CARM

Contents

2. Probability theory overview

Distributions

The Bernoulli distribution

$\mathrm{Ber}(p)$ with $0 \leq p \leq 1$:

$$F(x) = (1-p)\mathbb{1}_{[0,\infty)}(x) + p\mathbb{1}_{[1,\infty)}(x)$$

The binomial distribution $\mathrm{Bin}(n,p)$ with $n \in \mathbb{N}, 0 \leq p \leq 1$

$$F(x) = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \mathbb{1}_{[k,\infty)}(x)$$

For later use we set for every $m \in \{0, \ldots, n\}$

$$\beta_m(n,p) := \sum_{k=0}^{m} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Note that $\mathrm{Ber}(p) = \mathrm{Bin}(1,p)$ –

The **uniform distribution** $U(a,b)$ with $a < b$:

$$F(x) = \mathbb{1}_{[a,b]}(x) \frac{x-a}{b-a} + \mathbb{1}_{(b,\infty)}(x)$$

The **normal or Gaussian distribution** $N\left(\mu, \sigma^2\right)$ with $\mu \in \mathbb{R}, \sigma > 0$:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy$$

The t or **Student distribution** $t\left(n, \mu, \sigma^2\right)$ for $n \in \mathbb{N} \backslash \{1\}, \mu \in \mathbb{R}, \sigma > 0$:

$$F(x) = \Phi_n\left(\frac{x-\mu}{\sigma}\right), \quad \Phi_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{x} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}} \mathrm{d}y$$

where Γ is the Euler function defined for $a \in \mathbb{R}$ with $a > 0$ by

$$\Gamma(a) = \int_0^\infty z^{a-1} e^{-z} \, \mathrm{d}z$$

The exponential distribution $\mathrm{Exp}(\lambda)$ with $\lambda > 0$:

$$F(x) = \mathbb{1}_{[0,\infty)}(x)\left(1 - e^{-\lambda x}\right)$$

The lognormal distribution $\log N\left(\mu, \sigma^2\right)$ with $\mu \in \mathbb{R}, \sigma > 0$:

$$F(x) = \mathbb{1}_{(0,\infty)}(x) \Phi\left(\frac{\log(x) - \mu}{\sigma}\right)$$

The Pareto distribution $\mathrm{Par}(\theta)$ with $\theta > 0$:

$$F(x) = \mathbb{1}_{[0,\infty)}(x)\left(1 - (1+x)^{-\theta}\right)$$

Mean variance

Definition. Let $X \in \mathcal{X}$. (If well-defined) The **mean** of X is given by

$$E(X) := \int_\Omega X \, \mathrm{d}P = \int_{-\infty}^{\infty} x \, \mathrm{d}F_X(x)$$

The **variance** of X is defined by

$$\operatorname{var}(X) := \int_\Omega (X - E(X))^2 \, \mathrm{d}P = \int_{-\infty}^{\infty} (x - E(X))^2 \, \mathrm{d}F_X(x)$$

Mean and variance examples

	$E(X)$	$\operatorname{var}(X)$
$\operatorname{Ber}(p)$	p	$p(1-p)$
$\operatorname{Bin}(n,p)$	np	$np(1-p)$
$U(a,b)$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$
$N(\mu,\sigma^2)$	μ	σ^2
$t(n,\mu,\sigma^2)$	μ	$\begin{cases} \sigma^2 n(n-2)^{-1} & \text{if } n > 2 \\ \infty & \text{if } 1 < n \leq 2 \end{cases}$
$\operatorname{Exp}(\lambda)$	λ^{-1}	λ^{-2}
$\log N(\mu,\sigma^2)$	$e^{\mu+\frac{1}{2}\sigma^2}$	$\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$
$\operatorname{Par}(\theta)$	$\begin{cases} (\theta-1)^{-1} & \text{if } \theta > 1 \\ \infty & \text{if } \theta \leq 1 \end{cases}$	$\begin{cases} 2(\theta-1)^{-1}(\theta-2)^{-1} & \text{if } \theta > 2 \\ \infty & \text{if } \theta \leq 2 \end{cases}$

Figure 1: Mean variance examples

Quantiles

Proposition. For all $X \in \mathcal{X}$ and $\alpha \in (0,1)$ the following are equivalent: (1) For all $x, y \in \mathbb{R}$ with $F_x(x) = F_X(y) = \alpha$ we have $x = y$.

$$(2)\ q_\alpha^-(X) = q_\alpha^+(X)$$

Interpretation. The lower and upper α -quantiles are distinct only when the distribution function of X is flat at the level α. Corollary. Let $X \in \mathcal{X}$ and $\alpha \in (0,1)$. If F_X is strictly increasing and continuous in a neighborhood of $x \in \mathbb{R}$ with $F_X(x) = \alpha$, then

$$q_\alpha^-(X) = q_\alpha^+(X) = F_X^{-1}(\alpha)$$

Examples of quantiles

	$q_\alpha^-(X) = q_\alpha^+(X)$
$U(a,b)$	$(1-\alpha)a + \alpha b$
$N(\mu,\sigma^2)$	$\mu + \Phi^{-1}(\alpha)\sigma$
$t(n,\mu,\sigma^2)$	$\mu + \Phi_n^{-1}(\alpha)\sigma$
$\operatorname{Exp}(\lambda)$	$-\frac{1}{\lambda}\log(1-\alpha)$
$\log N(\mu,\sigma^2)$	$e^{\mu+\Phi^{-1}(\alpha)\sigma}$
$\operatorname{Par}(\theta)$	$(1-\alpha)^{-\frac{1}{\theta}} - 1$

	$q_\alpha^-(X)$	$q_\alpha^+(X)$
$Ber(p)$	$\begin{cases} 0 & \text{if } 0 < \alpha \leq 1-p \\ 1 & \text{if } 1-p < \alpha < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < \alpha < 1-p \\ 1 & \text{if } 1-p \leq \alpha < 1 \end{cases}$
$Bin(n,p)$	$\begin{cases} 0 & \text{if } 0 < \alpha \leq \beta_0(n,p) \\ 1 & \text{if } \beta_0(n,p) < \alpha \leq \beta_1(n,p) \\ \vdots & \vdots \\ n & \text{if } \beta_{n-1}(n,p) < \alpha < 1 \end{cases}$	$\begin{cases} 0 & \text{if } 0 < \alpha < \beta_0(n,p) \\ 1 & \text{if } \beta_0(n,p) \leq \alpha < \beta_1(n,p) \\ \vdots & \vdots \\ n & \text{if } \beta_{n-1}(n,p) \leq \alpha < 1 \end{cases}$

Correlation

Definition. Let $X \in \mathcal{X}^2$ and assume neither X_1 nor X_2 is null. (If well-defined) The correlation between X_1 and X_2 is given by

$$\operatorname{cor}(X_1, X_2) := \frac{E(X_1 X_2) - E(X_1)\, E(X_2)}{\sqrt{\operatorname{var}(X_1)\operatorname{var}(X_2)}}$$

Correlation provides an indication about linear dependence. Two RV can be uncorrelated and yet display strong nonlinear dependence!

3. Capital adequacy

Ω = set of all future *market scenarios*
$\mathcal{F}$ = set of all observable *market events*
P = historical probability

A = aggregate value of assets
L = aggregate value of liabiliites
X = RV, net capital, where X = A - L

For every terminal scenario $\omega\epsilon\Omega$, we have:

$$X(\omega) = A(\omega) - L(\omega)$$

Two main take-homes:

- **How to determine whether a company is adequately capitalized**
- **What to do if not**

Variance is unsifficient risk measure for CARM, because it doesn't tell what to do in case of inadequacy.

The higher the figure $\rho_A(X)$:
- the more costly to reach acceptability - the riskier position X

Acceptance set

allows us to define **capital adequacy tests**

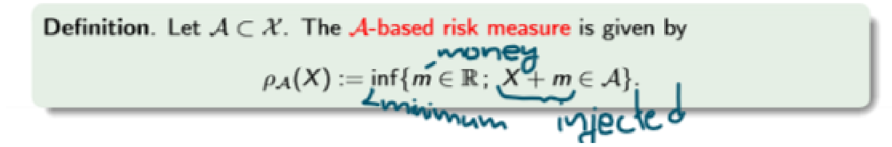


Figure 2: A-based risk measure

Examples of risk measures

Value at Risk (VaR)

Definition. Let $\alpha \in (0,1)$. The Value at Risk (VaR) at level α is defined by

$$\mathrm{VaR}_\alpha(X) := -q_\alpha^+(X)$$

where $-q_\alpha^+(X) \Rightarrow$ upper last quantile

Interpretation

The quantity $\mathrm{VaR}_\alpha(X)$ is the **worst realization of X that may occur in the** $100(1-\alpha)$ **cases**. Alpha is close to zero IRL, between 0.005-0.05.

Easy one : VaR is the **worst** outcome of the $100(1-\alpha)$ best cases. That is precisely the the q.

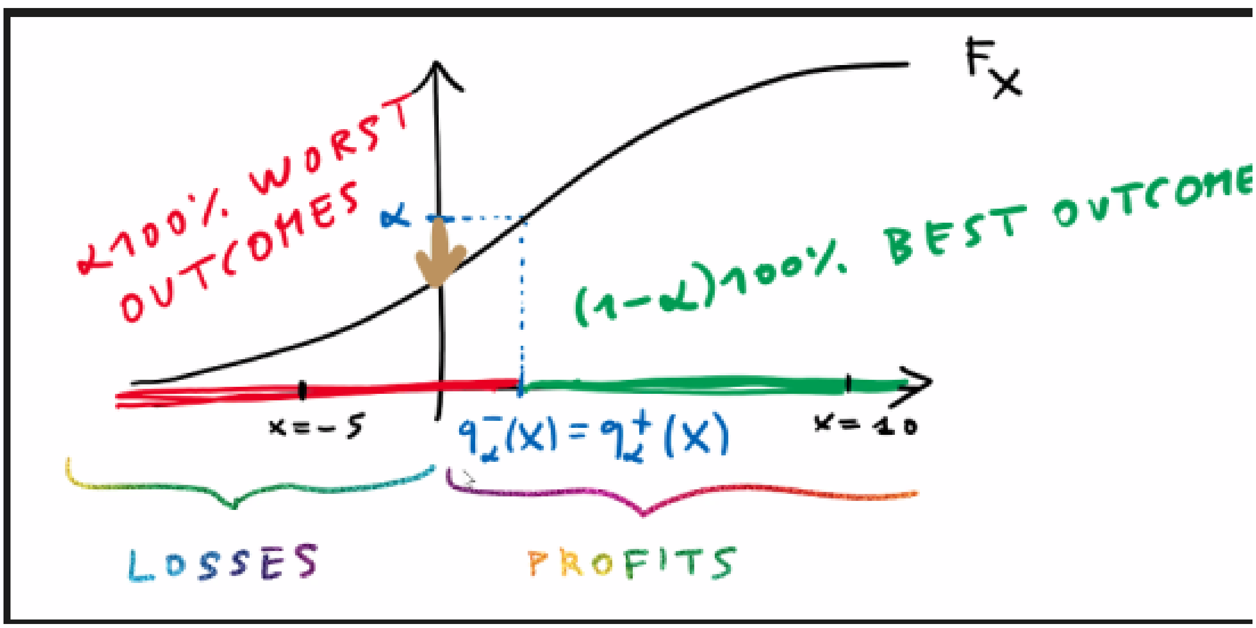


Figure 3: Plot title.

It tells us, how much capital to inject into X in order to **ensure that default prob does not exceed alpha**.

Pitfalls

- in discrete cases, the level might not be the worst cases
- because the way it is defined, it might not be the maximum loss- it can be also a profit.

Operational interpretation

$VaR_\alpha(X)$ represent the **smallest amount of capital** that has to be injected into X to **ensure that default prob. does not exceed level α $

Expected shortfall

$$\mathrm{ES}_\alpha(X) := \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_p(X)\mathrm{d}p$$

Interpretation

ES takes all VaR quantiles left of a given quantile and then normalizes is by α .

Operational interpretation

Recurring description: Acceptance set is such set, which has negative ES (because ES is itself negative, - times - = +)

ES can be interpreted as the smallest amount of capital that has to be injected into X to ensure NO DEFAULTS occur on average in the tail beyond the upper α-quantile of X. $\Rightarrow$ company is AQ under ES at level α if it is solvent on average over the tail beyond the upper α-quantile.

Alternative name - average value at risk

Question - How to choose α ?

With calibration. There is no a-priori theory, usually is set to app. 0.025. In regulation, it is prescribed to 0.01.

Worst-case risk measures

Definition. Let $A \in \mathcal{F}$. The worst-case risk measure based on A is defined by

$$\mathrm{WC}_A(X) := \inf\{m \in \mathbb{R}; m \geq -X \text{ on } A\}$$

Interpretation

WC can be interpretated as the best realization of -X over the event A, or up to a sign as the worst realization of X over the event A.

Operational interpretation

Smallest amount of capital injected into X in order to ensure that NO DEFAULTS occur in the scenarios belonging to A. $\Rightarrow$ company is adequately capitalized based on A if it solvent in each of the test scenarios prescribed by A.

Loss Value at Risk

Definition. Let $\lambda : (-\infty, 0] \to (0,1)$ be increasing. The Loss Value at Risk (LVaR) at level λ is defined by

$$\mathrm{LVaR}_\lambda(X) := \sup_{x<0} \left\{ \mathrm{VaR}_{\lambda(x)}(X - x) \right\}$$

*sup = supremum, i.e. highest value

Interpretation

The parameter x is loss level. For each x we apply VaR to the transformed cap. position X-x, we let the prob. level λ(x) depend on the loss size. Monotonicity ensure highers loss = lower prob. LVaR is the most conservative. It is generalization of Var. If λ=α, we get VaR.

Operational interpretation

Smallest amount of cap. that has to be injected into X that for every loss threshold x, prob. of incurring a default beyond size x does not exceed the level λ(x).

VaR controls only for prob. of default, LVaR controls for prob. of losing some specific value.

