9 | Risk Management

There are several types of risk in financial markets. Credit risk, Liquidity risk, operational risk, and market risk are the four main categories of financial risk. The two widely used risk measures are value at risk (VaR) and expected shortfall (ES) because they can be applied to all types of risks and securities, including complex portfolios.

Value at Risk and Expected shortfall

VaR uses two parameters, the time horizon and the confidence level, which are denoted by T and $1-\alpha$, respectively. Given this, the VaR is a bound such that the loss ofer the horizon is less than this bound with probability. For example, if the horizon is one week, the confidence coefficient is 99%, and the VaR is \$5 million, then there is only a 1% chance of a loss exceeding \$5 million over the next week. We sometimes use the notation $\text{VaR}(\alpha)$ or $\text{VaR}(\alpha, T)$ to indicate dependence of VaR on α or on both α and the horizon T. Usually, $\text{VaR}(\alpha)$ is used with T being understood.

In what follows, we define VaR under a probabilistic framework. If $\mathscr{L}$ is the loss over the holding period T, then $\text{VaR}(\alpha)$ is the αth upper quantile of $\mathscr{L}$. Equivalently, if $\mathscr{R} = -\mathscr{L}$ is the renue, then $\text{VaR}(\alpha)$ is the minus αth quantile of $\mathscr{R}$. For continuous loss distributions, VaR(α) solves

$$P\{\mathscr{L} > \mathrm{VaR}(\alpha)\} = P\{\mathscr{L} \geq \mathrm{VaR}(\alpha)\} = \alpha\,, \tag{9.1}$$

and for any loss distribution, continuous or not,

$$\mathrm{VaR}(\alpha) = \inf\{x : P(\mathscr{L} > x) \leq \alpha\}\,. \tag{9.2}$$

VaR is conceptual simplicity and its results are easy to be interpreted. However, VaR has at least two shortcomings. It focuses on the manageable risks near the center of the distribution but ignore the tails, as a consequence, it is impossible to estimate the risks of rare events. Another criticism of VaR, as will be discussed later, is that it discourages diversification. For these reasons it is being replaced by newer risk measures. One of these newer risk measures is the expected loss given that the loss exceeds VaR, which is called by a variety of names: expected shortfall, the expected loss given a tail event, tail loss, and shortfall. The name expected shortfall and the abbreviation ES will be used here.

For any loss distribution, continuous or not,

$$\mathrm{ES}(\alpha) = \frac{\int_0^\alpha \mathrm{VaR}(u)du}{\alpha}\,, \tag{9.3}$$

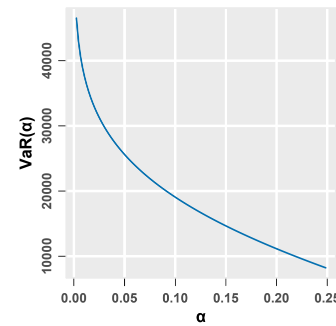
which is the average of VaR(α) over all u that are less than or equal to α. If $\mathscr{L}$ has a continuous distribution,

$$\mathrm{ES}(\alpha) = E\{\mathscr{L} \mid \mathscr{L} > \mathrm{VaR}(\alpha)\} = E\{\mathscr{L} \mid \mathscr{L} \geq \mathrm{VaR}(\alpha)\}\,. \tag{9.4}$$

Eg 9.1. VaR(α) with a normally distributed loss. Suppose that the

yearly return on a stock is normally distributed with mean 0.04 and standard deviation 0.18. If one purchases $100,000 worth of this stock, the VaR(α) with $T = 1$ year is

$$-4000 + z_\alpha\, 18{,}000\,,$$



where z_α is the α-upper quantile of the standard normal distribution. VaR depends heavily on α and in the figure at right ranges from 46,527 when α is 0.025 to 8,226 when α is 0.025.

In applications, risk measures are estimated, and estimation error is another source of uncertainty. This uncertainty can be quantified using a confidence interval for the risk measure.

Estimating VaR and ES with One Asset

We begin with the simple case of a single asset. These risk measures are estimated using historic data to estimate the distribution of returns. We make the assumption that returns are stationary and independent. In the later chapter we will remove the independence assumption by using GARCH models.

Nonparametric Estimation Suppose that we want a confidence coefficient of $1-\alpha$ for the risk measures, we estimate the α-quantile of the return distribution. In the nonparametric method, this quantile is estimated as the α-quantile of a sample of historic returns, which we will call $\hat{q}(\alpha)$. If S is the size of the current position, then the nonparametric estimate of VaR with confidence coefficient of $1-\alpha$ is

$$\hat{\text{VaR}}^{\text{np}}(\alpha) = -S \times \hat{q}(\alpha)\,.$$

To estimate ES with a confident coefficient $1-\alpha$, let $R_1, \ldots, R_n$ be the historic returns and define $\mathscr{L}_i = -S \times R_i$. Then

$$\begin{aligned} \hat{\text{ES}}^{\text{np}}(\alpha) &= \frac{\sum_{i=1}^{n} \mathscr{L}_i \mathbf{1}\{\mathscr{L}_i > \hat{\text{VaR}}^{\text{np}}(\alpha)\}}{\sum_{i=1}^{n} \mathbf{1}\{\mathscr{L}_i > \hat{\text{VaR}}^{\text{np}}(\alpha)\}} \\ &= -S \times \frac{\sum_{i=1}^{n} R_i \mathbf{1}\{R_i < \hat{q}(\alpha)\}}{\sum_{i=1}^{n} \mathbf{1}\{R_i < \hat{q}(\alpha)\}} \end{aligned} \tag{9.5}$$

which is the average of all $\mathscr{L}_i$ exceeding $\hat{\text{VaR}}^{\text{np}}(\alpha)$. Here $\mathbf{1}\{\cdot\}$ is the indicator function.

Eg 9.2. Nonparametric estimation VaR and ES. The data are S&P 500 index fund from Jan 1, 2015 to Oct 26, 2024. Suppose that an investor hold a \$20,000 position and he/she wants to know a 24-hour VaR. We estimate the VaR and ES with daily returns of the S&P 500 data. The sample size $n = 2,470$.

