probability that next year's annual maximum negative return exceeds all previous negative returns?

$$
\begin{aligned}
&\Pr\left(M_{260}^{(29)} > \max\left(M_{260}^{(1)}, \ldots, M_{260}^{(28)}\right)\right) \\
=&\Pr\left(M_{260}^{(29)} > 6.68\right) \\
=&1 - H_{\hat{\xi},\hat{\mu},\hat{\sigma}}(6.68) = 0.0267
\end{aligned}
$$

VaR Calculation based on BMM

- Assume that there are T observations of an asset return available in the sample period.
- We partition the sample period into g nonoverlapping subperiods of length n such that $T = ng$.
- If $T = n \times g + m$ with $1 \leq m < n$, then we delete the first m observations from the sample.
- Apply the extreme value theory to obtain estimates of the location, scale, and shape parameters μ_n, σ_n, and ξ for the block maxima $M_n^{(j)}$, $j = 1, ..., g$.

VaR Calculation based on BMM

- Let p^* be a small upper tail probability that indicates the potential loss.
- r_n^* be the $(1 - p^*)$th quantile of the block maxima under the limiting generalized extreme value distribution.

$$\Pr(M_n^{(j)} \le r_n^*) = 1 - p^*$$

VaR Calculation based on BMM

- We have

$$1-p^* = \begin{cases} \exp\left\{-\left[1+\frac{\xi(r_n^*-\mu_n)}{\sigma_n}\right]^{-1/\xi}\right\} & \text{if } \xi \neq 0 \\ \exp\left[-\exp\left(-\frac{r_n^*-\mu_n}{\sigma_n}\right)\right] & \text{if } \xi = 0 \end{cases}$$

where it is understood that $1+\xi\left(r_n^*-\mu_n\right)/\sigma_n > 0$ for $\xi \neq 0$.

VaR Calculation based on BMM

- Rewriting this equation as

$$\ln(1-p^*) = \begin{cases} -\left[1+\frac{\xi(r_n^*-\mu_n)}{\sigma_n}\right]^{-1/\xi} & \text{if } \xi \neq 0 \\ -\exp\left(-\frac{r_n^*-\mu_n}{\sigma_n}\right) & \text{if } \xi = 0 \end{cases}$$

we obtain the quantile as

$$r_n^* = \begin{cases} \mu_n - \frac{\sigma_n}{\xi}\left\{1-\left[-\ln(1-p^*)\right]^{-\xi}\right\} & \text{if } \xi \neq 0 \\ \mu_n - \sigma_n \ln\left[-\ln(1-p^*)\right] & \text{if } \xi = 0. \end{cases}$$

In financial applications, the case of $\xi \neq 0$ is of major interest.

VaR Calculation based on BMM

- For a given upper tail probability p^*, the quantile r_n^* is the VaR based on the extreme value theory for the subperiod maximum.
- The next step is to make explicit the relationship between block maxima and the observed return r_t series.

VaR Calculation based on BMM

- Because most asset returns are either serially uncorrelated or have weak serial correlations, we have

$$1 - p^* = P\left(M_n^{(j)} \leq r_n^*\right) = \left[P\left(r_t \leq r_n^*\right)\right]^n$$

- Assume for a specified small upper probability p, the $(1-p)$th quantile of r_t is r_n^*, $P\left(r_t \leq r_n^*\right) = 1 - p$.

VaR Calculation based on BMM

- For a given small upper tail probability p, the VaR of a financial position with log return r_t is

$$\mathrm{VaR} = \begin{cases} \mu_n - \frac{\sigma_n}{\xi}\left\{1 - [-n \ln(1-p)]^{-\xi}\right\} & \text{if } \xi \neq 0 \\ \mu_n - \sigma_n \ln[-n \ln(1-p)] & \text{if } \xi = 0 \end{cases}$$

where n is the length of the subperiod.

