VaR and CVaR analysis for risk management

Alexandre Carbonell,
Mathieu Hoff

Supervisor: Prof. Sara Van de Geer

Department of Mathematics
Eidgenössische Technische Hochschule Zürich

A semester project submitted for the Degree of
Master

D MATH

July 2023

Abstract

In the aftermath of the subprime financial crisis of 2007-2008, risk management has emerged as a critical aspect of the financial and insurance industry. The crisis exposed significant weaknesses in the banking system and the prudential framework, leading to excessive lending and risk-taking unsupported by effective transfer pricing and consistent valuation of complex products. The weaknesses in regulatory and supervisory frameworks have been exposed by the financial crisis, highlighting the need for action to prevent future crises. The Bank for International Settlements (BIS) [4] has been involved in implementing reforms to address these issues. The crisis had a dramatic impact on market valuation, and the resulting funding pressure was transmitted to major banks that had sponsored or provided funding guarantees to vehicles. As a result, risk management has become a crucial area of focus for financial institutions to ensure the stability and sustainability of the industry.

Table of contents

Chapter 1

Introduction

The process of managing risk in financial institutions and other areas heavily relies on measuring risk techniques. In the banking and insurance sectors, it is common practice to use probability distributions to model risk and represent it in the form of scalar-valued risk measures. In a formal sense, risk measures refer to the mapping of random variables that represent profits and losses (P&L) to real numbers that represent the amount of capital needed as a safeguard against insolvency. The two main risk measures used in financial institutions and regulation are Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). VaR is widely used as a measure of risk exposure by financial institutions, and the Basel Committee on Banking Supervision (BCBS) [4] has established it as the preferred approach for setting minimum capital requirements for market risk. VaR is defined as a quantile of the P&L distribution.

Definition 1 (VaR)**.** *Let L be a random variable and $\alpha \in (0,1)$, the Value-at-Risk of L at level α is then defined as:*

$$VaR_\alpha(L) = q_L^-(\alpha) = \min\{x \in \mathbb{R} \mid \mathbb{P}[L - x \leq 0] \geq \alpha\}.$$

CVaR, on the other hand, is a more recently employed measure of risk but equally interesting. It is the conditional expectation of the random variable L given exceedance of VaR at a given level α. Note that when the random variable L has continuous cumulative distribution function (CDF), it is also and more generally called Expected Shortfall (ES). In this paper, we will use both notations.

Definition 2 (CVaR and ES)**.** *Let L be a random variable and $\alpha \in (0,1)$, the Conditional Value-at-Risk of L at level α is then defined as:*

$$CVaR_\alpha(L) = \mathbb{E}[L \mid VaR_\alpha(L) \geq L].$$

The Expected Shortfall of L at level α is then defined as:

$$ES_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L)du.$$

The estimation and forecasting of risk exposure measures, such as VaR and CVaR, have been the subject of extensive research efforts, motivated by the global economic crisis and raised various questions in terms of efficiency and accuracy. Consequently, this paper aim at reviewing the main properties that are commonly used to identify and assess the accuracy of risk measures in a first time. Then we will redefine the principal methods of forecasting that are widely used in the literature. Finally we will apply those methods to derive VaR and CVaR from real-world data and do a simulation for portfolios risk management using the comparison elements that are employed for this aim. In summary, we will discuss in chapter 2, the various properties that allow to identify risk measures namely the coherence, the (conditional) elicitability and the robustness and we will see which properties are satisfied by VaR and CVaR and what they bring in term of risk analysis. Then we will slide in chapter 3 to the forecasting of VaR and CVaR through three principal methods from the parametric estimation and the historical data forecasting to analytical methods and the Monte Carlo simulation. In chapter 4, we will present the comparison and back-testing methods we can apply to test the forecasted VaR and CVaR. These methods fall into two main types: traditional back-testing and comparative back-testing. Lastly, in chapter 5 and after defining the mathematical support of our paper we plan to do a numerical experiment on the CAC40 index. Note that through chapter 3 and 4 we will mainly make use of the paper of Nolde, N. and Ziegel, J. (2017) [5]. Below we can see a graph we made with data found on Yahoo Finance showing the variations of the Morgan Stanley stock price particularly during crisis time.

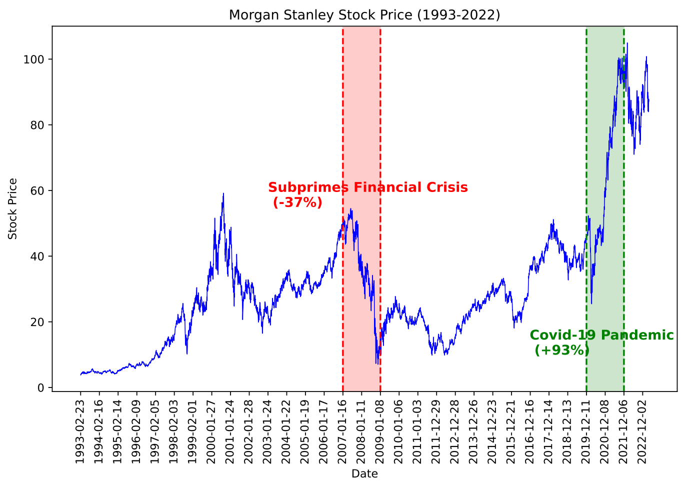

Fig. 1.1 Morgan Stanley stock price over the years.

Chapter 2

Risk measure analysis

While it is more comprehensive to consider multiple risk measurements for evaluating a portfolio's risk, in practice, it is common to report a single number as the basis for strategic decision-making. This chapter provides an overview of various risk measurements properties, their characteristics, and evaluates which properties are satisfied by VaR and CVaR, along with their financial implications for risk management.

2.1 Coherent, convex and averse risk measure

A coherent risk measure is a mathematical function that satisfies a set of desirable properties for measuring the risk associated with a portfolio of financial assets. One of the most crucial properties of a coherent risk measure is its ability to accurately capture the concept of risk aversion, where investors prefer portfolios with lower risk levels, all else being equal. This makes coherent risk measures a crucial tool in risk management, as they provide a framework for evaluating and comparing the riskiness of different portfolios and making informed investment decisions based on their risk preferences.

2.1.1 Definition and properties

Definition 3 (Coherent risk measure)**.** *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and ρ a risk measure. Then we say that $\rho : L^{\infty}(\Omega, \mathscr{F}, \mathbb{P}) \to \mathbb{R}$ is a coherent risk measure if it satisfies the following properties:*

(C1) Positive homogeneity: $\rho(\lambda L) = \lambda \rho(L)$, for $\lambda > 0$ and $L : \Omega \to \mathbb{R}$ a random variable.

(C2) Subadditivity: $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$, for all random variables $L_1, L_2 : \Omega \to \mathbb{R}$.

(C3) Monotonicity: $L_1 \leq L_2 \implies \rho(L_1) \leq \rho(L_2)$*, for all random variables* $L_1, L_2 : \Omega \to \mathbb{R}$.