The returns have been converted to percentage.

```
cat("Starting dates:");head(rtSP,2);cat("\nEnding dates:");tail(rtSP,2);
## Starting dates:
##                  GSPC
## 2015-01-05 -1.8447214
## 2015-01-06 -0.8933255
##
## Ending dates:
##                  GSPC
## 2024-10-24  0.2143473
## 2024-10-25 -0.0299492
```



```
n = dim(rtSP)[1];cat("Sample size: n =",n)
## Sample size: n = 2470
```

We will compute for $\alpha = 0.05$. The 0.05 quantile of the returns computed by R's `quantile()` function is -0.0191. A return of -1.6281% on a \$20,000 investment yields a revenue of $-\$325.61$ and therefore the estimated $\hat{\text{VaR}}^{\text{np}}(.05, 24\text{hr})$ is \$341.37.

ES(.05) is obtained by averaging all returns below -1.6281% and multiplying this average by \$20,000. The results is $\hat{\text{ES}}^{\text{np}} = \$558.33$.

```
S = 20000; alpha = .05; q = quantile(rtSP, alpha);
VaR.np = -S*q; ES.np = -S*mean(rtSP[rtSP < q])
VaR.np = VaR.np/100; ES.np = ES.np/100    ## Remove the factor of %
cat("Nonparametric estimates:"); c(quantile = q, VaR = VaR.np, ES = ES.np)
## Nonparametric estimates:
## quantile.5%      VaR.5%          ES
##   -1.706835  341.367042  558.328100
```

Nonparametric estimation using sample quantiles works best when the sample size is large and α are not very small. With smaller sample sizes or smaller values of α, we will consider parametric or semi-parametric estimation.

Parametric Estimation Parametric estimation of VaR and ES has a number of advantages. For example, risk measures can be easily computed for a portfolio of stocks if we assume that their returns have a joint parametric distribution, such as a multivariate t-distribution. Furthermore, parametric estimation allows the use of conditional mean and standard deviation models to adapt the risk measures to the current estimate of mean and volatility. This will be discussed in the later chapters.

Let $F(y|\boldsymbol{\theta})$ be a parametric family of distributions used to model the return distribution and suppose that $\hat{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$, such as the MLE computed from historic returns. Then $F^{-1}(\alpha|\hat{\boldsymbol{\theta}})$ is an estimate of the α-quantile of the return distribution and

$$\hat{\text{VaR}}^{\text{par}}(\alpha) = -S \times F^{-1}(\alpha|\hat{\boldsymbol{\theta}})$$

is a parametric estimate of $\text{VaR}(\alpha)$. As before, S is the size of the current position.

For example, suppose that the return has a t-distribution with mean μ, scale parameter λ and degrees of freedom ν, the VaR estimate is,

$$\hat{\text{VaR}}^{t}(\alpha) = -S \times \left\{\hat{\mu} + \hat{\lambda} F_{\nu}^{-1}(\alpha)\right\}, \tag{9.6}$$

where F_ν is Student's t-distribution function with ν degrees of freedom. In the case that the return has a normal distribution with mean μ and standard deviation σ, the VaR estimate,

$$\hat{\text{VaR}}^{N}(\alpha) = -S \times \left\{\hat{\mu} + \hat{\sigma}\Phi^{-1}(\alpha)\right\}, \tag{9.7}$$

where Φ is the standard normal distribution function.

Let $f(y|\boldsymbol{\theta})$ be the density of $F(y|\boldsymbol{\theta})$. Then the estimate of ES is

$$\hat{\text{ES}}^{\text{par}}(\alpha) = \frac{S}{\alpha} \cdot \int_{-\infty}^{F^{-1}(\alpha|\hat{\boldsymbol{\theta}})} u f(u|\hat{\boldsymbol{\theta}}) du .$$

Computing this integral is not always easy, but in the important

cases of normal and t distributions there are convenient formulas.

Suppose the return has a t-distribution with mean μ, scale parameter λ and degrees of freedom ν. Let f_ν and F_ν be the density and distribution function of Student's t distribution with ν degrees of freedom. The ES is

$$\text{ES}^{t}(\alpha) = S \times \left\{ -\mu + \lambda \cdot \frac{f_\nu\left\{F_\nu^{-1}(\alpha)\right\}}{\alpha} \cdot \frac{\nu + \left\{F_\nu^{-1}(\alpha)\right\}^2}{\nu - 1} \right\}. \tag{9.8}$$

Clearly, the formula of ES for t distribution in (9.8) is defined for the degrees of freedom $\nu > 1$, implying only finite first moment in return series. The formula is flexible enough to accommodate a return series without finite second moment which is occasionally found in empirical study.

Let ϕ and Φ be the standard normal density and distribution function. The formula for normal distributions is

$$\text{ES}^{N}(\alpha) = S \times \left\{ -\mu + \sigma \cdot \frac{\phi\left\{\Phi^{-1}(\alpha)\right\}}{\alpha} \right\}. \tag{9.9}$$

While stock returns are rarely normally distributed, the formulas of VaR and ES for normal distribution are easier to understand.

Eg 9.3. Parametric estimation of VaR and ES with the same data set as in Eg. 9.2. First we fit a t-distribution to the S&P daily returns using R's `fitdistr()` function. The estimates are $\hat{\mu} = 0.08135\%$, $\hat{\lambda} = 0.64548\%$ and $\hat{\nu} = 2.685$. These are labelled as `m`, `s` and `df` in R's return value `estimate`.

```
library(MASS)
alpha = 0.05; S = 20000;
fit.t = fitdistr(rtSP, "t"); fit.t$est;

##          m          s         df
## 0.08134975 0.64547747 2.68537998

mu = fit.t$est["m"];lambda = fit.t$est["s"];nu = fit.t$est["df"];
q.t = qt(alpha, df = nu); VaR.t = -S*(mu + lambda*q.t);
ES.t = S*(-mu+lambda*dt(q.t,nu)/alpha*(nu+q.t^2)/(nu-1));
VaR.t  = VaR.t /100; ES.t = ES.t/100    ## Remove the factor of %
cat("Parametric t estimates:\n "); c("5%quantile"
                                      = mu + lambda*q.t, VaR = VaR.t, ES = ES.t)

## Parametric t estimates:
##
## 5%quantile.m        VaR.m         ES.m
##    -1.513068   302.613668   535.428934
```