Example

- Consider the daily log return, in percentage, of IBM stock from July 3, 1962, to December 31, 1998.
- Given have $\hat{\sigma}_n = 0.945$, $\hat{\mu}_n = 2.583$, and $\hat{\xi} = 0.335$ for $n = 63$.
- For the left-tail probability $p = 0.01$ the corresponding VaR is

$$\begin{aligned} \mathrm{VaR} &= 2.583 - \frac{0.945}{0.335}\left\{1 - [-63\ln(1-0.01)]^{-0.335}\right\} \\ &= 3.04969 \end{aligned}$$

Example

- Thus, for daily negative log returns of the stock, the upper 1% quantile is 3.04969 .
- If one holds a long position on the stock worth \$10 million, then the estimated VaR with probability 1% is $\$10,000,000 \times 0.0304969 = \$304,969$.
- If the probability is 0.05, then the corresponding VaR is $\$166,641$

Example

- If we chose $n = 21$ (i.e., approximately 1 month), then $\hat{\sigma}_n = 0.823, \hat{\mu}_n = 1.902$, and $\hat{\xi} = 0.197$. The upper 1% quantile of the negative log returns based on the extreme value distribution is

$$\begin{aligned} \mathrm{VaR} &= 1.902 - \frac{0.823}{0.197}\left\{1 - [-21 \ln(1 - 0.01)]^{-0.197}\right\} \\ &= 3.40013 \end{aligned}$$

Example

- For a long position of $\$10,000,000$, the corresponding 1-day horizon VaR is $\$340,013$ at the 1% risk level. If the probability is 0.05, then the corresponding VaR is $\$184,127$. In this particular case, the choice of $n = 21$ gives higher VaR values.

Return Level

- The g n -block return level, $R_{n,g}$, is that level which is exceeded in one out of every g blocks of size n. That is, $R_{n,g}$ is the loss value such that
$$\Pr\{M_n > R_{n,g}\} = 1/g$$
- $R_{n,g}$ is simply the $1-1/g$ quantile of limiting GEV distribution:
$$\begin{aligned} R_{n,g} &\approx H^{-1}_{\xi,\mu_n,\sigma_n}(1-1/g) \\ &= \mu_n - \frac{\sigma_n}{\xi}\left(1-(-\log(1-1/g))^{-\xi}\right) \end{aligned}$$
provided that $g \neq 0$.

Modeling Extremes Over High Thresholds

- Idea: Modeling only block maxima data is wasteful of data (only the maxima of large blocks are used).
- A more efficient alternative approach that utilizes more data is to model the behavior of extreme values above some high threshold. This method is often called peaks over thresholds (POT).

Modeling Extremes Over High Thresholds

- An advantage of the POT approach is that common risk measures like Value-at-Risk (VaR) and expected shortfall (ES) may easily be computed.

Risk Measures

- Value-at-Risk (VaR). For $0 \leq q < 0.05$, say, VaR_{1-q} is the $(1-q)$ th quantile of the loss distribution F

$$VaR_{1-q} = F^{-1}(1-q)$$

where F^{-1} is the inverse of F.

- Expected Shortfall (ES). ES_{1-q} is the expected loss size, given that VaR_{1-q} is exceeded:

$$ES_{1-q} = E\left[X \mid X > VaR_{1-q}\right]$$

Note: ES_{1-q} is related to VaR_{1-q} via

$$ES_{1-q} = VaR_{1-q} + E\left[X - VaR_{1-q} \mid X > VaR_{1-q}\right]$$

The Generalized Pareto Distribution

- Let $X_1, X_2, \ldots$ be a sequence of $i.i.d$ random losses with an unknown CDF F
- Let $M_n = \max\{X_1, \ldots, X_n\}$.
- A natural measure of extreme events are values of the X_i that exceed a high threshold u.
- Define the exceedances above the high threshold u :

$$Y = X - u, X > u$$

The Generalized Pareto Distribution

Then the excess distribution (aka tail distribution) above the threshold u is the conditional probability

$$
\begin{aligned}
F_u(y) &= \Pr(Y \leq y) \\
&= \Pr\{X - u \leq y \mid X > u\} = \frac{F(y+u) - F(u)}{1 - F(u)}
\end{aligned}
$$

for $y > 0$

The Generalized Pareto Distribution

For the class of distributions F such that the CDF of the standardized value of M_n converges to a GEV distribution, for large enough u there exists a positive function $\beta(u)$ such that $F_u(y)$ is well approximated by the generalized Pareto distribution (GPD)

$$G_{\xi,\beta(u)}(y) = \begin{cases} 1-(1+\xi y/\beta(u))^{-1/\xi} \text{ for } \xi \neq 0 \\ 1-\exp(-y/\beta(u)) \text{ for } \xi = 0 \end{cases},$$

where $\beta(u) > 0$, $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\beta(u)/\xi$ when $\xi < 0$.