(C4) Translation invariance: $\rho(L - \lambda) = \rho(L) - \lambda$*, for all* $\lambda \in \mathbb{R}$ *and* $L : \Omega \to \mathbb{R}$ *a random variable.*

The coherence of a risk measure is a very important property to look at when assessing risk. Three additional risk measure properties are often used as a complementary property to subadditivity. Namely comonotonically additivity, law-invariance and convexity. First of all, let us define the concept of comonotonicity between two random variables.

Definition 4 (Comonotonicity)**.** *Two random variables* $L_1, L_2 : \Omega \to \mathbb{R}$ *are comonotonic if there exists a random variable* $X : \Omega \to \mathbb{R}$ *(the common risk factor) and two non-decreasing functions* f_1 *and* f_2 *such that:* $L_1 = f_1(X)$ *and* $L_2 = f_2(X)$ *in distribution.*

Moreover, we also bring one more axiomatic definition that are of great interest furthermore, the aversity. Then, we have the following additional properties:

Definition 5 (Additional properties)**.** *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space,* ρ *be a risk measure and* $L, L_1, L_2 : \Omega \to \mathbb{R}$ *random variables. Then we have the following additional axioms for* ρ*:*

(C5) Comonotonically additive: $\rho(L_1 + L_2) = \rho(L_1) + \rho(L_2)$*, for all comonotonic random variables* $L_1, L_2 : \Omega \to \mathbb{R}$.

(C6) Law-invariant: $\mathbb{P}(L_1 \leq x) = \mathbb{P}(L_2 \leq x),\ x \in \mathbb{R} \implies \rho(L_1) = \rho(L_2)$.

(C7) Convex: $\rho(\lambda L_1 + (1-\lambda)L_2) \leq \lambda\rho(L_1) + (1-\lambda)\rho(L_2)$*, for* $\lambda \in (0,1)$.

(C8) Aversity: $\rho(L) > \mathbb{E}[L]$*, for L non-constant.*

The interpretation of aversity is that the risk of loss in a non-constant random variable L is only considered acceptable if the expected value of L is less than zero. We can show some relation between these axioms.

Lemma 2.1.1 (Coherent vs convex)**.** *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space,* ρ *be a risk measure defined on this probability space. Then under positive homogeneity (C1), subadditivity (C2)* $\iff$ *convex (C7).*

Proof. The proof is straightforward.
$(\Longrightarrow)$ Suppose that ρ satisfies (C1) and (C2) and take $\lambda \in (0,1)$ and $L_1, L_2 : \Omega \to \mathbb{R}$. Then we have:

$$\rho(\lambda L_1 + (1-\lambda)L_2) \overset{(C2)}{\leq} \rho(\lambda L_1) + \rho((1-\lambda)L_2) \overset{(C1)}{\leq} \lambda\rho(L_1) + (1-\lambda)\rho(L_2).$$

($\Longleftarrow$) Suppose now that ρ satisfies (C1) and (C7), then for $\lambda = \frac{1}{2}$ we have:

$$\rho(L_1 + L_2) \overset{(C1)}{\leq} 2 \cdot \rho\left(\frac{1}{2}L_1 + \frac{1}{2}L_2\right) \overset{(C7)}{\leq} 2 \cdot \left(\frac{1}{2}\rho(L_1) + \frac{1}{2}\rho(L_2)\right) = \rho(L_1) + \rho(L_2).$$

$\square$

2.1.2 Relation to VaR and CVaR

The fact that VaR is represented as a quantile of a profit and loss (P&L) distribution implies that VaR possesses the properties of positive homogeneity (C1), monotonicity (C3) and translation invariance (C4). However, VaR fails to be a coherent risk measure as it fails to be subadditive (C2). In financial terms, it is difficult to implement a decentralized risk management approach using VaR since there is no assurance that combining VaR figures for different portfolios or business units will provide a dependable estimate for the overall risk of the organization. That was an argument of Susanne Emmer, Marie Kratz, Dirk Tasche in their paper [7]. Therefore, the reliability of the aggregated VaR numbers must be carefully assessed before using them as a basis for risk management decisions. Moreover, by looking at the lemma 2.1.1, we see that VaR also fails to be convex (C7) in general. Note that the averse property is satisfied by VaR but VaR is not an averse risk measure since it fails to be subadditive (C2). See Value-at-Risk vs. Conditional Value-at-Risk in Risk Management and Optimization [8] for the precise definition. Lastly, the Comonotonically additivity is a property of risk measure that is satisfied by VaR as well as the law-invariance.

Lemma 2.1.2 (Comonotonically additivity and law-invariance of VaR)**.** *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $L_1, L_2 : \Omega \to \mathbb{R}$ to comonotonic random variables and $\alpha \in (0,1)$. Then we have:*

i. Comonotonically additivity: $VaR_\alpha(L_1 + L_2) = VaR_\alpha(L_1) + VaR_\alpha(L_2)$

ii. Law-invariance: $\mathbb{P}(L_1 \leq x) = \mathbb{P}(L_2 \leq x),\ x \in \mathbb{R} \implies VaR_\alpha(L_1) = VaR_\alpha(L_2)$.

Proof. i. One can show that $(L_1, L_2) \overset{d}{=} (q^-_{L_1}(U), q^-_{L_2}(U))$ for $U \sim Uniform(0,1)$ and $q^-_L(a) := \min\{x \in \mathbb{R} : F_L(x) \geq a\}$ the left-a-quantile of a random variable L with CDF F_L and $a \in (0,1)$. Now, applying the continuous mapping theorem we obtain that $L_1 + L_2 \overset{d}{=} q^-_{L_1}(U) + q^-_{L_2}(U)$ and then:

$$\begin{aligned} VaR_\alpha(L_1 + L_2) &= VaR_\alpha(q^-_{L_1}(U) + q^-_{L_2}(U)) = (q^-_{L_1} + q^-_{L_2})(VaR_\alpha(U)) = (q^-_{L_1} + q^-_{L_2})(\alpha) \\ &= VaR_\alpha(L_1) + VaR_\alpha(L_2). \end{aligned}$$

We used in the third equality that $\mathbb{P}[(q^-_{L_1}+q^-_{L_2})(U)\leq(q^-_{L_1}+q^-_{L_2})(\alpha)]=\mathbb{P}[U\leq\alpha]=\alpha$ nowing that q_{L_i} is an increasing function for $i=1,2$.

ii. The law-invariance of VaR is straightforward. Suppose that $\mathbb{P}(L_1\leq x)=\mathbb{P}(L_2\leq x)$, for all $x\in\mathbb{R}$, then we have:

$$
\begin{aligned}
VaR_\alpha(L_1)=q^-_{L_1}(\alpha)&=\min\{x\in\mathbb{R}\mid\mathbb{P}[L_1\leq x]\geq\alpha\}=\min\{x\in\mathbb{R}\mid\mathbb{P}[L_2\leq x]\geq\alpha\}\\
&=q^-_{L_2}(\alpha)=VaR_\alpha(L_2).
\end{aligned}
$$

□

On the other hand, CVaR (ES when L has a continuous CDF) is a coherent, convex, comonotically additive and law-invariant risk measure. i.e: ES satisfies all the properties (C1)-(C8).