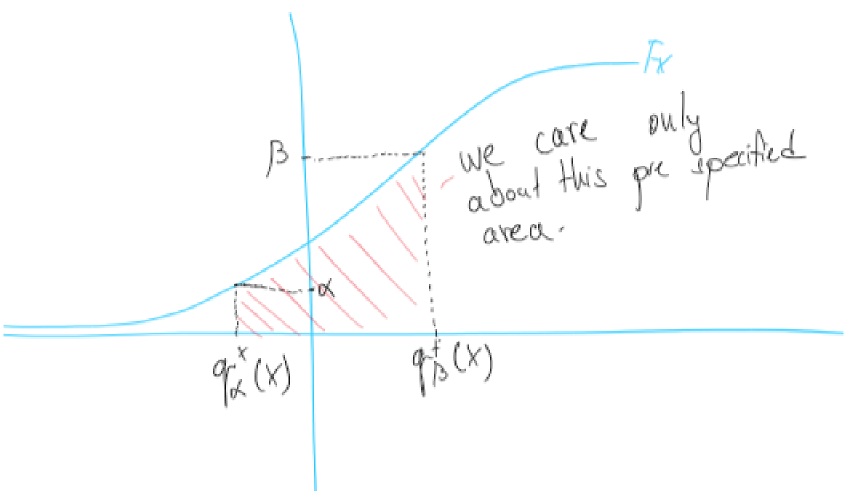
Range Value at Risk

Definition. Let $0 < \alpha < \beta < 1$. The Range Value at Risk (RVaR) at levels α and β is defined by

$$\mathrm{RVaR}_{\alpha,\beta}(X) := \frac{1}{\beta - \alpha} \int_\alpha^\beta \mathrm{VaR}_p(X) \mathrm{d}p$$

Interpretation

$\mathrm{RVaR}_{\alpha,\beta}(X)$ can be interpreted, up to a sign, as a kind of average realization of X among all the realizations that are contained between $q_\alpha^+(X)$ and $q_\beta^+(X)$



We take the cut range, because at extreme ends, we have no data

4. and 5. Properties of Risk Measures

Main take home: keep in mind what is a apt risk measure for CA

$$\left.\begin{aligned} X+m \notin \mathcal{A} \text{ for every } m \in \mathbb{R}, \text{ in which case we have} \\ \rho_{\mathcal{A}}(X)=\infty. \end{aligned}\right\} \begin{aligned} &\text{We have such a position that} \\ &\text{injection of any finite avount is insufficient} \end{aligned}$$

Proposition. Let $X \in \mathcal{X}$. For all $0<\alpha<\beta<1$ we have - $\mathrm{VaR}_\alpha(X)$ is always finite. - $\mathrm{RVaR}_{\alpha,\beta}(X)$ is always finite. - $\mathrm{ES}_\alpha(X)>-\infty$ - $\mathrm{ES}_\alpha(X)<\infty$ if and only if $\mathbb{E}(\min(X,0))>-\infty$. - consider only the negative amounts and check, whether it's finite or not

Impact of capital injections

KEY PROPERTY $\rho_{\mathcal{A}}$ satisfies the following cash-additivity property:

$$\rho_{\mathcal{A}}(X+m)=\rho_{\mathcal{A}}(X)-m$$

Interpretation: Cap. injections decrease cap. req. in a linear way $\Rightarrow$ cap. test are some defnites as **target capital**

$X_0 =$ avalaible capital $\geq$ target capital $= \rho_A(X-X_0)$

where X_0 is capital at time 0 . Indeed, by cash additivity, we have

$$X \in \mathcal{A} \Longleftrightarrow \rho_{\mathcal{A}}(X) \leq 0 \Longleftrightarrow X_0 \geq \rho_{\mathcal{A}}\left(X-X_0\right)$$

Monotonicty

$$X \in \mathcal{A}, Y \geq X \Longrightarrow Y \in \mathcal{A}$$

If company A has more capital than company B, which is adeq. cap., than A is also adeq. cap. Important for holdings, consolidation, etc.

Stability under scaling

We say that $\mathcal{A}$ is stable under scaling whenever

$$X \in \mathcal{A}, a \in[0, \infty) \Longrightarrow a X \in \mathcal{A}$$

Stab. under scaling means acceptability is **independent of the size** od the cap. positions.

!! LVaR is NOT homogenous, scaling doesnt work for it.

Stability under aggregation

If two positions are independtly adeq. cap., than their sum is also adeq. cap. Again important for consolidation, subsidiaries etc.

$\rho_A(X+Y) \leq \rho_A(X)+\rho_A(Y)$

The $\leq$ sign, meaning sometimes two positions aggregated yield an "overcapitalized" positions, meaning: - **incentivizes** aggregation (always true) - **implied** diversification (if certain assumption are full-filled - they have to be i.i.d)

Three main implications:

- sum of cap. req. of indi. subposition is conservative proxy for the cap. req. of the agg. position (challenging to estimate)
- CA can be enforec in a decentralized way
- no incentive to break up balance sheet

Following risk measures are (not) subadditive:

- VaR: NO
- ES: YES
- WC: YES
- LVaR: NO
- RVaR: NO

Coherence

We say that $\mathcal{A}$ is COHERENT iff it's BOTH stable under scaling AND aggregation.

Following risk measures are (not) coherent

- VaR: NO
- ES: YES
- WC: YES
- LVaR: NO
- RVaR: NO

!!! Careful:

If we have a fully-leveraged positions (zero-cost LS portfolio e.g.), we accumulate **loss peaks in the tail**, but because of subaddivity, the position even improves CA.

6. Tail risks

Sensitivity to tail risk

Definition

Definition. We say that $\mathcal{A}$ is sensitive to tail risk whenever
$X \in \mathcal{X}, \mathbb{P}(E) > 0 \Rightarrow X - a\mathbb{1}_E \notin \mathcal{A}$ for $a \in (0, \infty)$ large enough.

Interpretation

If $\mathcal{A}$ is sensitivity to tail risk, a large accum. of loss peaks will always result in a positive cap. req. There risk measures are sensitive to tail risk:

- VaR: NO
- ES: YES
- WC: NO
- LVaR: NO
- RVaR: NO

ONLY *ES* is sensitivity to tail risks!!

Acceptability and distributions

We say that $\mathcal{A}$ is law invariant whenever

$$X \in \mathcal{A}, F_Y = F_X \Longrightarrow Y \in \mathcal{A}$$

Interpretation

Law invariance says that acceptability depends only on the **probability distribution** of a capital position. Meaning we can use statistical estimation to compute capital requirements.

7. Surplus and default profiles

- shareholders' surplus: random var, $S_x = max(X, 0) = max(A - L, 0)$: what is extra from the acc. set, shareholders' payoff
- shareholders' default option, $D_x = -min(X, 0) = max(-X, 0) = max(L - A, 0)$: incorporates the limited liability.If L-A is negative, meaning there is less cash than liabilities to satisfy them, creditors get 0 (max condition) and can suck it

For every capital positiion we have X=Sx-Dx

Protection of liability holders

Definition. We say that $\mathcal{A}$ is surplus invariant if

$$X \in \mathcal{A}, D_Y \leq D_X \Longrightarrow Y \in \mathcal{A}$$

Interpretation

Surplus invariance = company whose default option is smaller than of an acceptable company, it's also acceptable.

Surplus invariance

Proposition. For every acceptance set $\mathcal{A}$ the following are equivalent: - For every capital position $X \in \mathcal{X}$ we have

$$X \in \mathcal{A} \Longleftrightarrow -D_x = \min(X, 0) \in \mathcal{A}$$

$\mathcal{A}$ is surplus invariant. Interpretation. The outcome of a surplus-invariant capital adequacy test is independent of (the size of) the surplus and is only driven by the default profile.

Summary (SUPER IMPORTANT)

	VaR	ES	WC	LVaR	RVaR
monotonicity	✓	✓	✓	✓	✓
positive homogeneity	✓	✓	✓	NO	✓
subadditivity	NO	✓	✓	NO	NO
tail sensitivity	NO	✓	NO(1)	NO(2)	NO
law invariance	✓	✓	NO(1)	✓	✓
surplus invariance	✓	NO	✓	✓	NO
currency formula	✓	NO	✓	NO	NO

Remark. We have the following exceptions:

(1) Unless $\mathbb{P}(A) = 1$ (for instance $A = \Omega$).

(2) Unless $\lambda(x) \to 0$ as $x \to -\infty$.

Review of Statistics

Statistical inference

The basic goal of statistical inference is to provide **assesments about the PD that is assumed to have "generated" the sample**. Basically, we infer some characteristics from the sample distribution, which should have generated the sample and assume those characteristics are to remain constant and keep generating sample subject to the underlying rules and mechanics.

Standard assumption is the sample is i.i.d., even though it might not be necessarily true.

Statistics (backtesting through hypothesis testing)

Definition. Let Θ be a set. Any function $S_n : \mathbb{R}^n \to \Theta$ is called d statistic. We say that S_n is a - (point) estimator if Θ consists of points in $\mathbb{R}^k$. - interval/region estimator if Θ consists of subsets of $\mathbb{R}^k$. - test statistic if Θ contains two elements. - rank statistic if Θ is finite and contains more than two elements.

Interpretation:

the set Θ (statistic) consists of **model parameters** (mean, variance) or **decision parameters** ("accept, reject").

A statistic gives us a **rule to link the data with a specific model parameter or decision variable**

Example. Let $\Theta = \mathbb{R}$. The statistic $M_n : \mathbb{R}^n \to \Theta$ given by

$$M_n\left(x_1, \ldots, x_n\right) := \frac{1}{n} \sum_{i=1}^{n} x_i$$

is called the sample mean. Example. Let $\Theta = [0, \infty)$. The statistic $V_n : \mathbb{R}^n \to \Theta$ given by (for $n > 1$)

$$V_n\left(x_1, \ldots, x_n\right) := \frac{1}{n-1} \sum_{i=1}^{n}\left(x_i - \frac{1}{n} \sum_{i=1}^{n} x_i\right)^2$$

is called the sample variance.

Statistical Estimation of VaR and ES

Algoritm for estimation

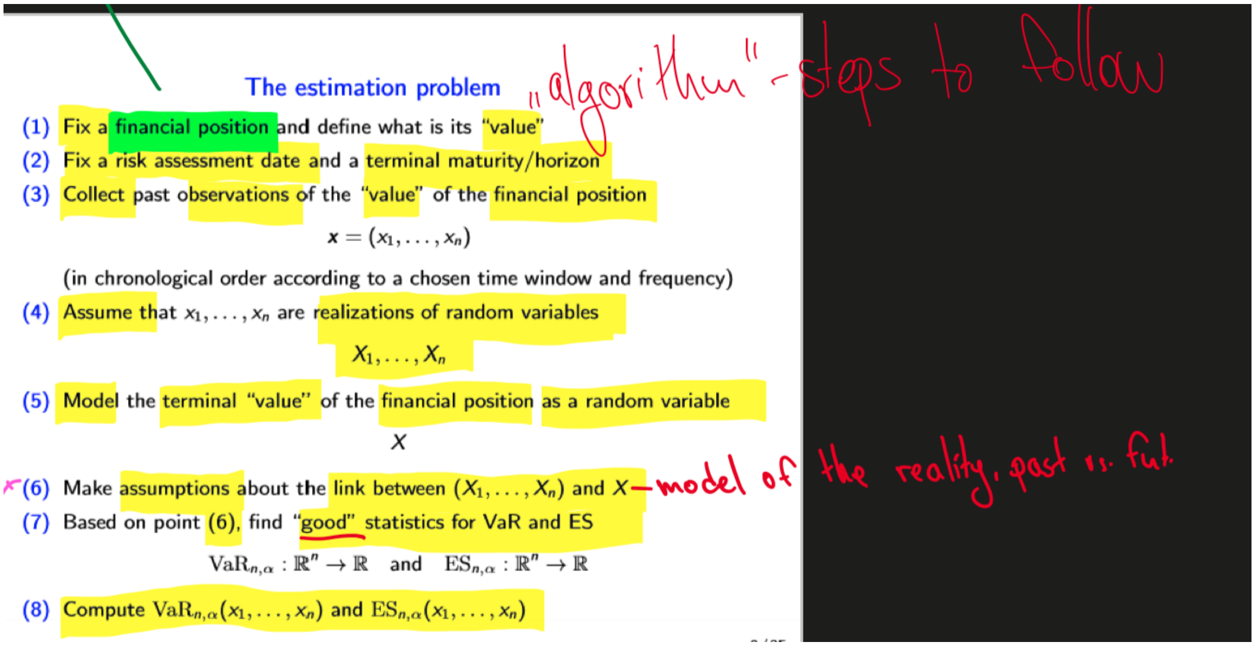


Figure 4: Plot title.

The estimation problem

(6) Make assumptions about the link

- **unconditional approach**: the random variables X1 .... Xn and X are assumed to be **independent and identically distributed**
- **conditional approach**: the random vars are assumed to be *dependent* and to display different distributions

(8) Choice of the sample size is crucial:

- **small sample size**: the RM statistics may be very volatile and hence encentivize too frequent portfolio rebalancing
- **large sample size**: the RM statistics may be too slow to react to losses or incorporate different volatility clusters

From CA to PM

we do not have data about capital position of a financial institution.

We consider

$$X = \text{portfolio's return}$$

Proposition. Let $\rho = \mathrm{VaR}_\alpha$ or $\rho = \mathrm{ES}_\alpha$ for some $\alpha \in (0,1)$.
$\rho\left(\frac{V_1 - V_0}{V_0}\right) = \frac{\rho(V_1 - V_0)}{V_0}$.
$\rho\left(\frac{V_1}{V_0}\right) = \frac{\rho(V_1 - V_0)}{V_0} - 1$.