Denote $q_{t,\nu}(\alpha) = F_{\nu}^{-1}(\alpha)$, the α-quantile of Student's t in (9.6) , then

$$\widehat{\mathrm{VaR}}^{t}(\alpha) = -S \times \left\{\hat{\mu} + \hat{\lambda}\, q_{t,\hat{\nu}}(\alpha)\right\}.$$

The parametric estimate $\widehat{\mathrm{VaR}}^{t}(.05) = 302.61$ is smaller than the nonparametric estimate $\widehat{\mathrm{VaR}}^{\mathrm{np}}(.05) = 341.37$. The parametric ES estimate of $\widehat{\mathrm{ES}}^{t}(.05) = 535.43$ is also smaller than the nonparametric estimate $\widehat{\mathrm{ES}}^{\mathrm{np}}(.05) = 558.33$. This discrepancy is due to the difference between the 0.05 quantile estimates of S&P returns, the nonparametric estimate is -1.7068%, while the parametric t estimate is -1.5131%.

The computation of ES for most of parametric models other than normal and t is done by simulation.

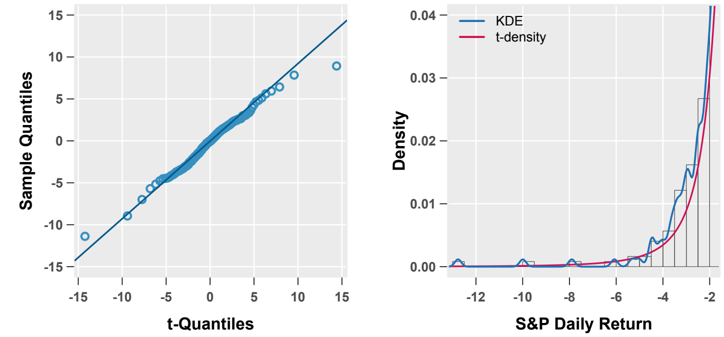


Figure 9.1: t-quantile plot of S&P returns and the left tail of fitted t-density. The parameters of t are estimated by the MLEs, $\hat{\mu} = 0.08135$, $\hat{\lambda} = 0.64548$ and $\nu = 2.685$.

Estimating VaR and ES for a Portfolio

When VaR is estimated for a portfolio of assets rather than a single asset, parametric estimation based on the assumption of multivariate normal or t-distributed returns is very convenient, because the portfolio's return have a univariate normal or t-distributed returns. The portfolio theory Handout 6 can be used to estimate the mean and variance of the portfolio's return.

Estimating VaR becomes complicated when the portfolio contains different financial products, stocks, bonds, options, foreign exchange positions and other assts. When a portfolio contains only stocks, then VaR is relatively straightforward to estimate. We will restrict our attention to this case.

Suppose that the portfolio contains N stocks with weights $\boldsymbol{w} = (w_1, \ldots, w_N)^T$. Denote the returns of the stocks by $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)^T$ and their $N \times N$ covariance matrix $\boldsymbol{\Sigma}$. For a multivariate normal returns, the estimated portfolio mean and standard deviation are

$$\hat{\mu}_P = \boldsymbol{w}^T\hat{\boldsymbol{\mu}} \qquad \text{and} \qquad \hat{\sigma}_P = \sqrt{\boldsymbol{w}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{w}}\ .$$

The formulas (9.7) and (9.9) for VaR and ES assuming normal distribution can be applied. Similarly, for a multivariate t returns, the portfolio mean and scale are estimated by

$$\hat{\mu}_P = \boldsymbol{w}^T\hat{\boldsymbol{\mu}} \qquad \text{and} \qquad \hat{\lambda}_P = \sqrt{\boldsymbol{w}^T\hat{\boldsymbol{\Lambda}}\boldsymbol{w}}\ . \tag{9.10}$$

The portfolio VaR and ES can be calculated by the parametric t formulas (9.6) and (9.8) .

Eg 9.4. VaR and ES for portfolios. The data in Eg. 4.1, pages 76-77 will be used in this example. These are weekly returns of 4 stocks, Adobe, Microsoft, Oracle and Qualcomm from Jan. 1, 2007 to Aug 31, 2024. The sample size $n = 922$.

```
head(yt,2)

##                  ADBE       MSFT      ORCL     QCOM
## 2007-01-05 -1.223407 -0.7394916  2.875422 2.353678
## 2007-01-12 -1.638160  5.1613376 -0.796784 2.375241

tail(yt,2)

##                  ADBE        MSFT     ORCL     QCOM
## 2024-08-23 0.8706909 -0.40226874 1.243420 0.763718
## 2024-08-30 2.8446950  0.08394137 1.497454 1.032125

n = dim(yt)[1]; N = dim(yt)[2]; c(n = n, N = N)

##   n   N
## 922   4
```

The returns of these 4 stocks were modeled with a multivariate t distribution. The parameters were estimated by the MLE. The degrees of freedom estimate is $\hat{\nu} = 3.489$. The MLE of mean and scale matrix are shown in the output.


```
## MLE
nu;  ## degrees of freedom

## [1] 3.89

mus = est$center; mus   ## mean vector

##      ADBE      MSFT      ORCL      QCOM
## 0.4311421 0.4276957 0.2755423 0.3637282

Lambda = est$cov; Lambda  ## scale matrix Lambda

##           ADBE     MSFT     ORCL      QCOM
## ADBE 10.196218 4.946401 4.818932  5.003304
## MSFT  4.946401 6.546973 3.877660  3.854161
## ORCL  4.818932 3.877660 7.094651  3.615896
## QCOM  5.003304 3.854161 3.615896 10.800534
```