Lemma 2.1.3 (Coherence of ES)**.** *ES is a coherent risk measure.*

Proof. By combining the properties of integrals together with those of VaR, we directly have the positive homogeneity, the monotonicity and the translation invariance of ES. We then bring our attention to the subadditivity. There exists various proofs of the subadditivity of ES using quite different approach (see Seven Proofs for the Subadditivity of Expected Shortfall [3]). For sake of simplicity in the rest of the proofs of the other properties, let us have an approach using the law of large numbers for order statistics.

Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of *iid* random variables define on a probability space $(\Omega,\mathscr{F},\mathbb{P})$. Using the strong law of large numbers we get that:

$$
ES_\alpha(X_1)=\lim_{n\to\infty}\frac{1}{n(1-\alpha)}\sum_{i=1}^{n\lfloor 1-\alpha\rfloor}X_{(i)}\quad\text{a.s,}
$$

where $X_{(1)}\geq\ldots\geq X_{(n)}$ are the order statistics of $X_1,\ldots,X_n$. This representation suggests a way of estimating expected shortfall in the situation when we have large samples and $n\lfloor 1-\alpha\rfloor$ is a relatively large number. A proof of this result can be found in (Proposition 4.1 of Acerbi and Tasche (2002)). Then, set $A^n_m:=\{(i_1,\ldots,i_m)\in\mathbb{N}^m:1\leq i_1\leq\ldots\leq i_m\leq n\}$ and note that $\sum_{i=1}^m X_{(i)}=\max\{X_{i_1}+\ldots+X_{i_m}:(i_1,\ldots,i_m)\in A^n_m\}$. Now, take $L_1,L_2:\Omega\to\mathbb{R}$ with joint CDF F, let $((X_i,Y_i))_{i\in\mathbb{N}}$ be a sequence of *iid* bivariate random vector with the same CDF F and set

$Z_i = X_i + Y_i,\ i \in \mathbb{N}$. Then we have:

$$\begin{aligned}
\sum_{i=1}^{m} Z_{(i)} &= \max\{Z_{i_1} + \ldots + Z_{i_m} : (i_1, \ldots, i_m) \in A_m^n\} \\
&= \max\{X_{i_1} + \ldots + X_{i_m} + Y_{i_1} + \ldots + Y_{i_1 m} : (i_1, \ldots, i_m) \in A_m^n\} \\
&\leq \max\{(X_{i_1} + \ldots + X_{i_m}) + (Y_{j_1} + \ldots + Y_{j_m}) : \{(i_1, \ldots, i_m), (j_1, \ldots, j_m)\} \in A_m^n\} \\
&= \max\{X_{i_1} + \ldots + X_{i_m} : (i_1, \ldots, i_m) \in A_m^n\} + \max\{Y_{j_1} + \ldots + Y_{j_m} : (j_1, \ldots, j_m) \in A_m^n\} \\
&= \sum_{i=1}^{m} X_{(i)} + \sum_{i=1}^{m} Y_{(i)}
\end{aligned}$$

By setting $m = n\lfloor 1 - \alpha \rfloor$, we have:

$$\frac{1}{n(1-\alpha)} \sum_{i=1}^{n\lfloor 1-\alpha \rfloor} Z_{(i)} \leq \frac{1}{n(1-\alpha)} \sum_{i=1}^{n\lfloor 1-\alpha \rfloor} X_{(i)} + \frac{1}{n(1-\alpha)} \sum_{i=1}^{n\lfloor 1-\alpha \rfloor} Y_{(i)}.$$

And now, letting $n \to \infty$ we have, using the result previously stated, that:

$$\begin{aligned}
ES_\alpha(L_1 + L_2) &= \lim_{n\to\infty} \frac{1}{n(1-\alpha)} \sum_{i=1}^{n\lfloor 1-\alpha \rfloor} Z_{(i)} \leq \lim_{n\to\infty} \frac{1}{n(1-\alpha)} \sum_{i=1}^{n\lfloor 1-\alpha \rfloor} X_{(i)} + \lim_{n\to\infty} \frac{1}{n(1-\alpha)} \sum_{i=1}^{n\lfloor 1-\alpha \rfloor} Y_{(i)} \\
&= ES_\alpha(L_1) + ES_\alpha(L_2).
\end{aligned}$$

□

Lemma 2.1.4 (Convexity of ES)**.** *ES is a convex risk measure.*

Proof. The only thing left to prove to get that ES is a convex risk measure is the convexity property (C7). However, using the limit representation of ES as in the previous proof, it is straightforward. □

Lemma 2.1.5 (Aversity of ES)**.** *ES is an averse risk measure.*

Proof. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $L : \Omega \to \mathbb{R}$ be a non-constant random variable. Again, we use the limit representation of ES and the standard strong law of large number to get the following result:

$$ES_\alpha(L) = \lim_{n\to\infty} \frac{1}{n(1-\alpha)} \sum_{i=1}^{n\lfloor 1-\alpha \rfloor} L_{(i)} \geq \lim_{n\to\infty} \frac{1}{n(1-\alpha)} \sum_{i=1}^{n} L_i > \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} L_i = \mathbb{E}[L].$$

□

Lemma 2.1.6 (Comonotonically additivity and law-invariance of ES). *ES is a comonotonically additive and law-invariant risk measure.*

Proof. The proof for those two properties is inherant from the definition of ES and the proof of comonotonically additivity and law-invariance for VaR. □

2.2 Elicitability

2.2.1 Definition and properties

Elicitability is a characteristic of a risk measure that allows it to be determined or assessed using empirical data without relying on subjective judgements or assumptions. In short, a risk measure is considered elicitable if it can be inferred or extracted from the actions or choices of the decision-maker, without the need to provide further information. In order to formally define the notion of an elicitable risk measure, it is necessary to introduce the concept of consistency of a scoring function.