Interpretation:

Quantity $\rho(V1 - V0)$ is the capital to be injected to ensure we have a profit V1>V0 at an acceptable level of confidence

Estimation setting

```
library ( zoo )
```

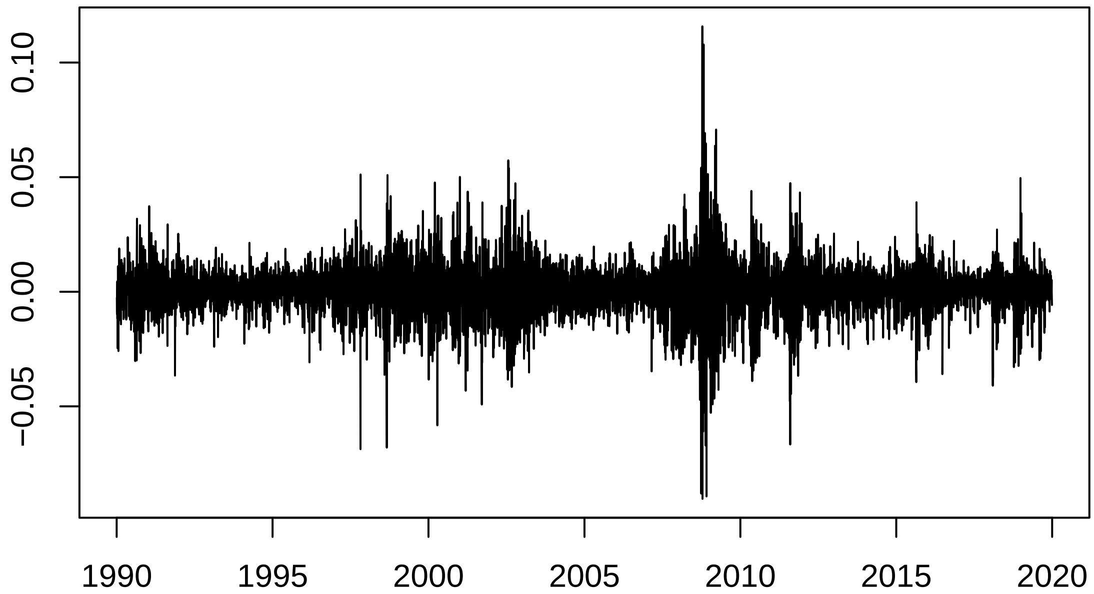
```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library ( pdfetch )
library (ggplot2)
data = pdfetch_YAHOO("^GSPC", fields="close", from="1990-01-01", to= "2019-12-31")
price = coredata ( data ) #very similar to xlsread, coredata takes only the
x = diff ( price )/ price [1:( length ( price ) -1)]


time = index ( data )
t = time [2: length ( time )]
plot (t ,x , type ="l" , xlab ="Relative returns" , ylab = "", main="S&P500 Relative Returns")
```

S&P500 Relative Returns



Relative returns

```
hVaR = function (x , alpha ){
n = length (x)
k = floor (n* alpha )
r = - sort (x )[ k +1]
return ( r)
}
compute_hVaR = function (x ,n , alpha ){
N = length (x)- n +1
r = rep (0 , N)
for (i in 1: N ){
r[i ] = hVaR (x[ i :( i+n -1)] , alpha )
}
return ( r)
}
# % x is the time series of our relative returns
hVarOutput = compute_hVaR (x ,250 ,0.01)
```

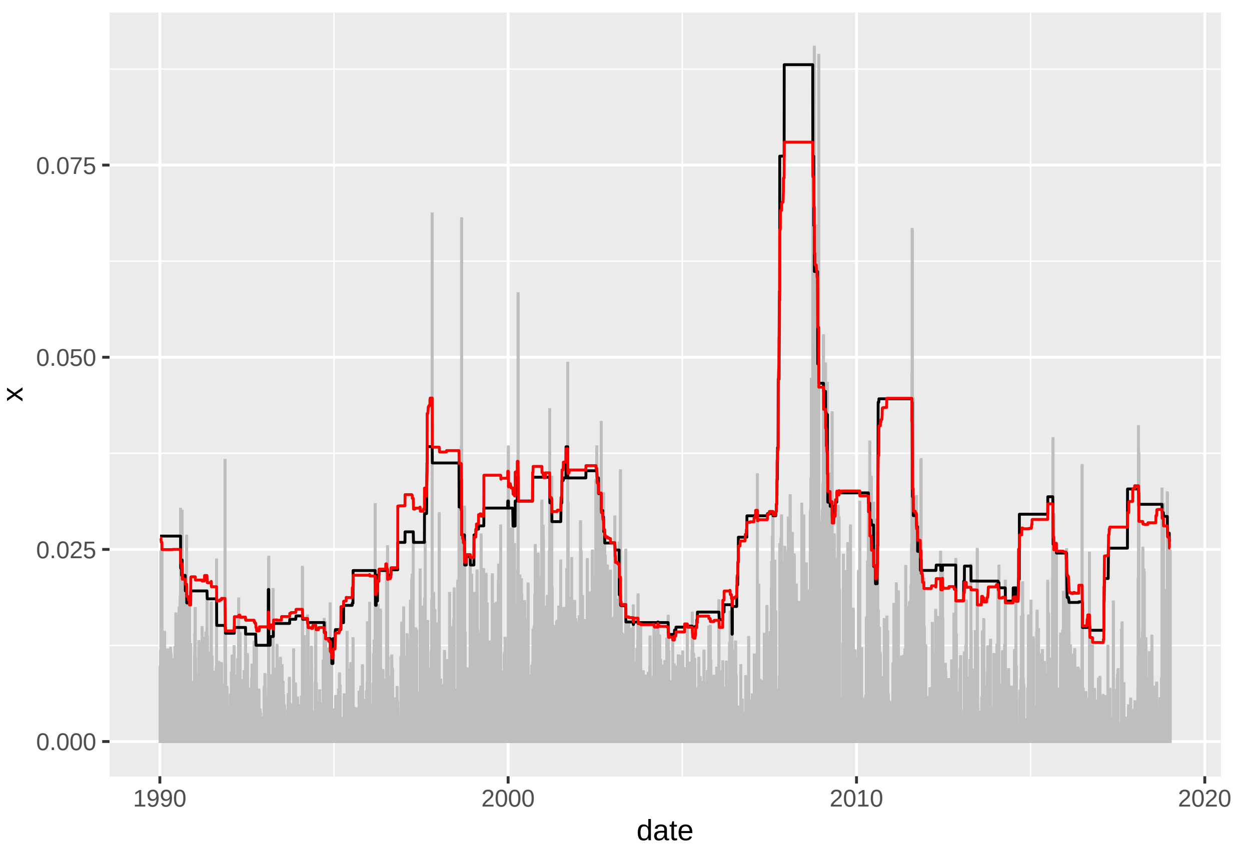
```
hES = function (x , alpha ){
n = length (x)
k = floor (n* alpha )
r = - sum ( sort ( x )[1: k ]) / (n* alpha )- sort ( x )[ k +1] *( alpha -k/n )/ alpha
return ( r)
}
compute_hES = function (x ,n , alpha ){
N = length (x)- n +1
r = rep (0 , N)
for (i in 1: N ){
r[i ] = hES (x[i :( i+n -1)] , alpha )
}
```

```r
return ( r)
}
#% x is the time series of ou
hESOutput = compute_hES(x ,250 ,0.025)

date = t[1:length(hESOutput)]

x = diff ( price )/ price [1:( length ( price ) -1)]
b = length(x)
for (i in 1 : b) {
  if (x[i] < 0) {
    x[i] = abs(x[i])}
    else {
      x[i] = 0
      }
    }
x = x[1:length(hESOutput)]

df = data.frame(date,hVarOutput,hESOutput,x)
ggplot(df, aes(date)) +
  geom_col(aes(y = x), color = "gray") +# basic graphical object
  geom_line(aes(y=hVarOutput), colour="black") +  # first layer
  geom_line(aes(y=hESOutput), colour="red") #second layer
```



Estimatoin of VaR and ES - The Weighted Historical Approach

The weighted sample distribution function

Definition. Let $x \in \mathbb{R}^n$ and consider a vector $\boldsymbol{\pi} = (\pi(x_1), \ldots, \pi(x_n)) \in \mathbb{R}^n$ satisfying: $\pi(x_1), \ldots, \pi(x_n) \in (0,1)$ $\sum_{i=1}^{n} \pi(x_i) = 1$ The weighted sample distribution function associated with x and π is the distribution function $F_{x,\pi} \in \mathcal{D}$ given by

$$F_{x,\pi}(x) := \sum_{i=1}^{n} \pi(x_i) \not\Vdash_{[x_i,\infty)}(x).$$

Interpretation

The sample distribution is a step-wise function constructed by **weighing past observations through the (probability) weights $\pi(x_{1\ldots n})$. It is a probability

For every $\alpha \in (0,1)$ we define the index

$$k(\alpha, \pi) := \min\left\{k \in \{1, \ldots, n\}; \sum_{i=1}^{k} \pi(x_{i:n}) > \alpha\right\} - 1.$$

Weighted historical estimators. The weighted historical VaR is given by

$$\mathrm{VaR}_{n,\alpha}(x_1, \ldots, x_n) = -x_{k(\alpha,\pi)+1:n}$$

The weighted historical ES is given by $\mathrm{ES}_{n,\alpha}(x_1, \ldots, x_n) = -\frac{1}{\alpha} \sum_{i=1}^{k(\alpha,\pi)} \pi(x_{i:n}) x_{i:n} - \frac{\alpha - \sum_{i=1}^{k(\alpha,\pi)} \pi(x_{i:n})}{\alpha} x_{k(\alpha,\pi)+1:n}$

Interpretation:

The weighted historical VaR and ES coincide with the VaR and ES of the weighted sample distribution function

Setting probability weights

```
pweights = function ( lambda ){
i = 1:250
pi = lambda ^(250 - i )* (1 - lambda )/ (1 - lambda ^250)
return(pi)
}
```

Compute historical weighted Var

```
wVaR = function (x , lambda , alpha ){
pi = pweights ( lambda )
k = min ( which ( cumsum ( pi [ order (x )]) > alpha )) -1
r = - sort (x )[ k +1]
return ( r)
}
```

```
hold4 = compute_wVaR = function (x ,n , lambda , alpha ){
N = length (x)- n +1
r = rep (0 , N)
for (i in 1: N ){
r[i ] = wVaR (x[ i :( i+n -1)] , lambda , alpha )
}
return ( r)
}
#% x is the time series of our relative returns
hold1 = compute_wVaR(x ,250 ,0.99 ,0.01)
hold2 = compute_wVaR(x ,250 ,0.98 ,0.01)
```

Historical weighted ES

```
wES = function (x , lambda , alpha ){
pi = pweights ( lambda )
q = pi [ order (x )]
k = min ( which ( cumsum (q) > alpha )) -1
r1 = (k ==0) * sort (x )[ k +1]
r2 = (k >0) *( sum ( sort (x )[1: k ]*q [1: k ])+ sort (x )[ k +1] * ( alpha - sum (q [1: k ]))))
r = -( r1 + r2 )/ alpha
return ( r)
}
holder3 = compute_wES = function (x ,n , lambda , alpha ){
N = length (x)- n +1
r = rep (0 , N)
for (i in 1: N ){
r[i ] = wES (x[i :( i+n -1)] , lambda , alpha )
}
return ( r)
}
#% x is the time series of our relative returns
holder = compute_wES(x ,250 ,0.99 ,0.025)
holder2 = compute_wES(x ,250 ,0.98 ,0.025)
```

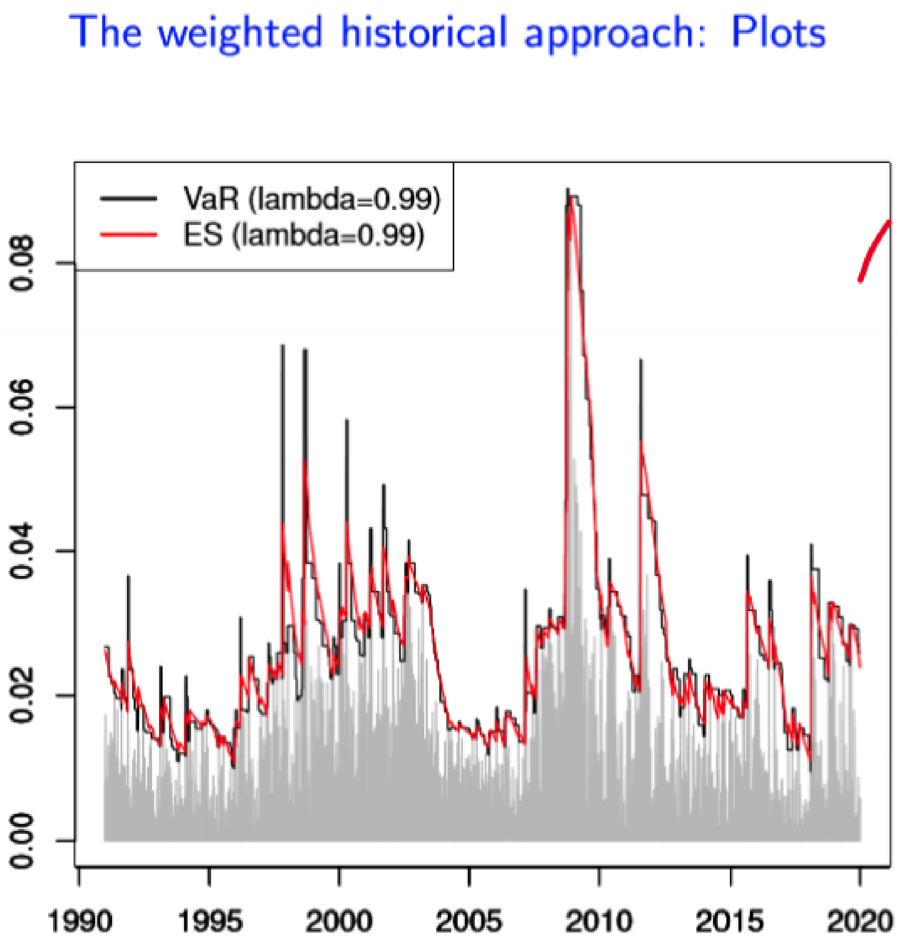
	mean	maximum	minimum	standard deviation
data	0.04%	11.58%	−9.03%	1.10%
VaR ($\lambda = 0.99$)	2.62%	9.03%	0.94%	1.34%
ES ($\lambda = 0.99$)	2.62%	8.93%	1.05%	1.27%
VaR ($\lambda = 0.98$)	2.54%	9.03%	0.76%	1.35%
ES ($\lambda = 0.98$)	2.51%	8.99%	0.81%	1.26%

Estimation of VaR and ES: The Risk Metrics Approach

VaR and ES under normality

Let Φ be the distribution function of a standard normal random variable, i.e.