We will compute the one-week VaR and ES of the minimum variance portfolios and tangency without constraints. These two portfolios were obtained by using the MLE of mean vector $\hat{\boldsymbol{\mu}}$ and scale matrix $\hat{\boldsymbol{\Lambda}}$. We will obtain the portfolio wights by the explicit formulas.

```
## Minimum Variance Portfolio
invL = solve(Lambda); ones = rep(1,N)
w.min = (invL%*%ones)[,1]/(t(ones)%*%invL%*%ones)[1,1];
cat("Minimum variance portfolio weights:");w.min;

## Minimum variance portfolio weights:
##       ADBE       MSFT       ORCL       QCOM
## 0.02488223 0.42714842 0.36869235 0.17927699
```

The portfolio is a t-distribution, its parameter estimates $\hat{\mu}_P = .3602\%$, $\hat{\lambda}_P = 2.2450\%$ and DF $\hat{\nu} = 3.89$. The VaR and ES can be estimated by plugging these estimates to (9.6) and (9.8). The 5th percentile is -4.465%.

```
## Computing VaR and ES of Minimum Variance Portfolio
alpha = 0.05; S = 20000;
mu = sum(w.min*mus);
lambda = sqrt(as.vector(t(w.min)%*%Lambda%*%w.min));
c(mu = mu, lambda = lambda) ## mean and scale of portfolio

##        mu    lambda
## 0.3602157 2.2450463

q.t = qt(alpha, df = nu); VaR = -S*(mu + lambda*q.t);
ES = S*(-mu+lambda*dt(q.t,nu)/alpha*(nu+q.t^2)/(nu-1));
VaR  = VaR /100; ES = ES/100    ## Remove the factor of %
out = c("5%quantile" =  mu + lambda*q.t, VaR = VaR, ES = ES)
cat("Minimum variance portfolio: ");out

## Minimum variance portfolio:
##  5%quantile         VaR          ES
##   -4.465031   893.006156 1388.094641
```

The computation of the one-week VaR and ES for the tangency portfolio are similar. The tangency portfolio is obtained with the annul risk free rate 4.70%.

```
## Tangency Portfolio computaton
mu.f = 4.7/52 ## weekly risky rate
m.ex = mus -mu.f; ## excess returns
w.T = (invL%*%m.ex)[,1]/(t(ones)%*%invL%*%m.ex)[1,1]
cat("Tangency Portfolio weights:"); w.T;

## Tangency Portfolio weights:
##       ADBE       MSFT       ORCL       QCOM
##  0.2640451  0.8031646 -0.1969646  0.1297549
```

The tangency portfolio mean $\hat{\mu}_p = 0.4503\%$, scale $\hat{\lambda} = 2.5928\%$ and DF $\hat{\nu} = 3.89$. The 5th percentile of the portfolio return distribution is -5.122%.


```
mu = sum(w.T*mus);
## Compute VaR and ES of tangency portfolio
lambda = sqrt(as.vector(t(w.T)%*%Lambda%*%w.T));
c(mu = mu, lambda = lambda) ## mean and scale of tangency portfolio

##        mu    lambda
## 0.4502744 2.5927701

q.t = qt(alpha, df = nu); VaR = -S*(mu + lambda*q.t);
ES = S*(-mu+lambda*dt(q.t,nu)/alpha*(nu+q.t^2)/(nu-1));
VaR  = VaR /100; ES = ES/100    ## Remove the factor of %
out = c("5%quantile" = mu + lambda*q.t, VaR = VaR, ES = ES)
cat("Tangency portfolio: "); out

## Tangency portfolio:
## 5%quantile        VaR         ES
##   -5.12233 1024.46604 1596.23628
```

A portfolio based on a multivariate t distribution is an exception, most of parametric approaches require simulations, for example, copula based simulations.

Semiparametric Estimation of VaR and ES with Heavy Tailed Distributions

In estimating VaR, the nonparametric estimator is feasible for larger α and sample sizes but not for smaller α. For example, if the sample had 1000 returns, then the 0.05-quantile is reasonably accurate, but not estimation of the 0.001 quantile. Parametric estimation can estimate VaR for any value of α but is sensitive to mis-specfication of the tail when α is small as we see in Eg. 9.3. Furthermore, it cannot be applied to a portfolio consisting of any collection of stock returns.

An alternative approach is semiparamatric estimation which assume that the left tail of the return density $f(\cdot)$ follows a power law,

$$f(y) \sim A|y|^{-(a+1)}, \qquad \text{as} \quad y \to -\infty\,, \tag{9.11}$$

where $A > 0$ is a constant, $a > 0$ is the tail index. The symbol "$\sim$" means the ration of the LHS and RHS converges to 1. Therefore,

$$P(R \le y) = \int_{-\infty}^{y} f(u)du \sim \frac{A}{a}|y|^{-a}, \qquad \text{as} \quad y \to -\infty\,,$$

and for $y_0 > 0$ and $y_1 > 0$ then

$$\frac{P(R < -y_1)}{P(R < -y_0)} \approx \left(\frac{y_1}{y_0}\right)^{-a}.$$

Now suppose that, for $0 < \alpha < \alpha_0$, $y_0 = \text{VaR}(\alpha_0)$ and $y_1 = \text{VaR}(\alpha)$ and without loss of generality, we use $S = 1$. Then

$$\frac{\alpha}{\alpha_0} = \frac{P\{R < -\text{VaR}(\alpha)\}}{P\{R < -\text{VaR}(\alpha_0)\}} \approx \left\{\frac{\text{VaR}(\alpha)}{\text{VaR}(\alpha_0)}\right\}^{-a}.$$

For smaller α, we can obtain $\text{VaR}(\alpha)$ using a larger value of α_0 by

$$\text{VaR}(\alpha) = \text{VaR}(\alpha_0)\left(\frac{\alpha_0}{\alpha}\right)^{1/a}. \tag{9.12}$$

The $\text{VaR}(\alpha_0)$ can be estimated well by a nonparametric estimate, the estimate of tail index will be discussed later.