We let a functional v on a class of a probability space $\mathbb{P}$ on $\mathbb{R}$:

$$\begin{aligned} v : \mathbb{P} &\to 2^{\mathbb{R}} \\ \mathbb{P} &\to v(\mathbb{P}) \subset \mathbb{R} \end{aligned}$$

Definition 6 ((Strictly) Consistency of a scoring function). *We say that a scoring function $s : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ is* ***consistent*** *relative to the class $\mathbb{P}$ if:*

$$\mathrm{E}_P[s(t,L)] \leq \mathrm{E}_P[s(x,L)]$$

$\forall P \in \mathbb{P}, \forall t \in v(\mathbb{P}), \forall x \in \mathbb{R}$. *Here, L has distribution P.*

We say that a scoring function is ***strictly consistent*** *if it is consistent and:*

$$\mathrm{E}_P[s(t,L)] = \mathrm{E}_P[s(x,L)] \Rightarrow x \in v(\mathbb{P})$$

Definition 7. *We say that a functional v is* ***elicitable*** *relative to $\mathbb{P}$ if and only if there is a scoring function s which is strictly consistent for v relative to $\mathbb{P}$.*

The concept of elicitability is quite useful in determining the optimal point forecasts. When a functional has a (strictly) consistent scoring function, it becomes the only function under which the functional is an optimal point forecast. Therefore, discovering a strictly consistent scoring function for a functional v enables us to determine the optimal forecast $\hat{x}$ for $v(\mathbb{P})$ by $\hat{x} = \arg\min \mathrm{E}_P[s(x,L)]$.

Elicitability of a functional of probability distributions means that the functional can be estimated by generalised regression. This property makes elicitability an important concept as it allows for comparing the performance of different forecast methods.

Theorem 2.2.1 (Osband)**.** *An elicitable functional v has convex level sets, i.e. : If $P_0 \in \mathbb{P}$ and $P_1 \in \mathbb{P}$, and $P^\star = p \cdot P_0 + (1-p) \cdot P_1 \in \mathbb{P}$ for some $p \in (0,1)$, then $t \in v(P_0)$ and $t \in v(P_1) \Rightarrow t \in v(P^\star)$.*

Proof. In order to have $t \in v(P^\star)$ we need to show that, for s a strictly consistent scoring function for v,

$$\mathrm{E}_{P^\star}[s(t,L)] \leq \mathrm{E}_{P^\star}[s(x,L)]$$

Let all the conditions stated in the theorem hold. Then we have for any $x \in \mathbb{R}$:

$$\begin{aligned}
\mathrm{E}_{P^\star}[s(t,L)] &= p \cdot \mathrm{E}_{P_0}[s(t,L)] + (1-p) \cdot \mathrm{E}_{P1}[s(t,L)] \\
&\leq p \cdot \mathrm{E}_{P_0}[s(x,L)] + (1-p) \cdot \mathrm{E}_{P1}[s(x,L)] \\
&= \mathrm{E}_{P^\star}[s(x,L)].
\end{aligned}$$

In the first equality we used that $P^\star = p \cdot P_0 + (1-p) \cdot P_1$. To move from the first line to the second we used the fact that s is consistent. And finally we used again the definition of $P^\star$. Thus by the definition of a strictly consistent scoring function we get that $t \in v(P^\star)$. □

2.3 Conditional elicitability

2.3.1 Definition and properties

Some risk measures do not verify the elicitability property but it turns out that they can be conditional elicitable.

Definition 8 (Conditional elicitability)**.** *Let two functionals $\tilde{\beta}$ and $\beta : S \to 2^{\mathbb{R}}$ with $S \subset \mathbb{P} \times 2^{\mathbb{R}}$. We say that a functional v of $\mathbb{P}$ is **conditionally elicitable** if we have:*

i) $\tilde{\beta}$ is elicitable relative to $\mathbb{P}$

ii) $\forall x \in \tilde{\beta}(P)$ the functional $\beta_x : P_x \to 2^{\mathbb{R}}$, $P \to \beta(P,x) \subset \mathbb{R}$ is elicitable relative to $P_x = \{P \in \mathbb{P} : (P,x) \in S)\}$.

iii) $v(P) = \beta(P, \tilde{\beta}(P))\ \forall P \in I\!P.$

To forecast some risk measures that are not elicitable, the conditional elicitability can be a useful concept. Indeed conditional elicitability provides the possibility to forecast it in two steps. For example, let β be elicitable, and $\beta(P,x)$ only be conditionnaly elicitable. Hence due to the elicitability of β we can first forecast $\beta(P)$. Then taking the previous result for $\beta(P)$, we can forecast $\beta(P,x)$ due to the elicitability of β_x.

2.3.2 Relation to VaR and CVaR

It has recently been discovered that the CVaR is not elicitable contrary to the risk measure VaR. This led to a debate about the best risk measure between VaR and CVaR. We will thus show that VaR is elicitable and then that CVaR is conditionally elicitable.

First of all:

Lemma 2.3.1. *VaR is elicitable.*

Proof. We discussed the fact that if ϕ is elicitable then it is the minimum of some strictly consistent scoring function $S(x,y)$ according to: $\phi = \arg\min_x \mathbb{E}[S(x,Y)]$. To prove the elicitability of VaR we will use the following scoring function:

$$S(x,y)) = (\mathbb{1}_{x \geq y} - \alpha)(x-y).$$

Thus we want to have that $VaR_\alpha(Y) = \arg\min_x \mathbb{E}[\mathbb{1}_{x \geq Y} - \alpha)(x-Y)]$. To find x we want to take the derivative of the expectation and set it to 0. For simplicity we write:

$$\mathbb{1}_{x \geq y} = H(x-y) = \begin{cases} 1 & \text{if } x-y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Now, considering that $\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x)\,dx$, where $f_X(x)$ represents the density of x, we have:

$$\mathbb{E}[S(x,y)] = \int_{-\infty}^{+\infty} (H(x-y) - \alpha)(x-y) f_Y(y)\,dy.$$

Given the definition of H we can split the integral with $y \leq x$ in which part we will have $H(x-y) = 1$ and in another half with $y > x$, in which we will have $H(x-y) = 0$:

$$\mathbb{E}[S(x,y)] = \int_{-\infty}^{x} (1-\alpha)(x-y) f_Y(y) - \int_{x}^{\infty} \alpha(x-y) f_Y(y)\,dy$$

We then take the derivative:

$$
\begin{aligned}
&\frac{d}{dx}\left(\int_{-\infty}^{x}(1-\alpha)(x-y)f_Y(y)-\int_{x}^{\infty}\alpha(x-y)f_Y(y)\,dy\right)\\
=&\frac{d}{dx}\left(\int_{-\infty}^{x}(1-\alpha)(x-y)f_Y(y)\right)-\frac{d}{dx}\left(\int_{x}^{\infty}\alpha(x-y)f_Y(y)\,dy\right)\\
=&(1-\alpha)\left[(x-x)f_Y(y)-0\cdot(x+\infty)f_Y(\infty)+\int_{-\infty}^{x}\frac{d}{dx}[(x-y)f_Y(y)]\,dy\right]\\
-&\alpha\left[0\cdot(\infty-x)f_Y(y)-(x-x)f_Y(x)+\int_{x}^{\infty}\frac{d}{dx}[(x-y)f_Y(y)]\,dy\right]\\
=&(1-\alpha)\int_{-\infty}^{x}f_Y(y)\,dy-\alpha\int_{x}^{\infty}f_Y(y)\,dy\\
=&\int_{-\infty}^{x}f_Y(y)\,dy-\alpha\left(\int_{-\infty}^{x}f_Y(y)\,dy+\int_{x}^{\infty}f_Y(y)\,dy\right)=\int_{-\infty}^{x}f_Y(y)\,dy-\alpha
\end{aligned}
$$