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} \, \mathrm{d}y$$



got rid of plateaus

Figure 5: Plot title.

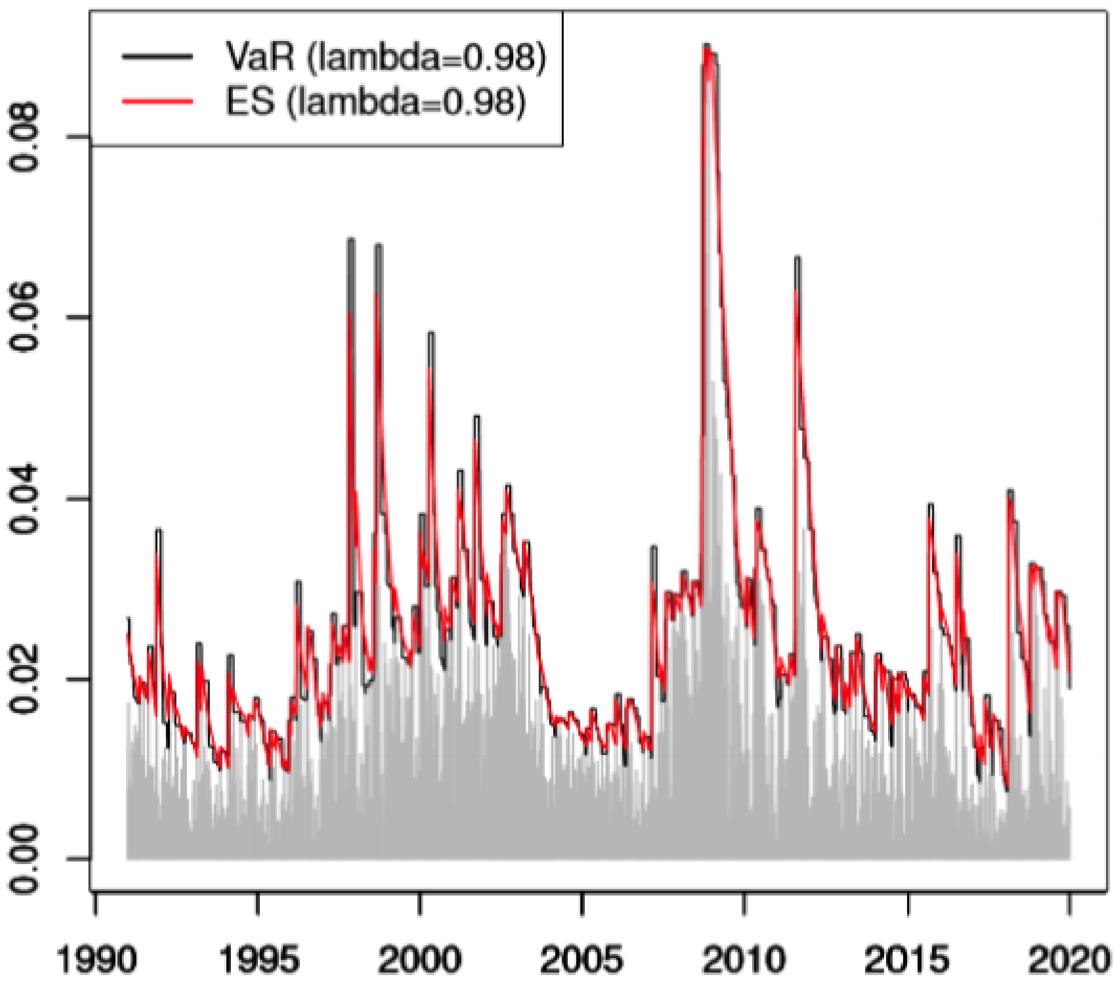


Figure 6: Plot title.

Proposition. Let $X \sim N\left(\mu, \sigma^2\right)$. Then, for every $\alpha \in (0,1)$ we have: - $\mathrm{VaR}_\alpha(X) = -\mu - \sigma\Phi^{-1}(\alpha)$ - $\mathrm{ES}_\alpha(X) = -\mu + \frac{\sigma}{\alpha}\Phi'\left(\Phi^{-1}(\alpha)\right)$

Estimating mean and variance

Definition. For a given $\lambda > 0$, the EWMA variance is defined by

$$V_n^{EWMA}\left(x_1, \ldots, x_n\right) := (1-\lambda)\sum_{i=1}^{n} \lambda^{n-i}\left(x_i - M_n\left(x_1, \ldots, x_n\right)\right)^2$$

or recursively by

$$\begin{cases} V_1^{EWMA}\left(x_1\right) = (1-\lambda)\left(x_1 - M_n\left(x_1, \ldots, x_n\right)\right)^2 \\ V_2^{EWMA}\left(x_1, x_2\right) = \lambda V_1^{EWMA}\left(x_1\right) + (1-\lambda)\left(x_2 - M_n\left(x_1, \ldots, x_n\right)\right)^2 \\ \ldots \\ V_n^{EWMA}\left(x_1, \ldots, x_n\right) = \lambda V_{n-1}^{EWMA}\left(x_1, \ldots, x_{n-1}\right) + (1-\lambda)\left(x_n - M_n\left(x_1, \ldots, x_n\right)\right)^2 \end{cases}$$

The EWMA standard deviation is given by

$$SD_n^{EWMA}\left(x_1, \ldots, x_n\right) := \sqrt{V_n^{EWMA}\left(x_1, \ldots, x_n\right)}$$

Interpretation. The acronym EWMA stands for exponential weighted moving average. This is akin to the variance in an (I)GARCH model (see Lecture 10).

We define variance by (as above)

$$V_n^{EWMA}\left(x_1, \ldots, x_n\right) := (1-\lambda)\sum_{i=1}^{n} \lambda^{n-i}\left(x_i - M_n\left(x_1, \ldots, x_n\right)\right)^2$$

Lambda is usually .99 or .98. Interpretation: the older the observation is (lower i), the less weight it gives into the EWMA (given by λ^{n-i}) and thus receives less importance

Setting EWMA weights

```
EWMAweights = function(lambda){ i = 1:250
pi = (1-lambda)*lambda^(250-i)
return(pi)
}
```

Computing VaR in Risk Metrics Approach

```
rmVaR = function(x,lambda,alpha){
n = length(x)
m = sum(x)/n
sd = sqrt(sum(EWMAweights(lambda)*(x-m)^2))
q = qnorm(alpha ,0,1)
r = -m-sd*q
return(r)
}
```

Computing ES in Risk Metrics Approach

```
compute_rmVaR = function(x,n,lambda,alpha){ N = length(x)-n+1
r = rep(0,N)
for (i in 1:N){
r[i] = rmVaR(x[i:(i+n-1)],lambda ,alpha)
}
return(r)
}
# x is the time series of our relative returns compute_rmVaR(x,250,0.99,0.01) compute_rmVaR(x,250,0.97,
```

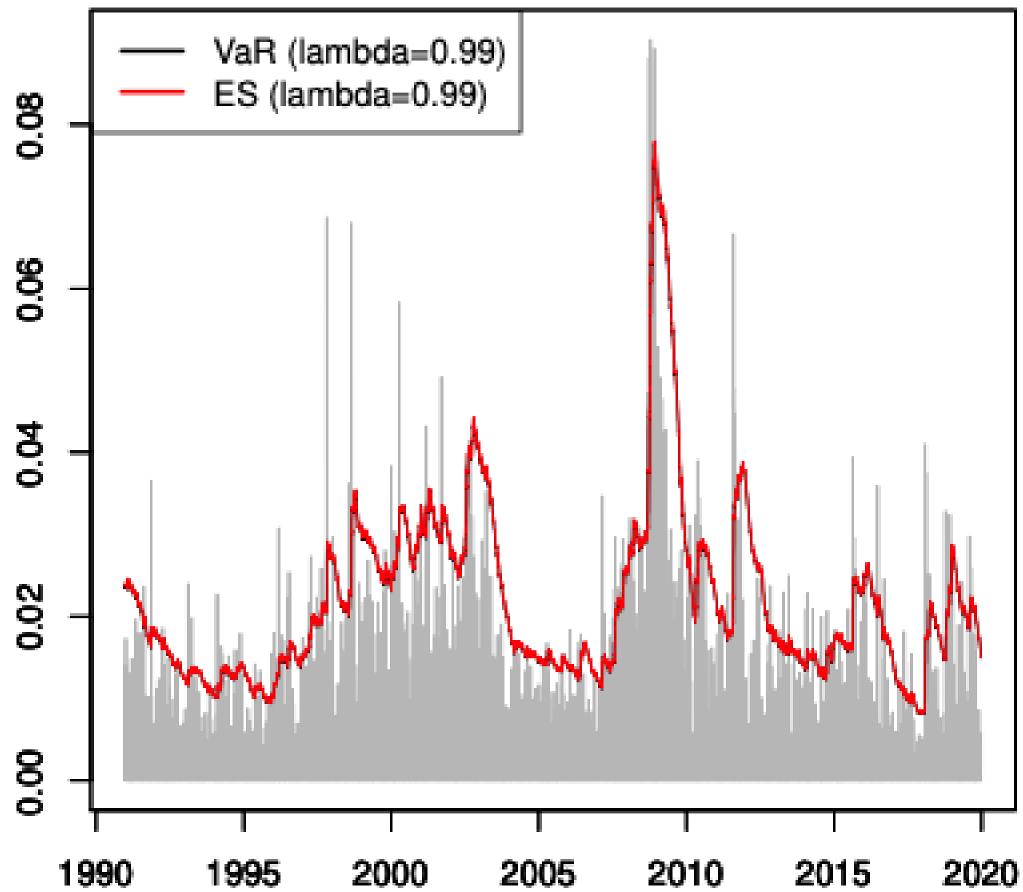


Figure 7: Plot title.

Estimation of VaR and ES: Simulation Approach

Generating distribution using quantiles

Proposition. Let $U \sim U(0,1)$. For every $F \in \mathcal{D}$ define the random variable

$$q_U^+(F) := \inf\{x \in \mathbb{R}; F(x) > U\} \overset{F \text{ invertible}}{=} F^{-1}(U)$$

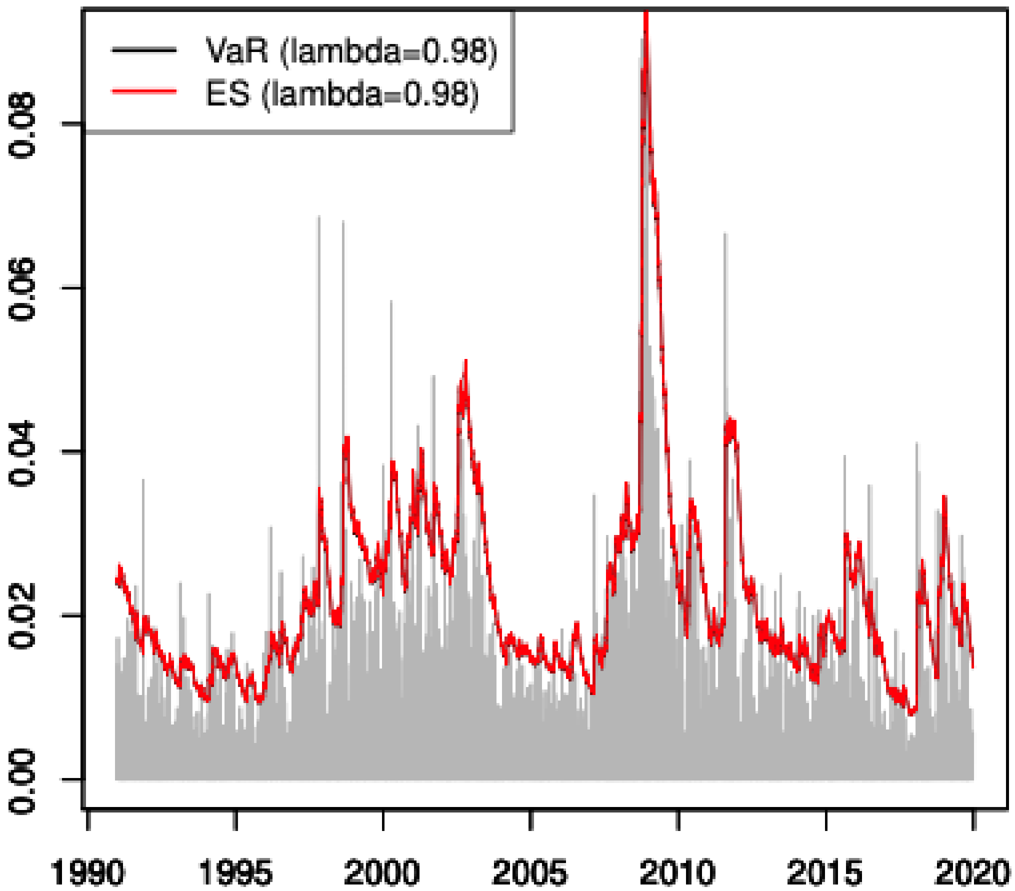


Figure 8: Plot title.

	mean	maximum	minimum	standard deviation
data	0.04%	11.58%	-9.03%	1.10%
VaR ($\lambda = 0.99$)	2.19%	7.75%	0.82%	1.08%
ES ($\lambda = 0.99$)	2.20%	7.79%	0.82%	1.09%
VaR ($\lambda = 0.98$)	2.25%	9.51%	0.79%	1.18%
ES ($\lambda = 0.98$)	2.26%	9.55%	0.79%	1.19%

Figure 9: Plot title.

Then, we have

$$q_U^+(F) \sim F$$

This process is called **“quantile inversion”**

Idea: By taking (upper) quantile of a DF F to a uniform random variable, we obtain a random variable with the same distribution F.

That shows that if we want to simulate realizations of a random variable with distribution F, it suffices to apply the upper quantile of F to simulated realizations of a uniform random variable.

The simulation approach

The idea is that if we have a set of reutrns Z, even if we konw their distribution, it’s no trivial to know distribution of the whole set X. That is why we combine them with *f*, the DF from before.