To find a formula for $\text{ES}(\alpha)$, we assume further that for some $c < 0$,

$$f(y) = A|y|^{-(a+1)}, \qquad y \le c\,,$$

so we have equality in (9.11). Then for any $s \le c$, $P(R \le s) = A/a|s|^{-a}$ and the conditional density for R given that $R \le s$ is

$$f(y|R \le s) = \frac{A|y|^{-(a+1)}}{P(R \le s)} = a|s|^{a}|y|^{-(a+1)}. \tag{9.13}$$

It follows that the for $a > 1$,

$$E\left\{|R|\,\middle|\,R \le s\right\} = a|s|^a \int_{-\infty}^{s} |y|^{-a} dy = \frac{a}{a-1}|s|\,.$$

The formula for $\text{ES}(\alpha)$ can be otained by letting $s = -\text{VaR}(\alpha)$,

$$\text{ES}(\alpha) = \frac{a}{a-1}\text{VaR}(\alpha)\,.$$

This formula gives an estimate of $\text{ES}(\alpha)$ from the estimate of $\text{VaR}(\alpha)$.

Estimating the Tail Index

Two estimators of tail index will be introduced, the regression estimator and the Hill estimator. The former is the least squares estimation, the latter is the likelihood estimation. The Hill estimator is heavily dependent on the choice of bandwidth.

Regression Estimator Equation (9.11) gives the linear function bu log transformation,

$$\log\{P(R \le -y)\} = \log(A/a) - a\log(y)\,.$$

Let $R_{(1)}, \ldots, R_{(n)}$ be the order statistics of the returns, $R_1, \ldots, R_n$. The nonparametric estimate of $P(R \le R_{(k)})$ is k/n, thus,

$$\log(k/n) \approx \log(A/a) - a\log\{-R_{(k)}\}\,.$$

Equivalently, with rearranging of the last equation,

$$\log\{-R_{(k)}\} \approx \frac{1}{a}\log\left(\frac{A}{a}\right) - \frac{1}{a}\log\left(\frac{k}{n}\right). \tag{9.14}$$

The approximation (9.14) is expected to be accurate only if $|R_{(k)}|$ is large, that is, k's are small. The regression estimator is $\hat{a} = -1/\hat{\beta}_1$ where $\hat{\beta}_1$ is the least squares estimator of the slope in (9.14) by fitting $\log\{-R_{(k)}\}$ on $\log(k/n)$, $k = 1, \ldots, m$ for $m = o(n)$, that is, a smaller order of n.

Eg 9.5. Estimating the left tail index. The data in this example are the daly returns of S&P 500 as in Eg. 9.3 and Eg. 9.2 .

R function `sort()` will be used for sorting the data. The regression estimator of the tail index is calculated by regressing $\log(R_{(k)})$ on $\log(k/n)$, $k = 1, \ldots,, m$ for bandwidths $m = n^s$, $s = 0.5, 0.55, \ldots, 0.8$.

In addition to $\hat{a}$, we also list the slope, standard errors of the slope estimates and standard deviations of residuals for reference.These values can be obtained by applying `coef()`,`vcov()` and `sigma()` on the `lm()` object.

```
yt = sort(as.numeric(rtSP)) # sort from smallest to largest
s = seq(0.5, 0.8, 0.05); s  ## a sequence of exponents

## [1] 0.50 0.55 0.60 0.65 0.70 0.75 0.80

m = round(n^s); names(m) = paste0("n^",s); m

##  n^0.5 n^0.55  n^0.6 n^0.65  n^0.7 n^0.75  n^0.8
##     50     73    109    160    237    350    518

out = matrix(nrow = 3, ncol = length(m))
rownames(out)= c("slope","se", "sig.e")
colnames(out)= paste("m", m, sep = " = ")
```



```
for(i in 1:length(m)){
  x = log((1:m[i])/n); y = log(-yt[1:m[i]])
  lse = lm(y~x)
  out[,i] = c(coef(lse)[2], sqrt(vcov(lse)[2,2]), sigma(lse))
}
out = rbind(out,ahat = -1/out["slope",] );   ## ahat = 1/slope
round(out,5)

##          m = 50   m = 73  m = 109  m = 160  m = 237  m = 350  m = 518
## slope -0.36090 -0.36351 -0.37882 -0.39764 -0.42697 -0.48330 -0.55563
## se     0.00995  0.00672  0.00486  0.00389  0.00381  0.00510  0.00573
## sig.e  0.06194  0.05201  0.04704  0.04648  0.05610  0.09232  0.12728
## ahat   2.77085  2.75097  2.63974  2.51487  2.34207  2.06910  1.79977
```

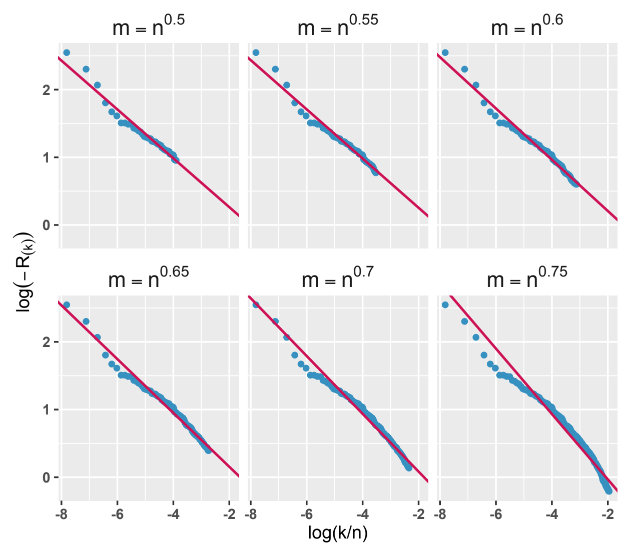


Figure 9.2: Regression estimators of the left tail distribution slop of S&P daily returns, with $m = n^s, s = 0.5, \ldots, 0.75$. The tail index estimator is $\hat{a} = -1/\text{slope}$.

Figure 9.2 shows the first fix regression plots. Both $m = n^{0.65} = 160$ and $m = n^{0.7} = 237$ are seen to be the best two. We select the larger bandwidth $m = 237$, which also has the smallest residual standard deviation indicating the best linear fit among the 7 regressions. The slope of the line with $m = 237$ is -0.4270, the tail index estimator is $\hat{a} = 1/0.4270 = 2.342$.