Here we used the Leibniz rule for both integrals to get the result. And in the final step we used the fact that the integral of a density is equal to 1. So now we solve the following equation:

$$\int_{-\infty}^{x}f_Y(y)\,dy-\alpha=0\Leftrightarrow\alpha=\int_{-\infty}^{x}f_Y(y)\,dy$$

But then we get:

$$x=F_Y^{-1}(\alpha)$$

which means that the minimizer we were looking for is $VaR_\alpha(Y)$, which proves that VaR is elicitable through its scoring function. □

Gneiting proved in 2011 that CVaR is not elicitable. This means that it is not possible to find a scoring function s which is strictly consistent for CVaR. As said above this finding has provoked debate and misinterpretation, as some thought that the non-elicitability of CVaR would imply its non-backtestability, which is wrong.

So Expected Shortfall is not elicitable but we can show that it is conditionally elicitable:

Lemma 2.3.2. *For continuous distributions with finite means, ES is conditionally elicitable.*

Proof. Let's prove that all points of the definition of the conditionally elicitable are satisfied by the ES. First of all, for $\alpha\in[0,1]$, we have that for continuous distributions ES simplifies to $ES_\alpha(X)=\mathrm{E}[X\mid X\geq q_\alpha]$ where X is a random variable with distribution P. Let define S and $\mathscr{P}$ as follow:

$$
\begin{aligned}
&P=\{\text{continuous distributions on }\mathbb{R}\text{ with finite means}\}\\
&S=\{(P,c)\in P\times\mathbb{R}:P[c,\infty)>0\}
\end{aligned}
$$

Now we can write β and $\tilde{\beta}$ such that ES satisfies our conditions: $\beta : S \mapsto \mathbb{R}$

$$\beta(P,c) = \mathrm{E}_P[X \mid X \geq c]$$

and $\tilde{\beta} : P \mapsto \mathbb{R}$

$$\tilde{\beta}(P) = q_\alpha(X).$$

Then, as quantiles are elicitable (using the fact that VaR is elicitable), the first property is satisfied. By the definition of β and $\tilde{\beta}$ we trivially have that

$$ES_\alpha(X) = \beta(P, q_\alpha(X)) = \beta(P, \tilde{\beta}(P))$$

showing that the third property is satisfied. So we still have to prove that $\forall x \in \tilde{\beta}(P)$ the functional $\beta_x : P_x \to 2^{\mathbb{R}},\ P \to \beta(P,x) \subset \mathbb{R}$ is elicitable relative to $P_x = \{P \in \mathbb{P} : (P,x) \in S)\}$. The idea is to see that for a fixed x, β_x is elicitable with this strictly consistent function:

$$s(z,y) = \begin{cases} \varphi(y) - \varphi(z) - \varphi'(z)(y-z) & \text{if } x \leq z, \\ 0 & \text{otherwise,} \end{cases}$$

with $\varphi(z) = \dfrac{z^2}{1+|z|}$.

$\square$

Another powerful result that will not be proved here, is that CVaR is jointly elicitable with VaR. We will see later that this an important result to be able to backtest Expected Shortfall. Indeed as Expected Shortfall is not elicitable, it can not be backtested on its own but has to be backtested with VaR.

2.4 Robustness

The ability of a risk measure to offer consistent and trustworthy evaluations of risk in various conditions or scenarios is what is meant by its robustness. Chapter 3 will review various estimators and forecasting methods for both VaR and CVaR (ES). It is therefore crucial to evaluate the robustness of these estimations. If there is no robustness (as defined appropriately), the results may be insignificant. This is because even minor errors in measuring the P&L distribution can significantly affect the estimated risk measure. By a more robust estimator we mean an estimator whose performance is not so susceptible to the presence of outlying data values.

2.4.1 Definition and properties

In this paper, we present two types of robustness used for risk management. The first one is the robustness with respect to the weak topology and the second one, a more improved concept, is the robustness with respect to the Wasserstein metric. To address this, we examine risk functionals that handle risk measures for distribution levels and utilize existing statistical principles. However, these functionals are generally not continuous concerning the weak topology since it disregards the distribution's tails. This was identified by Weber in 2006, who proposed employing a more potent topology. You can also look at the paper of R. Kieselb, R. Ruhlickeb, G. S. J. Z.[6]. Therefore, we will employ robustness, meaning continuity with respect to a suitable metric. First, let us define the appropriate mathematical objects for our purpose. When we speak of estimation of risk measure for instance, we mathematically look at statistical functionals.

Definition 9 (Statistical functional / Estimator)**.** *Let $\mathscr{F}$ be a set of CDF defined on $\mathbb{R}$. Then a statistical functional or estimator is a map $T : \mathscr{F} \to \mathbb{R}$ that is applied to an empirical CDF F_n of a sequence of random variables $(L_i)_{i=1,...,n}$, iid, with theoretical CDF $F \in \mathscr{F}$. I.e the estimator is then $T_n = T(L_1,...,L_n)$ and if we can additionally write it this way $T_n = T(F_n)$, then it is a statistical functional.*

Now, the robustness of a risk measure ρ can be related to the continuity of T_ρ since we have, in the case of law-invariant risk measure, the following relation: $\rho(L) = T_\rho(F)$ for L a random variable with CDF F. And thus, we need to specify how to quantify the effect of a small deviation from the true distribution on the risk measure. For this, the concept of continuity is important.

Definition 10 (Continuity of statistical functional)**.** *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\mathscr{F}_c \subset \mathscr{F}$ be a convex class of CDF. A functional $T : \mathscr{F}_c \to \mathbb{R}$ is continuous at $F \in \mathscr{F}_c$ if we have:*

$$T(\tilde{F}) - T(F) = o_{\mathbb{P}}(1) \text{ when } d(\tilde{F}, F) = o_{\mathbb{P}}(1), \text{ for } \tilde{F} \in \mathscr{F}_c.$$

Where d is some appropriate distance between two CDF.

The robustness of some risk measures can thus be assessed by looking at the continuity of the corresponding statistical functional with respect to the weak topology which is a metric space in our settings. However, we will see more in detail in the next section that there are interesting risk measures that are not continuous with respect to the weak topology. Hence, it becomes imperative to establish an alternative metric that is well-suited for maintaining the continuity of the corresponding statistical functionals of the risk measure. This is especially necessary when those risk measure exhibit a lack of continuity concerning the weak topology.

Let us then endeavor to define a metric that is better equipped to handle this situation and thus describe the robustness of more risk measures.

We then come to the Wasserstein distance.