Algorithm

The simulation approach is based on the following steps: - Let $h \in \mathbb{N}$ (number of iterations) and $s \in \mathbb{N}$ (number of simulations per iteration). - For every $k \in \{1, \ldots, h\}$ let $u_1^k, \ldots, u_{ms}^k$ be realizations of independent uniform random variables on $(0, 1)$. - For every $j \in \{1, \ldots, m\}$ and $i \in \{1, \ldots, s\}$ set

$$z_i^{k,j} = q^+_{u^k_{j-1)s+i}}\left(Z^j\right)$$

- For every $i \in \{1, \ldots, s\}$ set

$$x_i^k = f\left(z_i^{k,1}, \ldots, z_i^{k,m}\right).$$

Remark. Note that we have to ensure independence also across the different iterations.

Then we perform VaR and ES for each iteration h and average across all iterations.

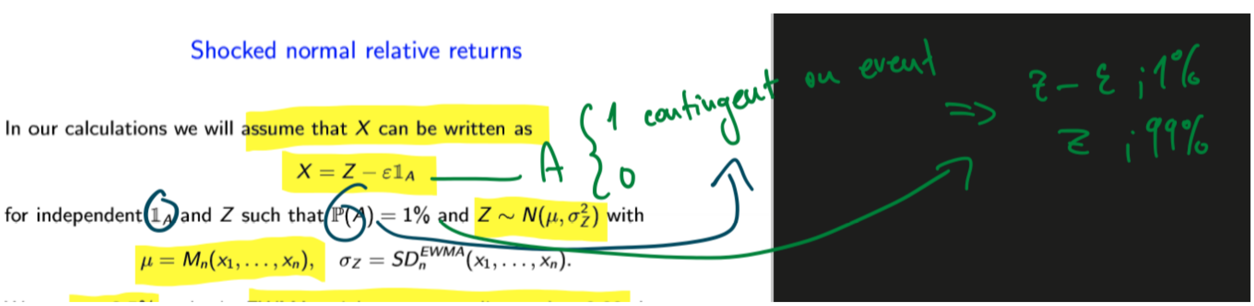


Figure 10: Plot title.

Shocked normal relative returns We set $\varepsilon = 2.5$ and take EWMA weights corresponding to $\lambda = .99$.

Interpretation: We assume that X follows a normal distr. subject to external shocks with probability 1%. The shock lowers relative returns by *2.5%*. Note that X is not normal.

Simulation VaR

```r
simVaR = function(x,h,s,lambda,alpha){
m = sum(x)/length(x)
sd = sqrt(sum(EWMAweights(lambda)*(x-m)^2))
hr = rep(0,h)
u = runif(2*s*h)
sim_z = qnorm(u[1:(s*h)],m,sd)
sim_ind = qbinom(u[(s*h+1):(2*s*h)],1,0.01)

for (k in 1:h){
z = sim_z[((k-1)*s+1):(k*s)]
ind = sim_ind[((k-1)*s+1):(k*s)]
x = z-0.025*ind
hr[k] = hVaR(x,alpha)
}
r = sum(hr)/h

return(r)
}
compute_simVaR = function(x,n,h,s,lambda,alpha){ N = length(x)-n+1
r = rep(0,N)
for (i in 1:N){
r[i] = simVaR(x[i:(i+n-1)],h,s,lambda ,alpha)
}
return(r)
} # x is the time series of our relative returns compute_simVaR(x,250,1,500,0.99,0.01)
 simulatedVaR = compute_simVaR(x,250,10,500,0.99,0.01)
```

```r
simES = function(x,h,s,lambda,alpha){
m = sum(x)/length(x)
sd = sqrt(sum(EWMAweights(lambda)*(x-m)^2))
hr = rep(0,h)
u = runif(2*s*h)
sim_z = qnorm(u[1:(s*h)],m,sd)
sim_ind = qbinom(u[(s*h+1):(2*s*h)],1,0.01)
for (k in 1:h){
z = sim_z[((k-1)*s+1):(k*s)]
ind = sim_ind[((k-1)*s+1):(k*s)]
x = z-0.025*ind
hr[k] = hES(x,alpha)
}
r = sum(hr)/h
return(r)
}
compute_simES = function(x,n,h,s,lambda,alpha){ N = length(x)-n+1
r = rep(0,N)
for (i in 1:N){
r[i] = simES(x[i:(i+n-1)],h,s,lambda ,alpha)
}
return(r)
}   #% x is the time series of our relative returns
simES1 = compute_simES(x,250,1,500,0.99,0.025)
simES2 = compute_simES(x,250,10,500,0.99,0.025)
```

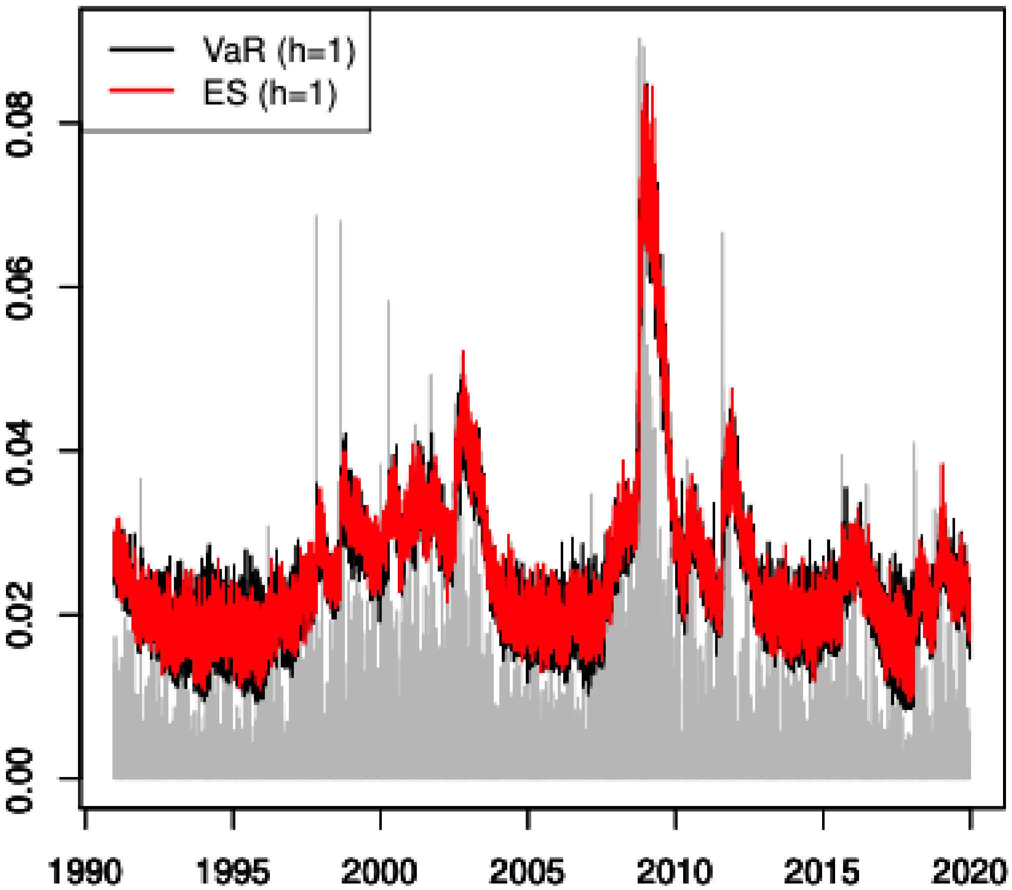


Figure 11: Plot title.

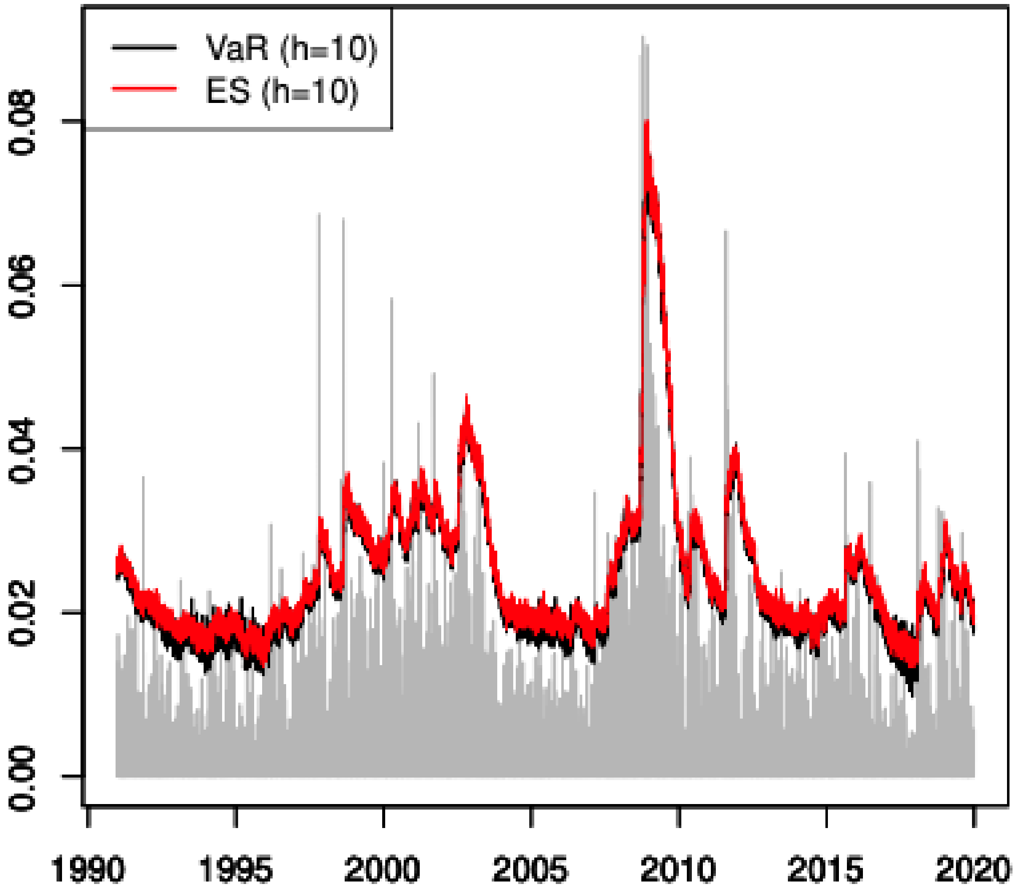


Figure 12: Plot title.

	mean	maximum	minimum	standard deviation
data	0.04%	11.58%	-9.03%	1.10%
VaR ($h = 1$)	2.44%	8.49%	0.83%	1.00%
ES ($h = 1$)	2.53%	8.46%	0.93%	0.97%
VaR ($h = 10$)	2.44%	7.91%	0.97%	0.97%
ES ($h = 10$)	2.53%	8.01%	1.34%	0.96%

Figure 13: Plot title.

Estimation VaR and ES - Conditional Approach

So far, we assumed i.i.d. Empirically we know wolatility clusters, so no good.

Filtrations

Definition. An (n -dimensional) filtration is any sequence of σ -fields

$$\mathcal{F} = (\mathcal{F}_1, \dots, \mathcal{F}_n)$$

such that $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$. We always assume that $\mathcal{F}_n \subset \mathcal{F}$.

Interpretation: filtration models the *flow of information* - at each point in time we get new information and we want to discard some outcomes as impossible.

Definition. Let $X \in \mathcal{X}^n$ and consider an n -dimensional filtration $\mathcal{F}$. We say that X is $\mathcal{F}$ -adapted whenever X_i is $\mathcal{F}_i$ -measurable for every $i \in \{1, \dots, n\}$.

The general time series model

Setup. Let $t \in \mathbb{Z}$ and take a filtration $(\mathcal{F}_t)$. Set $X_{n+1} = X$ and assume that $X_1, \dots, X_n, X_{n+1}$ are part of an $(\mathcal{F}_t)$ -adapted process (X_t) such that

$$X_t = \mu_t + \sigma_t Z_t$$

where μ_t and σ_t are $\mathcal{F}_{t-1-\text{ measurable and }} U_t = \sigma_t \overline{Z_t}$ is the innovation term. We stipulate the following assumptions: - the elements of (Z_t) are independent - the elements of (Z_t) are identically distributed as Z - $\mathbb{E}[Z] = 0$ and $\text{var}(Z) = 1$ - σ_t is strictly positive (almost surely) Remark: The time series (Z_t), with standardized i.i.d. elements and zero unconditional mean, is also known in time series analysis as **"white noise"** - that contains no information

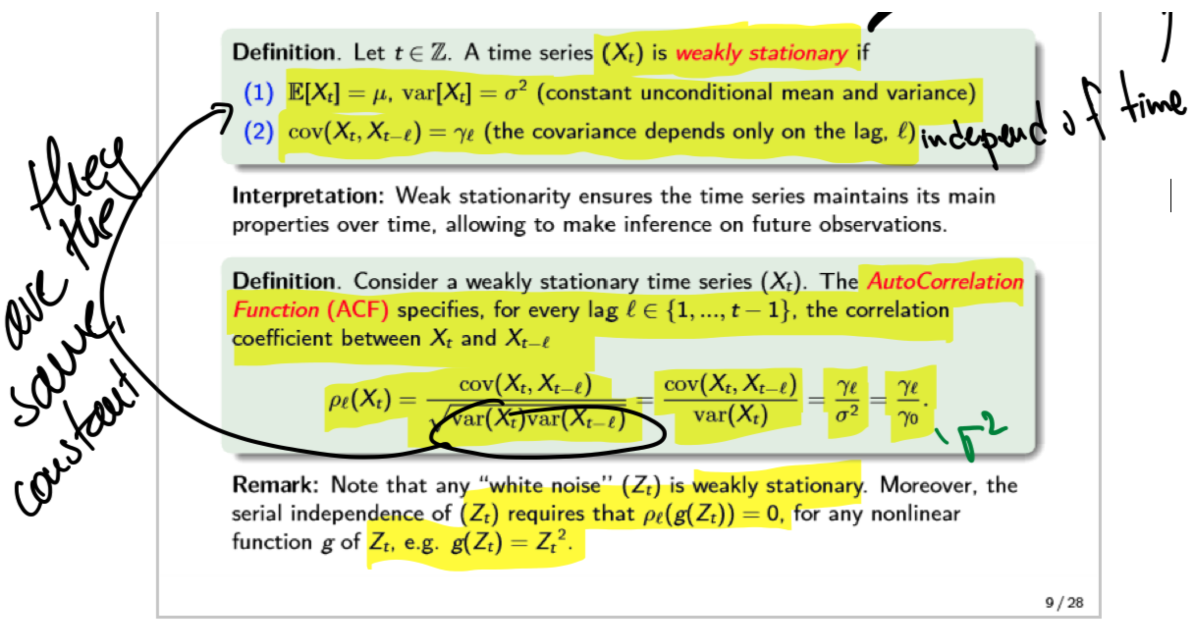


Figure 14: Plot title.