Suppose that we have invested \$20,000 in $S\&P$ 500 index fund. We will use $\alpha_0 = 0.1$ in equation (9.12), the nonparamatric estimate of 0.1-quantile is -1.1103%, thus $\widehat{\text{VaR}}^{\text{np}}(\alpha_0) = 20000 \times 1.1103/100 =$ \$222.0625. Using (9.12), the estimate of the 24-hour VaR(α) and ES(α) for $\alpha < \alpha_0 = 0.1$ are

$$\widehat{\text{VaR}}(\alpha) = 222.0625\left(\frac{0.1}{\alpha}\right)^{1/\hat{a}}, \quad \widehat{\text{ES}}(\alpha) = \frac{\hat{a}}{\hat{a}-1}\widehat{\text{VaR}}(\alpha). \quad (9.15)$$

```
a = out["ahat", which(m == 237)]
alpha0 = 0.1; alpha = seq(0.01,0.05,0.01)
VaR0 = -S*quantile(yt,alpha0)/100  ## yt is in %
cat("Nonparametric estimates: ", paste(c("0.1-quantile","\t VaR(.1)"),
    round(c(quantile(yt,alpha0), VaR0), 4), sep = " = "))

## Nonparametric estimates:  0.1-quantile = -1.1103   VaR(.1) = 222.0625

VaR = VaR0*(alpha0/alpha)^(1/a); names(VaR) = paste(alpha)
ES = a/(a-1)*VaR
cat("Semiparametric Risk Estimate for alpha = ", alpha, ":");rbind(VaR,ES)

## Semiparametric Risk Estimate for alpha =  0.01 0.02 0.03 0.04 0.05 :
##          0.01     0.02     0.03     0.04     0.05
## VaR  593.5375 441.4856 371.3046 328.3862 298.5430
## ES  1035.7931 770.4445 647.9704 573.0727 520.9928
```

Hill Estimator The Hill estimator of the left tail index a uses (9.11) with all data less than a constant c that is sufficiently small. The choice of c is crucial and will be discussed later. Let $n(c)$ be the number of R_i less than or equal to c, by (9.13), the conditional density of R_i given that $R_i \le c$ is $a|c|^a|y|^{-(a+1)}$. Therefore, the like-

lihoodd for $R_{(1)}, \ldots, R_{(n(c))}$ is

$$L(a) = \left(\frac{a|c|^a}{|R_{(1)}|^{a+1}}\right)\cdots\left(\frac{a|c|^a}{|R_{(n(c))}|^{a+1}}\right),$$

and the log-likelihood is

$$\log\{L(a)\} = \sum_{i=1}^{n(c)}\left\{\log(a) + a\log|c| - (a+1)\log\left|R_{(i)}\right|\right\}. \quad (9.16)$$

Differentiating the RHS with respect to a and setting the derivative equal to 0 gives the equation

$$\frac{n(c)}{a} = \sum_{i=1}^{n(c)}\log\left\{R_{(i)}/c\right\}.$$

The local MLE of a, which is called the *Hill estimator*, is

$$\hat{a}^{\text{H}}(c) = \left\{\frac{1}{n(c)}\sum_{i=1}^{n(c)}\log(R_{(i)}/c)\right\}^{-1}. \quad (9.17)$$

The choice of c and thus $n(c)$ has been investigated by several researchers, but there is no general consensus on the best choice available. Choice of c can be achieved by Hill plots. In practice, one may plot the Hill estimator $\hat{a}^H(c)$ against $n(c)$ and find a proper $n(c)$ such that the estimate appears to be stable.

From the Hill plots in Figure 9.3, the Hill estimates varies very little, ranging from 2.453 to 2.512, when $n(c)$ in the range of $[285, 304]$, the corresponding a-estimates are within 1.656 and 1.676. The estimate is taken to be the average of these estimates, that is, $\hat{a}^{\text{H}} =$

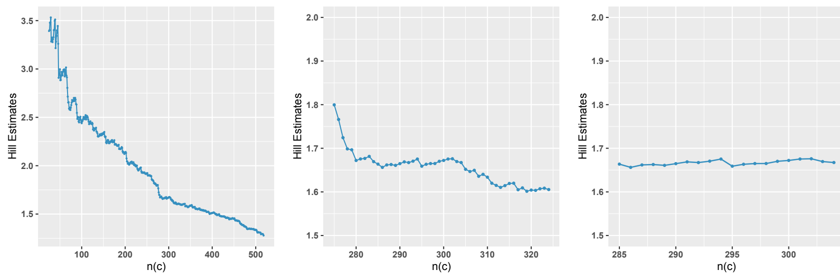


Figure 9.3: Left: Hill estimates for $n(c)$ between [25 : 518]; Center: Hill estimates for $n(c)$ in [275 : 324]; and Right: Hill estimates for $n(c)$ in [285 : 304].

1.667. This is much lower than the regression estimate $\hat{a} = 2.342$.

```
n0 = round(n*0.01); n1 = round(n^0.8)
nc = n0:n1; cat("range of n(c):", range(nc)); cat("range of c:", range(yt[nc]))

## range of n(c): 25 518
## range of c: -3.341639 -0.5330535

hill = c()
for(i in nc) hill = c(hill, 1/mean(log(yt[1:i]/yt[i])))
ind = which(nc >= 285 & nc <= 304);
cat("range of ahat between nc in [285, 304]:", range(hill[ind]))

## range of ahat between nc in [285, 304]: 1.656236 1.675975

a.hill = mean(hill[ind])
cat("Hill estimate:", a.hill)

## Hill estimate: 1.666731
```

We use the same $\alpha_0 = 0.1$, the nonparametric VaR at 0.1 is 222.0625, the estimator of VaR and ES with Hill estimator can be computed,

$$\hat{\text{VaR}}^{\text{H}}(\alpha) = 222.0625\left(\frac{0.1}{\alpha}\right)^{1/\hat{a}^{\text{H}}}, \quad \hat{\text{ES}}^{\text{H}}(\alpha) = \frac{\hat{a}^{\text{H}}}{\hat{a}^{\text{H}} - 1}\hat{\text{VaR}}(\alpha).$$