Definition 11 (Wasserstein distance)**.** *Let $(\Omega, \mathscr{F}, I\!P)$ be a probability space. Let L_1 and L_2 two random variables defined on this probability space with respective CDF F and $\tilde{F}$. The Wasserstein distance $d_W(\cdot,\cdot) : \mathscr{F} \times \mathscr{F} \to \mathbb{R}$, between the two CDFs is then defined for $p \geq 1$ by:*

$$d_W(L_1, L_2) := d_W(F, \tilde{F}) = \left(\int_0^1 |F^{-1}(x) - \tilde{F}^{-1}(x)|^p dx \right)^{\frac{1}{p}}.$$

Note that we use the abuse of notation $d_W(L_1, L_2)$ instead of $d_W(F, \tilde{F})$. One important property of this distance for risk management is the scaling property. It allows, for instance, the change of currency in assessing the risk without altering the computations.

Lemma 2.4.1 (Scaling property of the Wasserstein distance)**.** *Let L_1, L_2 and p as in the above definition. Then the Wasserstein distance satisfy, for $\lambda \in \mathbb{R}$:*

$$d_W(\lambda L_1, \lambda L_2) = |\lambda| d_W(L_1, L_2).$$

Proof. For a complete proof, see the related literature of Wasserstein distance. In the framework of risk measure, the case $p = 1$ is enough and thus we prove it. We then have:

$$d_W(L_1, L_2) = \int_0^1 |F^{-1}(x) - \tilde{F}^{-1}(x)| dx = \int_{\mathbb{R}} |F(x) - \tilde{F}(x)| dx,$$

with a suitable change of variable and thus:

$$d_W(\lambda L_1, \lambda L_2) = \int_{\mathbb{R}} |\lambda F(x) - \lambda \tilde{F}(x)| dx = |\lambda| \int_{\mathbb{R}} |F(x) - \lambda \tilde{F}(x)| dx = |\lambda| d_W(L_1, L_2).$$

□

We have now in our hand two ways of assess the robustness of the risk measures we are interested in. Let us explore which type of robustness is satisfied by VaR and CVaR.

2.4.2 Relation to VaR and CVaR

Since VaR and CVaR are distribution-based values and law-invariant, it is interesting to exploit the two concepts of robustness defined above. First of all, by the discussion above, we can show that VaR at level α is continuous with respect to the weak topology for $F \in \mathscr{F}$ if the quantile function F^{-1} is continuous in α. Additionally, we discuss the Wasserstein distance because both CVaR and VaR are continuous with respect to this metric. In summary, the concept of Wasserstein distance is a good way to assess robustness of our risk measures of interest given some assumptions that will not be discussed here (see: Conceptualizing Robustness in Risk Management from Stahl [6]). Certainly, in order to interpret these statistics, it is essential that they are consistent estimates for the corresponding parameter. Consistency would be confirmed by a convergence in terms of the Wasserstein distance.

Chapter 3

Forecasting of VaR and CVaR

There are various ways of estimating risk measure. From the wide range of estimation methods, we find the parametric estimation assuming statistical properties of the underlying P&L random variable L. Also, it is possible to estimate VaR and CVaR by looking into the historical data. More developed and somewhere accurate methods involve Monte Carlo simulations. Although these techniques of forecasting differ in the procedure and in the assumptions made, they all need to be computed on a fixed time horizon h so that we can compare them in a good way. VaR and CVaR are typically estimated for a specific time period, such as a daily, weekly or monthly period. In term of quantitative risk management, it is of prior importance to have forecasting techniques that are accurate but also simple to implement and to maintain. When making assumptions on our financial data, the more we capture the underlying factors that drive the financial instrument, the better. In this paper we focus our attention of univariate financial instruments and assume that they are defined by time series.

3.1 Parametric estimation

There exists numerous parametric approach for estimating VaR and CVaR in the literature. The objective in these approach is to find the best way to model the distribution of the returns or log-returns (P&L) of a given stock S_t given its historical values to assess the adequately the risks using risk measures. That explains why, in this set-up, potential model misspecification as well as estimation uncertainty in small samples data can be detrimental for forecasting.

$AR(m)$-$GARCH(p,q)$ processes have been widely used and demonstrated to be effective in modeling the log-returns of stocks across various markets and time periods, making them a suitable choice for financial modeling and analysis. We will then assume this structure for the evolution of the negated log-returns of the underlying stock i.e., the random variable $(L_t)_{t\in\mathbb{Z}}$

follows a *AR*(m)-*GARCH*(p,q) process for which we will need to find the optimal parameters p,q and m as well as the internal parameters of a *GARCH* process. For this aim, let us recall the mathematical expression of a *AR*(m)-*GARCH*(p,q) process.

Definition 12 (*AR*(m)-*GARCH*(p,q) process)**.** *An Autoregressive combined with Generalized Autoregressive Conditionally Heteroskedastic AR*(m)*-GARCH*(p,q) *process is a time series defined on a probability space* $(\Omega,\mathscr{F},\mathbb{P})$*. In mathematical terms, a random variable* $(L_t)_{t\in\mathbb{Z}}$ *evolves as a AR*(m)*-GARCH*(p,q) *if it is a stationary process and satisfies:*

i. $L_t = \mu_t + \sigma_t \cdot Z_t$.

ii. $\mu_t = \gamma_0 + \sum_{k=1}^{m} \gamma_k \cdot L_{t-k}$.

iii. $\sigma_t^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k \cdot L_{t-k}^2 + \sum_{k=1}^{q} \beta_k \cdot \sigma_{t-k}^2$.

Where $\alpha_0 > 0$ *and* $\alpha_k \geq 0,\ \ \forall k = 1,...,p$*. Also,* $\gamma_0,...,\gamma_m,\beta_1,...,\beta_q \in \mathbb{R}$*. Moreover,* $(\mu_t)_{t\in\mathbb{Z}}$ *is a* $\mathscr{F}_{t-1}$*-measurable process denoted as location,* $(\sigma_t)_{t\in\mathbb{Z}}$ *is a* $\mathscr{F}_{t-1}$*-measurable process denoted as the scale or volatility and* $(Z_t)_{t\in\mathbb{Z}}$ *is a* $\mathscr{F}_t$*-measurable process known as a white noise process (WN).*

3.1.1 Methodology for parametric risk measure forecasting

Now that we have established the structure for the negated log-returns, we can move on to discussing the methodology for estimating VaR and CVaR using parametric estimation techniques.

1. Define the time horizon h, the moving window of size m_w and the test size n_{test}: VaR and CVaR are typically estimated for a specific time period, such as one day, one week, or one month. In terms of risk management, we fix this time horizon to see on a given period, if the returns of a portfolio or a stock become risky or not. The moving window and the test size n_{test} are used for forecasting purpose. An alternative to the moving window is the use of a rolling window.

2. Investigating the AR-GARCH structure of the negated log-returns on the available data (in-sample) and forecast on the out-of-sample: we need to estimate the volatility and the mean on the out-of sample data and make assumption on the distribution that models the better the WN process $(Z_t)_{t\in\mathbb{Z}}$ to get the forecasted log-returns.