Maximum-Likelihood estimation

Definition. Let $t \in \mathbb{Z}$ and take a filtration $(\mathcal{F}_t)$. Consider an $(\mathcal{F}_t)$ -adapted process (X_t) with joint density function $f_\theta(x_1, \ldots, x_t)$, with θ an unknown $1 \times p$ parameter vector. The likelihood function and log-likelihood function are defined respectively as

$$
\begin{aligned}
L(\theta) &= f_\theta(X_1, \ldots, X_t) \\
\ell(\theta) &= \log(f_\theta(X_1, \ldots, X_t))
\end{aligned}
$$

The maximum likelihood estimator (MLE) θ^{MLE} of θ is the maximizer of $\ell(\theta)$.

Conditional VaR and ES

Definition. Let $\alpha \in (0,1)$. The conditional VaR at time t is

$$\mathrm{VaR}_\alpha(X_{t+1}|\mathcal{F}_t) := -\inf\{x \in \mathbb{R}\,;\ \mathbb{P}(X_{t+1} \leq x|\mathcal{F}_t) > \alpha\}.$$

Similarly, the conditional ES at time t is given by

$$\mathrm{ES}_\alpha(X_{t+1}|\mathcal{F}_t) := \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_p(X_{t+1}|\mathcal{F}_t)\mathrm{d}p.$$

Proposition. For every $\alpha \in (0,1)$ the following statements hold:

- $\mathrm{VaR}_\alpha(X_{t+1}|\mathcal{F}_t) = -\mu_{t+1} - \sigma_{t+1}\mathrm{VaR}_\alpha(Z)$.
- $\mathrm{ES}_\alpha(X_{t+1}|\mathcal{F}_t) = -\mu_{t+1} + \sigma_{t+1}\mathrm{ES}_\alpha(Z)$.

Remark. To compute $\mathrm{VaR}_\alpha(X_{t+1}|\mathcal{F}_n)$ and $\mathrm{ES}_\alpha(X_{t+1}|\mathcal{F}_n)$ we have to estimate μ_{t+1} and σ_{t+1} as well as $\mathrm{VaR}_\alpha(Z)$ or $\mathrm{ES}_\alpha(Z)$.

Figure 15: Plot title.

Modeling AR, MA and ARMA models

Choosing the model

Modelling conditional heteroskedastic models

$$\operatorname{var}(X_{t+1} \mid \mathcal{F}_t) = \operatorname{var}(X_{t+1}) = \sigma^2$$

It conditional and unconditional variance is the same, it's called *conditional homoskedasticity*.

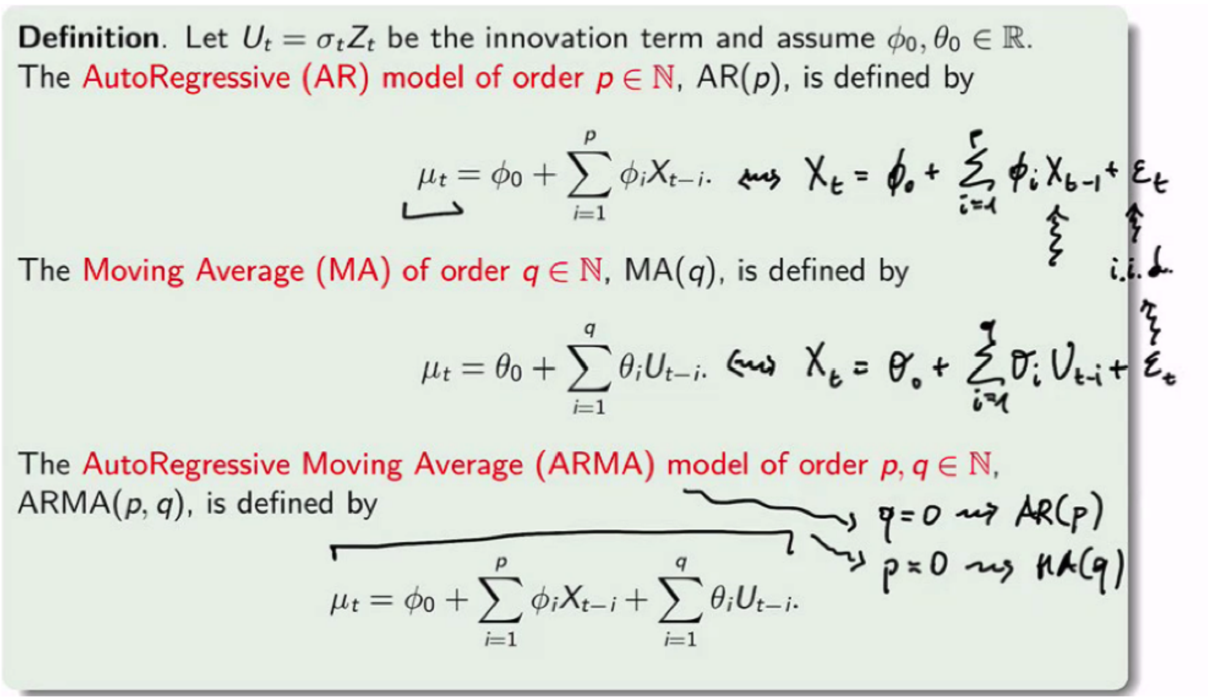


Figure 16: Plot title.

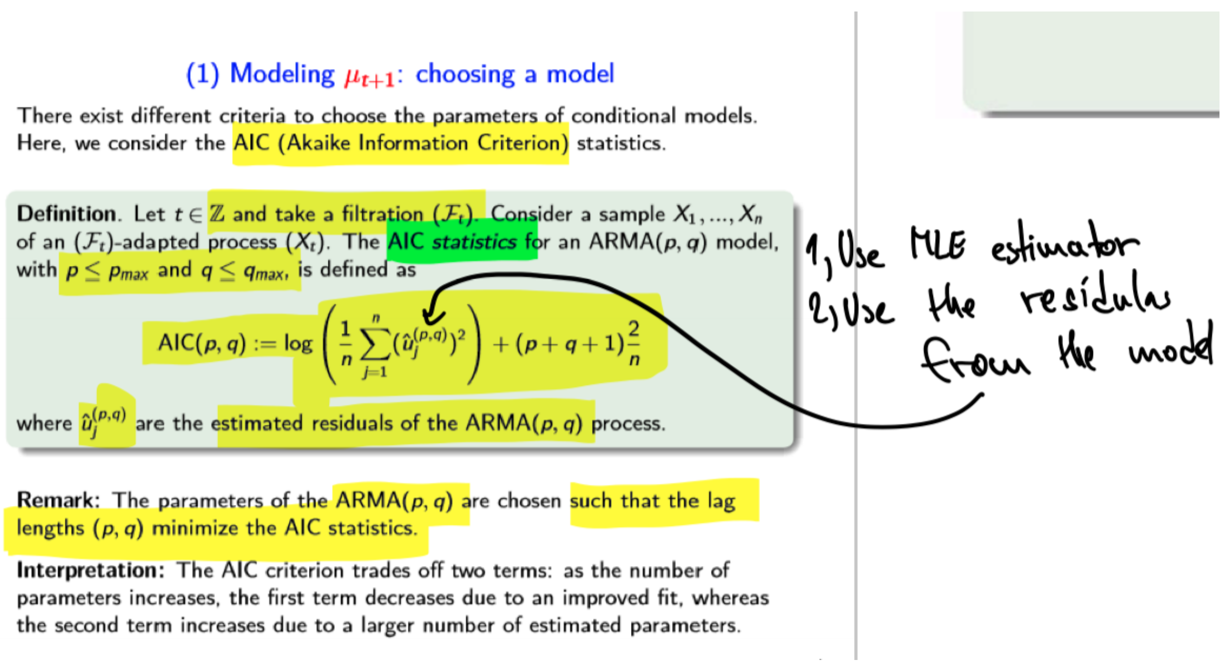


Figure 17: Plot title.

the elements of (X_t) are serially uncorrelated but dependent, i.e. if we emove the 2nd "i" in i.i.d., their conditional variance is not constant anymore

$$\operatorname{var}\left(X_{t+1} \mid \mathcal{F}_t\right)=\operatorname{var}\left(U_{t+1} \mid \mathcal{F}_t\right)=\mathbb{E}\left(U_{t+1}^2 \mid \mathcal{F}_t\right)=\sigma_{t+1}^2$$

This property is called conditional heteroskedasticity: the dependence across X_t nakes the flow of information relevant to determine their variance, which varies ser time as new information becomes available.

Risk Factor Mapping

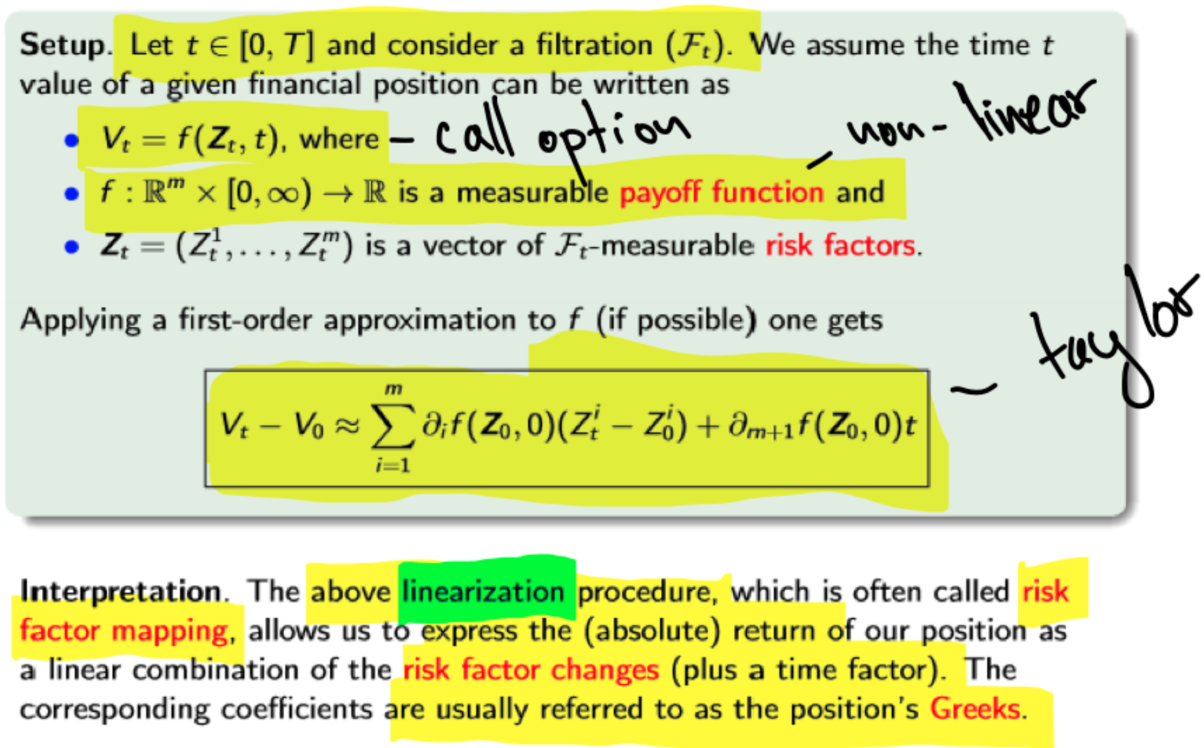


Figure 18: Plot title.

Linearization is regression

Delta VaR and ES

Definition. Let $\alpha \in (0,1)$ and set $X_t = \frac{V_t - V_0}{V_0}$. The Delta VaR is given by

$$\operatorname{VaR}_\alpha^\Delta\left(X_t\right):=\frac{1}{V_0} \operatorname{VaR}_\alpha\left(\sum_{i=1}^m \partial_i f\left(Z_0, 0\right)\left(Z_t^i-Z_0^i\right)+\partial_{m+1} f\left(Z_0, 0\right) t\right)$$

Similarly, the Delta ES is given by

$$\operatorname{ES}_\alpha^\Delta\left(X_t\right):=\frac{1}{V_0} \operatorname{ES}_\alpha\left(\sum_{i=1}^m \partial_i f\left(Z_0, 0\right)\left(Z_t^i-Z_0^i\right)+\partial_{m+1} f\left(Z_0, 0\right) t\right)$$

Interpretaion: Delta approach is useful when risk factors can be estimated - we have a lot of data. Then we believe payoff of our position can be safely approximated in a linear way.