Mean Excess Function

- Suppose the threshold excess $X - u_0$ follows a GPD with parameters $\xi < 1$ and $\beta(u_0)$.
- Then the mean excess over the threshold u_0 is

$$E[X - u_0 \mid X > u_0] = \frac{\beta(u_0)}{1 - \xi}$$

Mean Excess Function

- For any $u > u_0$, define the mean excess function $e(u)$ as

$$e(u) = E[X - u \mid X > u] = \frac{\beta(u_0) + \xi(u - u_0)}{1 - \xi}$$

- Alternatively, for any $y > 0$

$$e(u_0 + y) = E[X - (u_0 + y) \mid X > u_0 + y]$$
$$= \frac{\beta(u_0) + \xi y}{1 - \xi}$$

- The mean excess function is a linear function of $y = u - u_0$.

Graphical Diagnostic for determining u_0

- Define the empirical mean excess function

$$e_n(u) = \frac{1}{n_u} \sum_{i=1}^{n_u} \left(x_{(i)} - u \right)$$

where $x_{(i)}$ $(i = 1, \ldots, n_u)$ are the values of x_i such that $x_i > u$

- The mean excess plot is a plot of $e_n(u)$ against u and should be linear in u for $u > u_0$

Maximum Likelihood Estimation

Let $x_1, \ldots, x_n$ be iid sample of losses with unknown CDF F.

- For a given u, extreme values are those x_i values for which $x_i - u > 0$. Denote these values $x_{(1)}, \ldots, x_{(k)}$
- Define the threshold excesses as $y_i = x_{(i)} - u$ for $i = 1, \ldots, k$.

Maximum Likelihood Estimation

- If u is large enough, then $\{y_1, \ldots, y_k\}$ may be thought of as a random sample from a GPD with unknown parameters ξ and $\beta(u)$.
- For $\xi \neq 0$, the log-likelihood function based on the GPD is

$$
\begin{aligned}
l(\xi, \beta(u)) = &-k \ln(\beta(u)) \\
&- (1 + 1/\xi) \sum_{i=1}^{k} \ln\left(1 + \xi y_i/\beta(u)\right)
\end{aligned}
$$

Estimating the Tails of the Loss Distribution

- For a sufficiently high threshold $u, F_u(y) \approx G_{\xi,\beta(u)}(y)$. Seting $x = u + y$ an approximation to the tails of $F(x)$ for $x > u$ is given by

$$F(x) = (1 - F(u))G_{\xi,\beta(u)}(y) + F(u)$$

- Estimate $F(u)$ using the empirical CDF

$$\hat{F}(u) = \frac{(n-k)}{n}$$

$k =$ number of exceedences over u

Estimating the Tails of the Loss Distribution

- Combine $\hat{F}(u)$ with $G_{\hat{\xi},\hat{\beta}(u)}(y)$ to give

$$\hat{F}(x) = 1 - \frac{k}{n}\left(1 + \hat{\xi} \cdot \frac{x-u}{\hat{\beta}(u)}\right)$$

- This approximation is used for VaR computations based on the fitted GPD

VaR and ES for GPD

- VaR for GPD. Compute $\hat{F}^{-1}(1-q)$ using

$$\widehat{VaR}_{1-q} = u + \frac{\hat{\beta}(u)}{\hat{\xi}}\left(\left(\frac{n}{k}(1-q)\right)^{-\hat{\xi}} - 1\right)$$

- ES for GPD

$$\widehat{ES}_{1-q} = \frac{\widehat{VaR}_{1-q}}{1-\hat{\xi}} + \frac{\hat{\beta}(u) - \hat{\xi}u}{1-\hat{\xi}}$$