3. Estimating VaR and CVaR: once the volatility and the fitted distribution of $(Z_t)_{t\in\mathbb{Z}}$, given by its CDF F_Z, estimated adequately, we can compute the estimated VaR and CVaR, denoted $\widehat{VaR}_{\alpha}^{p}$ and $\widehat{CVaR}_{\alpha}^{p}$ respectively, on the out-of-sample data for the given period.

Methodology: Step 1.

In order to ensure effective risk management for a specific objective, it is important to fix and optimize the time horizon, moving window, and out-of-sample size (point 1. and 2. in the methodology review above). This is because the choice of these parameters can significantly impact the accuracy and reliability of the estimated risk measures.

- The *time horizon h* refers to the length of time over which the risk is being assessed. This may depend on the specific objective and risk appetite of the investor or institution. For example, a short-term trader may be more concerned with daily or weekly VaR estimates, while a long-term investor may be more interested in monthly estimates.
- The *moving window* refers to the number of observations used to forecast the future value of the log-returns volatility and mean by forecasting the conditional volatility and mean of the AR-GARCH process in order to compute VaR and CVaR. This number of data in the first window size is important because it determines the amount of information used to estimate those parameters and then the risk measures. A larger rolling window for the first forecasts may provide a more accurate estimate of risk, but it may also be more sensitive to changes in market conditions.
- The *out-of-sample size* n_{test} refers to the number of observations used to evaluate the accuracy and reliability of the VaR and CVaR estimates. This parameter is important because it ensures that the estimated risk measures are not overfit to the historical data used for estimation. A larger out-of-sample size can provide a more robust evaluation of the model, but it may also reduce the amount of data available for estimation.

Ultimately, the choice of these parameters will depend on the specific risk management objective and the characteristics of the data being used for estimation. It is important to carefully consider these parameters and optimize them to ensure that the estimated risk measures are accurate and reliable for the intended purpose.

Methodology: Step 2.

Once the time horizon, moving window beginning size and out-of-sample established, we can move to the step that aims to investigate the mathematical structure of the log-returns of the underlying portfolio or stock. This step can be challenging and is often the subject of much discussion and debate when forecasting risk measures. The goal of this step is to identify the appropriate AR-GARCH model parameters and distribution assumption that best fit the data, and ultimately obtain reliable estimates of VaR and CVaR. In this paper, we will focus solely on a standard interpretation of the log-returns as a AR-GARCH process as defined above. Specifically, we will consider three possible distribution assumptions for the WN component $(Z_t)_{t\in\mathbb{Z}}$: normal, t-student and skewed t-student. While these distributions differ in their shape and tail behavior, they are all commonly used in financial risk management and have been shown to provide reasonable fits to empirical data. It is worth noting that other interpretations of log-returns are possible, such as those that involve heavy-tailed or asymmetric distributions, but these will not be considered in the present work. AR-GARCH models are known to capture the volatility clustering and leptokurtosis often observed in financial returns data. However, the choice of AR-GARCH parameters and the distribution assumption can also affect the VaR and CVaR estimates.

We fix $\boldsymbol{\theta} = (\gamma_0, ..., \gamma_m, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)$, the vector of parameters we need to estimate. We assume that the WN component $(Z_t)_{t\in\mathbb{Z}}$ is either normal, t-student or skewed t-student . The assumptions on the distribution for the component $(\mu_t)_{t\in\mathbb{Z}}$ and $(\sigma_t)_{t\in\mathbb{Z}}$ that lead us to consider a $AR(m)$-$GARCH(p,q)$ process can be explained using a deep investigation on the empirical properties of the financial data using stylized facts. The main stylized facts tools are volatility clustering and auto-correlation. The main tests for univariate distributions for a general CDF are the Kolmogorov-Smirnov, Cramér-Von Mises and Anderson-Darling tests. Kurtosis and skewness are also strong tools to test for normal assumptions. There exists numerous ways to undertake this investigation. We recommend you to read the corresponding section in the book of Quantitative Risk Management [1].

Now that we are much more confident in our assumption as AR-GARCH model, we can fit our parameter of interest $\boldsymbol{\theta}$ using (conditional) maximum likelihood procedure given the in-sample available data and the moving window size. This estimator was found to be consistent under regularity conditions (see, e.g., Gourieroux, 1997), so even for non-normal WN with existing second moment, we obtain a quasi maximum likelihood parameter estimate $\widehat{\boldsymbol{\theta}}$. Hence, we can generate our AR-GARCH process with the fitted parameter to reproduce the log-returns of our portfolio on the out-of-sample data.

Methodology: Step 3.

The last procedure is to estimate VaR and CVaR on the generated data. Once we have the estimated parameters $\widehat{\boldsymbol{\theta}}$, we then have $(\widehat{\mu}_t)_{t\in\mathbb{Z}}$ and $(\widehat{\sigma}_t)_{t\in\mathbb{Z}}$ on the out-of-sample data. Hence we use the standardized residuals $\widehat{Z}_t = \frac{L_t - \widehat{\mu}_t}{\widehat{\sigma}_t}$ to compute the parameters of the assumed distribution for the WN $(Z_t)_{t\in\mathbb{Z}}$, F_Z, using again the Maximum likelihood method. We obtain finally $\widehat{F}_Z$ and we define:

$$\widehat{VaR}^{p}_{\alpha,t+h} = \widehat{\mu}_t + \widehat{\sigma}_t \cdot \widehat{q}_{\alpha,t+h}, \text{ and,}$$
$$\widehat{CVaR}^{p}_{\alpha,t+h} = \widehat{\mu}_t + \widehat{\sigma}_t \cdot \mathbb{E}[L_t | L_t \geq \widehat{VaR}^{p}_{\alpha,t+h}],$$

where $\widehat{q}_{\alpha,t+h}$ is the α-quantile of the fitted distribution $\widehat{F}_Z$ on the window period for the h-ahead observation.

3.2 Historical data

3.2.1 Historical simulation

Historical data can be used to estimate the distribution of the loss operator using empirical distribution of data. We can create a distribution of potential outcomes using a time period in the past and using the market data from that period. With this method we assume that we can predict the near future using the past. The historical data simulation is the easiest method to estimate $\widehat{\text{VaR}(L)}$. The main idea to estimate $\widehat{\text{VaR}(L)}$ is to take the negative log returns, to sort all of the values and then to just take the VaR of this empirical distribution of these negative log returns to get the VaR of $t+1$. The historical simulation makes the assumption that the negative log returns are i.i.d.. But when the volatility changes, the negative log returns are no longer i.i.d.. Actually this is kind of a naive approach for estimating VaR of $t+1$ because of course it will not predict any extreme value for the year $t+1$. So in the sense of the expectation it is logically a pretty good estimator but it will never predict any extreme values nor exceed any extreme values observed during the period of past years taken. There is actually a second version of this method, which is the Filtered Historical Simulation. This other version is more accurate because it filters the data and permits to estimate more efficiently VaR.