Delta VaR and ES under normal risk factors Proposition. For $m \in \mathbb{N}$ let $\mu \in \mathbb{R}^m$ and $\sigma \in \mathbb{R}^{m \times m}$ be an invertible positive definite (symmetric) matrix. Then, for every $\boldsymbol{X} = (X_1, \ldots, X_m) \in \mathcal{X}^m$ the following statements are equivalent: - $\sum_{i=1}^m a_i X_i \sim N\left(\sum_{i=1}^m a_i, \sum_{i,j=1}^m a_i a_j \sigma_{ij}\right)$ for every $a_1, \ldots, a_m \in \mathbb{R}$ - $\boldsymbol{X} \sim N^m(\boldsymbol{\mu}, \boldsymbol{\sigma})$

Proposition. Let $\mu \in \mathbb{R}^m$ and $\sigma \in \mathbb{R}^{m \times m}$ be an invertible positive definite (symmetric) matrix. Assume that $Z_t - Z_0 \sim N^m(\mu, \sigma)$. Then, we have:

- $\mathrm{VaR}_\alpha^\Delta(X_t) = -\frac{1}{V_0}\left(\sum_{i=1}^m \partial_i f(Z_0, 0)\mu_i + \partial_{m+1} f(Z_0, 0) t\right) - \frac{1}{V_0}\sqrt{\sum_{i,j=1}^m \partial_i f(Z_0, 0)\partial_j f(Z_0, 0)\sigma_{ij}}\phi^{-1}(\alpha)$
- $\mathrm{ES}_\alpha^\Delta(X_t) = -\frac{1}{V_0}\left(\sum_{i=1}^m \partial_i f(Z_0, 0)\mu_i + \partial_{m+1} f(Z_0, 0) t\right) + \frac{1}{V_0}\sqrt{\sum_{i,j=1}^m \partial_i f(Z_0, 0)\partial_j f(Z_0, 0)\sigma_{ij}}\frac{\Phi'\left(\Phi^{-1}(\alpha)\right)}{\alpha}$

Estimating covariance

Definition. For a given $\lambda > 0$, the EWMA covariance is defined by $COV_n^{EWMA}(x, y) := (1-\lambda)\sum_{i=1}^n \lambda^{n-i}\left(x_i - M_n(x_1, \ldots, x_n)\right)\left(y_i - M_n(y_1, \ldots, y_n)\right)$

Risk Metrics estimators. The Risk Metrics VaR estimator is given by

$$\mathrm{VaR}_{\alpha,n} := -\frac{1}{v_n}\left(\sum_{i=1}^m \partial_i f(z_n, 0) M_n\left(z^i\right) + \partial_{m+1} f(z_n, 0)\right) - \frac{1}{v_n}\sqrt{\sum_{i,j=1}^m \partial_i f(z_n, 0)\partial_j f(z_n, 0)\operatorname{cov}_n^{EWMA}\left(z^i, z^j\right)\Phi^{-1}(\alpha)}$$

The Risk Metrics ES estimator is given by

$$\mathrm{ES}_{\alpha,n} := -\frac{1}{v_n}\left(\sum_{i=1}^m \partial_i f(z_n, 0) M_n\left(z^i\right) + \partial_{m+1} f(z_n, 0)\right) + \frac{1}{v_n}\sqrt{\sum_{i,j=1}^m \partial_i f(z_n, 0)\partial_j f(z_n, 0)\operatorname{cov}_n^{EWMA}\left(z^i, z^j\right)}\frac{\phi'\left(\Phi^{-1}(\alpha)\right)}{\alpha}$$

Interpretation: above esimators coincide with the Delta VaR and ES computed under the assumption that the risk factor changes are jointly normal

Time conversoin rules

- if we want 1- day risk measure, we use daily data
- if we want 5-day RM, we use weekly

Warning 1: if chose horizon is too far, we have sparse data set. By doing so - We hide loss peaks between consecutive data points - include data points not include

Warning2: If we have a multi-day risk measure using a rolling window, we have to roll the window according to the chosen horizon

Let $\alpha \in (0,1)$ and $0 < s < t \leq T$. In this case $\mathrm{VaR}_\alpha(X_s)$ is safer to estimate - $\mathrm{VaR}_\alpha(X_t)$ is trickier to estimate We ask whether we can express $\mathrm{VaR}_\alpha(X_t)$ in terms of $\mathrm{VaR}_\alpha(X_s)$ as in

$$\mathrm{VaR}_\alpha(X_t) = f\left(\mathrm{VaR}_\alpha(X_s), s, t\right)$$

In this case, we could rely on the estimation of $\mathrm{VaR}_\alpha(X_s)$ and derive the estimates of $\mathrm{VaR}_\alpha(X_t)$ by exploiting the above link.

Easy interpretation: if we want longer horizons, because of the time thing, we esimate a shorter, i.e. safer VaR and then transform it to longer horizon with the distribution. In practice, it is done by the square-root rule

Square root rule Done in practice but impossible to prove mathematically.

The square-root rule. Let $\alpha \in (0,1)$ and $0 < s < t \leq T$. The square-root rule for VaR postulates that

$$\mathrm{VaR}_\alpha(X_t) = \sqrt{\frac{t}{s}}\,\mathrm{VaR}_\alpha(X_s)$$

The square-root rule for ES postulates that

$$\mathrm{ES}_\alpha(X_t) = \sqrt{\frac{t}{s}}\mathrm{ES}_\alpha(X_s)$$

Overview

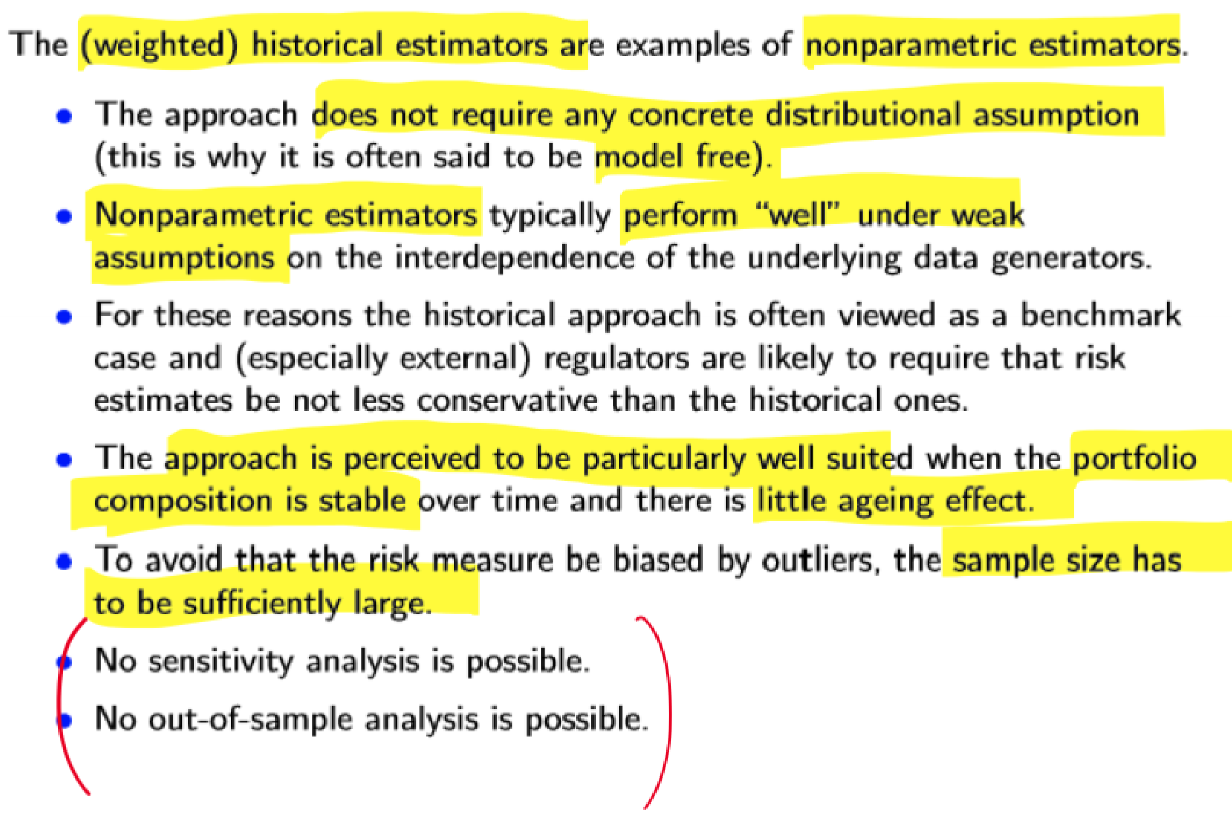


Figure 19: **Nonparametric**

Backtesting under VaR and ES

The term **backtesting** is a *process of assesing if ex-ante estimation are aligned with ex-post observations.*

Model selection

Setup. Let $\rho : \mathcal{X} \to [-\infty, \infty]$ be a law-invariant risk measure. To perform model selection using ρ one can do the following: - Collect past observations $x_1, \ldots, x_n$. - Use $x_1, \ldots, x_n$ to come up with competing models for F_X, say

$$\widehat{F}^{(1)}_{(x_1,\ldots,x_n)}, \ldots, \widehat{F}^{(m)}_{(x_1,\ldots,x_n)}.$$

The Risk Metrics estimator is an example of a parametric estimator.

- The approach relies on concrete distributional assumptions that...
- typically lead to simple and easy-to-communicate parametric formulas.
- The sample size has to be sufficiently large to ensure a good parameter estimation.
- The approach is perceived to be particularly well suited when the portfolio composition is linear and stable over time.
- In the presence of nonlinear instruments one typically relies on linearization.

derivs.

- The easiest analytical formulas typically require strong distributional assumptions (e.g. normality).
- Sensitivity analysis is possible (within the chosen distribution class).
- Out-of-sample analysis is possible.

Figure 20: **Parametric**

The simulation-based estimator is an example of a Monte Carlo estimator.

- The approach relies on concrete distributional assumptions.
- Because of simulation there is no problem of small sample size.
- The approach is perceived to be particularly well suited when the portfolio composition is neither linear nor stable over time.
- The approach typically requires high computational costs (especially in the presence of derivative instruments whose price cannot be determined in closed form).
- The approach is sometimes perceived as a black box whose outcomes are difficult to understand for external users/stakeholders.
- Sensitivity analysis is possible (within the chosen distribution class).
- Out-of-sample analysis is possible.

Figure 21: **Monte Carlo**

The estimators in the setting of ARMA-GARCH models are examples of conditional estimators.

- The approach relies on concrete distributional assumptions.
- The approach requires familiarity with time series models.
- Risk estimates are typically sensitive to volatility clusters.
- Risk estimates are sometimes perceived to be too sensitive.
- Sensitivity analysis is possible (within the chosen distribution class).
- Out-of-sample analysis is possible.

Figure 22: **Conditional estimators**

- Risk factor mapping allows to approximate a complex payoff structure by a linear combination of risk factor changes.
- In many circumstances the (joint) distribution of the underlying risk factors can be safely estimated.
- This simplifies the problem of computing a risk measure for the overall position and is particularly useful for financial positions containing
  - a large number of different instruments
  - instruments with nonlinear payoffs
- The quality of the linear approximation is better on short time horizons and for risk factors showing relatively small changes.
- However, the presence of highly nonlinear payoff profiles may make the approximation quite gross even if we have a short time horizon and small risk factor changes.

Figure 23: **Risk factor mapping**

- Compute the risk estimates $\rho\left(\widehat{F}^{(1)}_{(x_1,\ldots,x_n)}\right),\ldots,\rho\left(\widehat{F}^{(m)}_{(x_1,\ldots,x_6)}\right)$. - Assume that X materializes into x. - We want to use x to select the best model out of $\hat{F}^{(1)}_{(x_1,\ldots,x_n)},\ldots,\hat{F}^{(m)}_{(x_1,\ldots,x_n)}$-

By doing backtest for all models, we look which performs the best. More specifically, we are selecting our model **based on the performance of the risk measure ρ .

Question: How to decide which model is better?

Model validation

Setup. Let $\rho : \mathcal{X} \to [-\infty, \infty]$ be a law-invariant risk measure. To perform model validation using ρ one can do the following: - Collect past observations $x_1,\ldots,x_n$ - Use $x_1,\ldots,x_n$ to come up with a model, say $\widehat{F}_{(x_1,\ldots,x_n)}$, for F_X. - Compute the risk estimate $\rho\left(\widehat{F}_{(x_1,\ldots,x_n)}\right)$. - Assume that X materializes into x. - We want to use x to validate the choice of $\widehat{F}_{(x_1,\ldots,x_n)}$ or not.

We compare ex-post observations with ex-ante model forecasts to assess whether our model assumption should be kept or not. This is **model valiatoin**.

Question: How to decide whether the model is validaated or not?

Elicitability

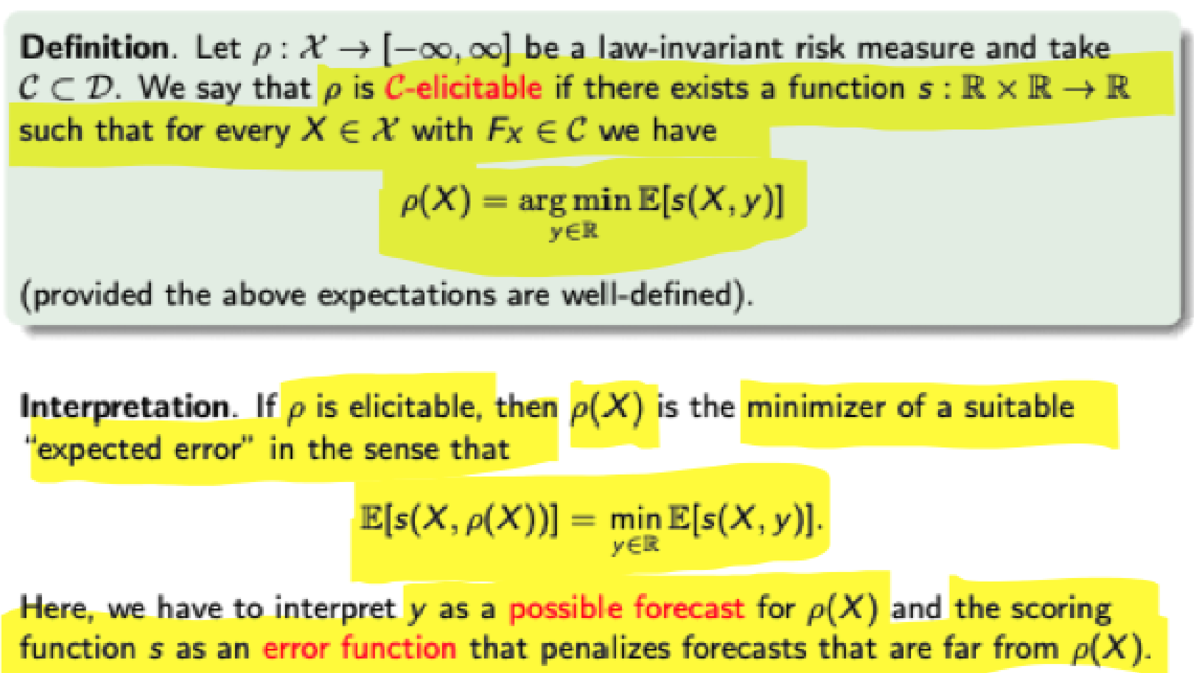


Figure 24: Plot title.