We compute both semiparametric estimates of the 24-hour VaR and

ES at $\alpha = 0.01, 0.02, \ldots, 0.05$. The Hill estimates all uniformly higher due to the its tail index is lower comparing to that of regression tail index estimate. For $\alpha = 0.05$, both semiparametric estimates of VaR are close to but lower than the t estimate. The semiparametric estimates of ES are much lower than either the nonparametric or t-estimates.

```
VaR = VaR0*(alpha0/alpha)^(1/a.hill); names(VaR) = paste(alpha)
ES = a.hill/(a.hill-1)*VaR
cat("Risk with Hill Estimate for alpha = ", alpha, ":");rbind(VaR,ES)

## Risk with Hill Estimate for alpha =  0.01 0.02 0.03 0.04 0.05 :
##          0.01      0.02      0.03     0.04     0.05
## VaR  883.9994  583.2315  457.2885 384.7955 336.5784
## ES  2209.8699 1457.9939 1143.1548 961.9328 841.3970
```

For $\alpha = 0.05$, nonparametric and Hill-based semiparametric estimates for VaR are similar, they also have highest two ES estimates but values are quite different. The Hill-based estimate is much higher due to its tail index estimate being much lower than those of t and regression. The parametric -t and regression based semiparametric have close estimates for both VaR and ES.

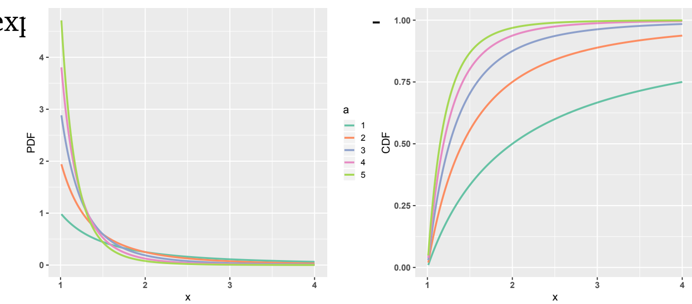
	Nonpara	t-distribtn	Regression	Hill-est
Tail index		2.685	2.42	1.667
VaR	341.37	302.61	298.54	336.58
Es	558.33	535.43	520.99	841.40

Table 9.1: Nonparametric, parametric and semiparametric estimates of 1-day VaR and ES ousing S&P 500 daily returns from 1/1/2015 to 10/26/2024.

Pareto Distributions A random variable Y has a Pareto distribution, named after the Swiss economist Vilfredo Pareto(1948-1923), if its CDF and PDF

$$F(y) = \begin{cases} 1 - \left(\dfrac{c}{y}\right)^a & , \; y > c \\ 0 & , \text{ otherwise} \end{cases} , \quad f(y) = \begin{cases} \dfrac{ac^a}{y^{a+1}} & , \; y > c \\ 0 & , \text{ otherwise} . \end{cases}$$

The exp



ise.

Figure 9.4: The PDF and CDF of Pareto distributions with parameters $c = 1$ and $a = 1, \ldots, 5$.

Equation 9.13 states that the loss, conditional on being above $|c|$, has a Pareto distribution. Furthermore, if Y has a Pareto distribution with parameter a and c, then conditional distribution of Y, give en that $Y > d > c$, is a Pareto with parameters a and d .

Generalized Pareto Distributions The generalized Pareto distribution has three parameters, location α, scale β and shape ξ . The CDF of the generalized Pareto distribution is given by,

$$P(Y \le y) = \begin{cases} 1 - \left(1 + \dfrac{\xi(y-\alpha)}{\beta}\right)^{-1/\xi} & , \quad \xi \ne 0 \\ 1 - \exp\left(-\dfrac{y-\alpha}{\beta}\right) & , \quad \xi = 0 . \end{cases}$$

where it is understood that $1 + \xi(y-\alpha)/\beta > 0$ for $\xi \ne 0$. The expectation for a $\xi > 1$ is $E(Y) = \alpha + \beta/(1-\xi)$. When $\xi > 0$ and $\alpha = 0$, the GDP is the Pareto distribution with $a = 1/\xi$ and $c = \beta/\xi$. The GDP random variables can be generated by the formula,

$$Y = \alpha + \frac{\beta(U^{-\xi} - 1)}{\xi} \sim GDP(\alpha, \beta, \xi)$$

for a uniform random variable, $U \sim \text{Unif}(0,1)$.

Estimating VaR and ES with Generalized Pareto Distributions

Extreme value theory is of particular interest in financial risk analysis because its specialty in extreme quantiles and probabilities. Under the independent assumption, the limiting CDF of the normalized maximum is given by

$$F_*(x) = \begin{cases} \exp\{-(1+\xi x)^{-1/\xi}\} & , \quad \xi \ne 0 \\ \exp\{-\exp(-x)\} & , \quad \xi = 0 \end{cases} . \tag{9.18}$$

The extreme value approach to VaR and ES calculation however has faced some difficulties. An alternative approach proposed is a threshold method which, instead of focusing on extremes, focuses on exceedances of the loss over some high threshold and the times at which the exceedances occur. This approach is also referred as *Peaks over Thresholds* (PoT).