Elicitability under VaR Lemma. Let $\alpha \in (0,1)$. For every $X \in \mathcal{X}$ we have

$$\mathrm{ES}_\alpha(X) = \frac{1}{\alpha}\inf_{q\in\mathbb{R}}\{\mathbb{E}[\max(q-X,0)] - \alpha q\}$$

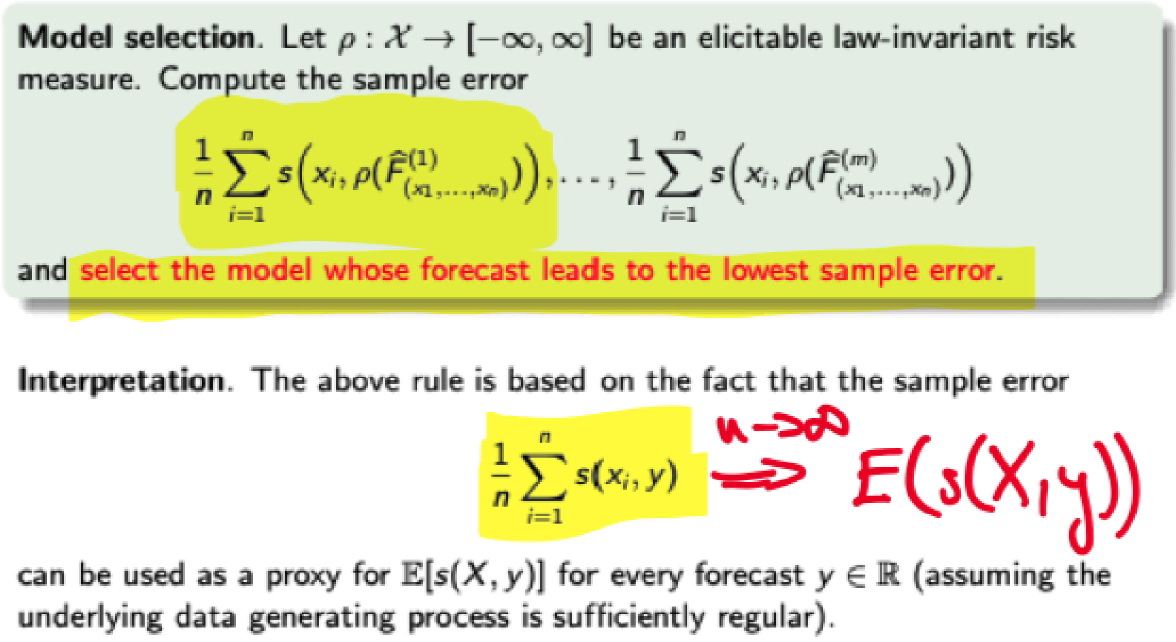


Figure 25: Plot title.

Moreover, the infimum is attained at any α -quantile of X.

** Intepreation**: VaR is the minimizer and ES is the results. That also leads to the fact that ES is not elicitable.

```
compute_scoreVaR = function(x,risk,n,alpha){ N = length(x)-n+1
s = rep(0,N)
for (i in 1:N){
s[i] = mean(risk[i]-(risk[i]+x[i:(i+n-1)])*(-x[i:(i+n-1)]-risk[i]>=0)/alpha) }
m = mean(s)
return(m)
}
#% x is the time series of our relative returns
#VaR is the time series of our VaR estimates compute_scoreVaR(x,VaR,250,0.01)
```

Identifiability

Definition. Let $\rho : \mathcal{X} \to [-\infty, \infty]$ be a law-invariant risk measure and take $\mathcal{C} \subset \mathcal{D}$. We say that ρ is $\mathcal{C}$ -identifiable if there exists a function $p : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that for every $X \in \mathcal{X}$ with $F_X \in \mathcal{C}$ we have: - $\mathbb{E}[p(X,y)] = 0$ if and only if $y = \rho(X)$ - $\mathbb{E}[p(X,y)]$ is strictly increasing (in y) (provided the above expectations are well-defined).

Interpretation: If ρ is identifiable, the $\rho(X)$ is the unique zero of a suitable "expected distance" in the sense that:

$$\mathbb{E}[p(X,y)] \begin{cases} < 0 & \text{if } y < \rho(X) \\ = 0 & \text{if } y = \rho(X) \\ > 0 & \text{if } y > \rho(X) \end{cases}$$

We have to interpret y as possible forecast for $\rho(X)$ and the function p as a *penalty function* that penalizes forcasts that are from from $\rho(X)$. p assigns a negative average value to underestimated forecasts and a psoitive average value to overestimated ones.

Identifiability and model validation/selection Model validation and selection. Let $\rho : \mathcal{X} \to [-\infty, \infty]$ be an identifiable law-invariant risk measure. Compute the sample penalty

$$\frac{1}{n}\sum_{i=1}^{n} p\left(x_i, \rho\left(\widehat{F}_{(x_2,\ldots,x_n)}\right)\right)$$

and validate the model if the sample penalty is sufficiently close to zero. Similarly, compute the sample penalties

$$\left|\frac{1}{n}\sum_{i=1}^{n} p\left(x_i, \rho\left(\widehat{F}^{(1)}_{(x_1,\ldots,x_n)}\right)\right)\right|, \ldots, \left|\frac{1}{n}\sum_{i=1}^{n} p\left(x_i, \rho\left(\widehat{F}^{(m)}_{(x_1,\ldots,x_n)}\right)\right)\right|$$

and select the model whose forecast leads to the smallest sample penalty. Interpretation. The above rules are based on the fact that the sample distance

$$\frac{1}{n}\sum_{i=1}^{n} p\left(x_i, y\right)$$

can be used as a proxy for $\mathbb{E}[p(X, y)]$ for every forecast $y \in \mathbb{R}$ (assuming the underlying data generating process is sufficiently regular).

VaR is identifiable, ES is not.

```
compute_penaltyVaR = function(x,risk,n,alpha){ N = length(x)-n+1
p = rep(0,N)
for (i in 1:N){
p[i] = abs(mean(1-(1/alpha)*(-x[i:(i+n-1)]-risk[i]>0)))
m = mean(p)
}
return(m)
}
# x is the time series of our relative returns
# VaR is the time series of our VaR estimates compute_penaltyVaR(x,VaR,250,0.01)
```

Backtesting under ES

ES is neither identifiable nor elicitable - does **not** mean we cannot perform backtests.

We can use a “scoring function” that is yero on average in correspondence of ES.

```
compute_scoreES = function(x,risk1,risk2,n,alpha){ N = length(x)-n+1
s = rep(0,N)
for (i in 1:N){
s[i] = abs(mean(x[i:(i+n-1)]*(x[i:(i+n-1)]+risk1[i]<=0)/(alpha*risk2[i])+1)) }
m = mean(s)
return(m)
}
#x is the time series of our relative returns
#VaR is the time series of our VaR estimates
# ES is the time series of our ES estimates compute_scoreES(x,VaR,ES,250,0.025)
```

Proposition. Let $\alpha \in (0,1)$. For every $X \in \mathcal{X}$ with continuous distribution and such that $0 \neq \mathrm{ES}_\alpha(X) < \infty$ we have

$$\mathbb{E}\left[\frac{X\mathbb{1}_{\{X \leq -\mathrm{VaR}_\alpha(X)\}}}{\alpha \mathrm{ES}_\alpha(X)} + 1\right] = 0.$$

The above result suggests the "scoring function" $s : \overset{x}{\mathbb{R}} \times \overset{y}{\mathbb{R}} \times \overset{z}{(\mathbb{R} \setminus \{0\})} \to \mathbb{R}$

$$s(x,y,z) = \frac{x\mathbb{1}_{(-\infty,-y]}(x)}{\alpha z} + 1.$$

Figure 26: Plot title.

Model selection. For every model $j \in \{1,\dots,m\}$ compute the sample score

$$\left|\frac{1}{n}\sum_{i=1}^{n} s(x_i, \mathrm{VaR}_\alpha(\widehat{F}^{(j)}_{(x_1,\dots,x_n)}), \mathrm{ES}_\alpha(\widehat{F}^{(j)}_{(x_1,\dots,x_n)}))\right|$$

and select the model whose forecast leads to the smallest sample score.

Interpretation. The above rule is based on the fact that the sample score

$$\frac{1}{n}\sum_{i=1}^{n} s(x_i, y, z)$$

can be used as a proxy for $\mathbb{E}[s(X,y,z)]$ for all forecasts $y \in \mathbb{R}$ and $z \in \mathbb{R} \setminus \{0\}$ (assuming the underlying data generating process is sufficiently regular).

Note that the above sample scores depend on the estimation of two risk measures, namely VaR and ES!

Figure 27: Plot title.

Backtesting Under VaR : The Basel Framework

The Traffic Light Rule (TLR)

TLR is based on VaR and says a good model for **underlying data generating process** (data generation bscl.) should not give rise to too many losses exceeding the corresponding VaR level.

Testing of model is testing of the underlying distribution itself.

We want to calculate the **coverage ratio**, which test how often the above conditon (arrised default) is violated.

Determine the frequency of VaR violations, also called coyerage ratio, by

$$B_n^{freq}(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty, -\mathrm{VaR}_{i-1})}(x_i)$$

counter $\begin{cases} 1 & x_i + \mathrm{VaR}_{i-1} < 0 \Leftrightarrow x_i < -\mathrm{VaR}_{i-1} \\ 0 & \text{—}\,\text{||}\,\text{—} \geq 0 \end{cases}$

Figure 28: Plot title.

- green light if freq $\leq 1.6\%$
- yellow light if $1.6\% \leq$ freq. $\leq 3.6\%$
- red light if freq $> 3.6\%$

These percentages, or rather absolute number in relation to 252 (basel framework) come from Biniomal distributions.

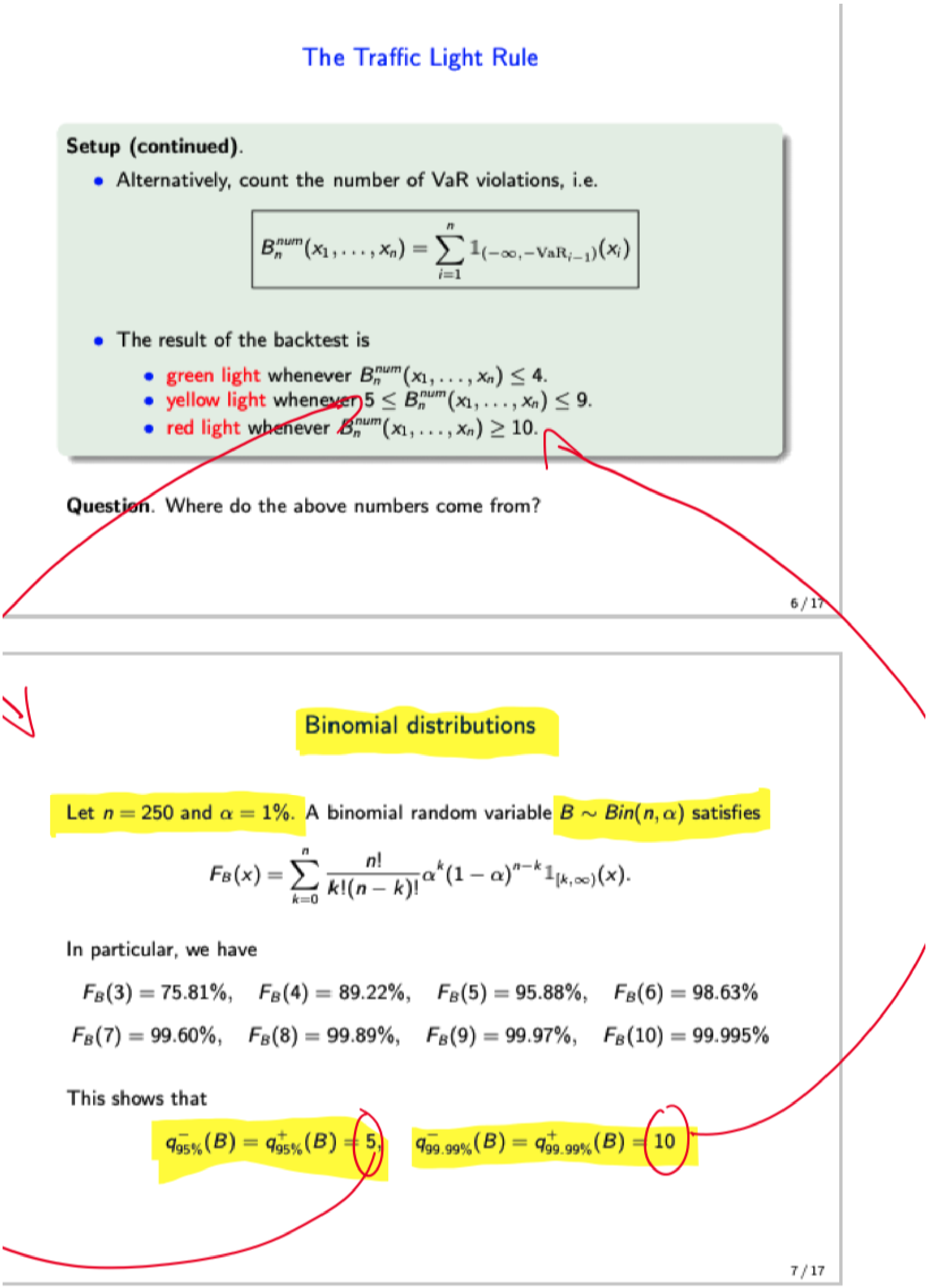


Figure 29: Plot title.

Interpretation. The Traffic Light Rule gives green light when the number of VaR violations does not reach the 95% quantile of a binomial distribution in the class $Bin(250, 1\%)$ and gives yellow light when it does not reach the 99.99% quantile of the same distribution.

Figure 30: Plot title.