Suppose η is our choice for threshold and x_t is the loss variable of an asset. Suppose that the ith exceedance occurs at t_i. We will focus on the exceedance $y = x_t - \eta$ and t_i and consider the conditional distribution of y given $x_t > \eta$. The basic theory of the PoT approach is to consider the conditional distribution of $x = y + \eta$ given $x > \eta$,

$$P(x \le y + \eta \,|\, x > \eta) = \frac{P(\eta \le x \le y + \eta)}{P(x > \eta)} = \frac{P(x \le y + \eta) - P(x \le \eta)}{1 - P(x \le \eta)}$$

Using $F_*(\cdot)$ of (9.18), the above equation becomes

$$\frac{F_*(y+\eta) - F_*(\eta)}{1 - F_*(\eta)} \tag{9.19}$$

$$= \frac{\exp\left\{-\left(1 + \frac{\xi(y+\eta-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}\right\} - \exp\left\{-\left(1 + \frac{\xi(\eta-\mu)}{\sigma}\right)^{-\frac{1}{\xi}})\right\}}{1 - \exp\left\{-\left(1 + \frac{\xi(\eta-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}\right)}$$

With the approximation $e^{-z} \approx 1 - z$ and a few algebraic steps, we obtain that

$$P(x \le y + \eta | x > \eta) \approx 1 - \left(1 + \frac{\xi y}{\sigma + \xi(\eta - \mu)}\right)^{-\frac{1}{\xi}}, \tag{9.20}$$

where $y > 0$ and $1 + \xi(\eta - \mu)/\sigma > 0$. The case of $\xi = 0$ is taken as the limit of $\xi \to 0$ so that

$$P(x \le y + \eta \,|\, x > \eta) \approx 1 - \exp(-y/\sigma)$$

This is the generalized Pareto distribution with $\alpha = 0$ and $\beta = \psi(\eta) = \sigma + \psi(\eta - \mu)$, that is,

$$G_{\xi,\psi(\eta)}(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi y}{\psi(\eta)}\right)^{-1/\xi} & , \quad \xi > 0, \quad y \ge 0 \\ & , \quad \xi < 0, \quad 0 \le y \le -\psi(\eta)/\xi \\ 1 - \exp\{-y/\psi(\eta)\} & , \quad \xi = 0\,. \end{cases} \tag{9.21}$$

Equations (9.20) and (9.21) shows that the conditional distribution of x given $x > \eta$ is well approximated by a GPD. The GPD share the same important property of the Pareto distribution about a different threshold. Suppose that the excess distribution of x given a threshold η_0 is a GPD with shape parameter ξ and scale parameter $\psi(\eta_0)$. Then, for an arbitrary threshold $\eta > \eta_0$, the excess distribution over the threshold η is also a GPD with the same shape parameter ξ and scale parameter $\psi(\eta) = \psi(\eta_0) + \xi(\eta - \eta_0)$.

Choice of threshold

The main difficulty of the PoT-method is selecting an appropriate threshold. If it's too high there will be too few data to model such that the standard errors will be large. On the other hand a low threshold will give a high bias because the approximation is valid for the tail part. One way to approach the problem is using the graphic tools to select an appropriate η_0.

Mean excess plots Given a high threshold η_0, suppose that the excess $x - \eta$ follows a GPD with parameters ξ and $\psi(\eta_0)$, when $0 < \xi < 1$. The mean excess over the threshold η_0 is

$$E(x - \eta_0, |, x > \eta_0) = \frac{\psi(\eta_0)}{1 - \xi}.$$

For any $\eta > \eta_0$, define the mean excess function $e(\eta)$ as

$$e(\eta) = E(x - \eta \,|\, x > \eta) = \frac{\psi(\eta)}{1 - \xi} = \frac{\psi(\eta_0) + \xi(\eta - \eta_0)}{1 - \xi}$$

In other words, for any $y > 0$,

$$e(\eta_0 + y) = E(x - (\eta_0 + y) \mid x > \eta_0 + y) = \frac{\psi(\eta)}{1-\xi} = \frac{\psi(\eta_0) + \xi y}{1-\xi}.$$

For a fixed ξ, the mean excess function is a linear function of $y = \eta - \eta_0$. This result leads to a simple graphical method to infer the appropriate threshold value η_0. Define the empirical mean excess function as

$$e_n(\eta) = \frac{1}{n_\eta} \sum_{i=1}^{n_\eta} (x_{t_i} - \eta)$$

where n_η is the number of returns that exceed η and x_{t_i} are the corresponding returns. The scatterplot of $e_n(\eta)$ against η is called *the mean excess plot* or *mean residual life plot*, which should be linear in η for $\eta > \eta_0$.

The assumption is that when the plot starts showing a linear behavior a suitable threshold can be estimated. The plot is likely to lose its linear behavior when the threshold gets too high due to the variance of the few extremes left will cause the plot to jump.

Parameter Stability Plots The plot shows the estimates of the shape and scale parameters to find a suitable threshold. Over some range of η, shape and scale parameter stability plots can be made by plotting these estimators against the interval η with the 95% confidence interval for the estimators. Looking at the plots one can find a suitable threshold at the lowest value where the plots are approximately constant.

VaR Calculation based on GPD

From the conditional distribution (9.19) and the GDP in (9.21), we have

$$\frac{F(x) - F(\eta)}{1 - F(\eta)} \approx G_{\eta, \psi(\eta)}(y)$$

where $x = y + \eta$ with $y > 0$. If we estimate CDF $F(\eta)$ of the losses by the empirical CDF, then

$$\hat{F}(\eta) = \frac{n - n_\eta}{n},$$

where n_η is the number of exceedances of threshold η. Consequently,

$$F(x) = F(\eta) + G(y)\{1 - F(\eta)\} \approx 1 - \frac{n_\eta}{n}\left\{1 + \frac{\xi(x-\eta)}{\psi(\eta)}\right\}^{-1/\xi} \tag{9.22}$$

This leads to an alternative estimate of the quantile of $F(x)$ for use in VaR calculation. Specifically for a small upper tail probability α. Let $q = 1 - \alpha$, by solving for x, we can estimate the qth quantile of $F(x)$, denoted by VaR_q,

$$\mathrm{VaR}_q = \eta - \frac{\psi(\eta)}{\xi}\left\{1 - \left[\frac{n}{n_\eta}(1-q)\right]^{-\xi}\right\}.$$

The method to VaR calculation can be done by `R` package `evir`. For the generalized Pareto distribution, ES assumes a simple form.

$$\mathrm{ES}_q = E(x \mid x > \mathrm{VaR}_q) = \mathrm{VaR}_q + E(x - \mathrm{VaR}_q \mid x > \mathrm{VaR}_q).$$