3.2.2 Bootstrap approach

Before introducing the Filtered Historical Simulation (FHS), we will present the Bootstrap technique that is used for the FHS. The idea of the Bootstraped historical simulation is to generate new returns by sampling, with replacement, the original negative log returns. In other words, we create a new vector of returns by selecting randomly a return of the original vector of negative log returns, and we repeat this process until we get a vector of the same size as the original vector of negative log returns. This process allows for the possibility of selecting the same observation multiple times and omitting others. We create a certain number of samples, and compute the VaR for each of them. At the end our estimate of VaR is the mean of these computed VaR. Note that the more samples we create, the more accurate we become. Indeed by obtaining more samples in the bootstrap method, we can better understand the various possible variations in the data, leading to improved accuracy in estimating the desired statistic or parameter. This approach relies on the Law of Large Numbers, which suggests that as the sample size increases, the sample mean approaches the population mean. In the bootstrap context, as we generate more bootstrap samples, our estimates tend to converge towards the true distribution of the statistic we are investigating. The same method applies to estimate ES, by computing the average of the losses in each resample exceeding the resample VaR, and then by taking the the mean of these ES estimates.

3.2.3 Filtered Historical Simulation (FHS)

Filtered historical simulation is a method that uses filtering techniques to increase the accuracy of VaR estimates. This method is based on the fact that extreme events and outliers in historical data can have a significant impact on risk estimates. FHS achieves this by performing a semi-parametric bootstrap within a GARCH framework, where the bootstrap preserves the non-parametric nature of HS, while the volatility model provides a sophisticated treatment of volatility dynamics.

In the first step we use the $AR(1)$-$GARCH(1,1)$ model defined in the section 3.1 .

In the second step, the fitted model is used to forecast volatility for each day in a sample period. The realized negative log returns are then divided by these volatility forecasts to obtain a set of standardized negative log returns. The standardized negative log returns should be i.i.d and thus suitable for HS.

Let's consider a 1-day VaR holding period. In the third step, we use bootstrapping on our dataset of standardized negative log-returns. To simulate future market conditions, we multiply each randomly selected sample by the GARCH forecast of tomorrow's volatility. By repeating this process and generating K simulated returns, we can capture a range of potential outcomes that reflect the current market conditions, as adjusted by the volatility forecast. Finally, each simulated return represents a possible value for the portfolio at the end of tomorrow, along with the corresponding potential loss. To estimate the VaR, we identify the loss that corresponds to our chosen confidence level. And we apply the same method as always for the CVaR.

Table 3.1 Pros and Cons of Non-parametric Approaches

Pros	**Cons**
• Intuitive and conceptually simple. • Can accommodate fat tails, skewness, and non-normal features. • Perceived empirical effectiveness, although mixed evidence. • Provide easily reportable and communicable results. • Confidence intervals for non-parametric VaR and ES can be produced.	• Results highly dependent on historical data set. • Difficulty handling shifts and slow to reflect major events. • Extreme losses in the data can dominate risk estimates. • Make no allowance for plausible but unobserved events. • Constrained by the largest historical loss, limiting handling of extremes.

3.3 Monte Carlo method

We are now going to present a methodology that bears resemblances to historical simulation, known as the Monte Carlo simulation. However, there exists a significant distinction between the two approaches. Unlike historical simulation, which utilizes observed market factor changes from the past, the Monte Carlo simulation involves the selection of a statistical distribution that effectively captures or approximates potential market factor variations. This process generates thousands of hypothetical changes in market factors. These changes are then employed to construct numerous potential profits and losses for the present portfolio, resulting in a distribution of possible outcomes.

Here, as in the parametric method and the FHS we use a $AR(1)$-$GARCH(1,1)$ to model our return. We will repeat our process with three distributions of the negative log returns in our training set: a Normal distribution, a Student's t distribution and a Skewed Student's t distribution. So for each of these three experiments, we will use the training set to determine the parameters of our $AR(1)$-$GARCH(1,1)$. Using the negative log return formula of this model we will be able to deteremine the negative log return of the test set. The Monte Carlo method implies that we will simulate the path of the negative log returns of the test set a large number of time, let say M simulations. As in the FHS method we will use previous returns to determine the possible returns for each of the following days. Finally we use the training set to determine the VaR of the first day of the test set. We just compute the VaR on the training set. For the second day we create M vectors composed of the training set values and every possible return of the first day as the last entry of each vector. We thus have now M new "training sets" with the possible values of the first day. Then we compute the VaR for the M vectors, and a possible VaR of the second day would be the average of these VaR. For the third day we just add the vector of possible returns of the second day to our new training set and apply again the same process. We repeat this action until we have computed the VaR of all days in the test set. For the Conditional Value at Risk we would just compute the CVaR for the M possible returns for each day and take take the average. Mathematically we get:

$$\widehat{VaR}^{MC}_{\alpha,t+h} = \frac{1}{M}\sum_{k=1}^{M} VaR^{k}_{\alpha,t+h}$$

$$\widehat{CVaR}^{MC}_{\alpha,t+h} = \frac{1}{M}\sum_{k=1}^{M} \mathbb{E}[L_t | L_t \geq VaR^{k}_{\alpha,t+h}]$$

where $VaR^{k}_{\alpha,t+h}$ is the value at risk of the k-th possible path of log returns at time $t+h$, and M the number of iterations used.

However, it's important to understand that the accuracy and dependability of VaR estimates derived from Monte Carlo simulation rely on various factors. These factors include the accuracy of the input parameters, the choice of a suitable statistical distribution, and the calibration of the model.

Chapter 4

Comparison and back-testing for risk management

Once a risk model has been developed, it is crucial to conduct thorough validation before implementing it in practical settings. Subsequently, regular evaluation of the model's performance is necessary. An integral aspect of model validation is back-testing, which involves using quantitative techniques to assess whether the forecasts produced by a VaR or ES forecasting model align with the underlying assumptions on which the model is built. Back-testing can also be used to compare and rank a set of such models. We will define in this chapter some of the techniques used in practice. The first set of techniques, so called traditional back-testing aims at analysing the goodness of fit of each model but not at comparing them. They test for the hypothesis H_0 : " The model is correct ". It is often the case that we are not just interested in how individual models perform, but also in how different models compare to each other. This is why we will define a second set of techniques, known as comparative back-testing which will allow comparitive analysis between the forecasting models employed.

4.1 Traditional Back-testing

4.1.1 The Basic Frequency Backtest (Kupiec)

Before presenting the basic frequency backtest also called the Kupiec proportion of failure test, we will shortly present an even simpler test. Basically the simplest backtest is to count the number of losses larger than the estimated VaR and comparing it to the number of losses larger than the actual VaR on a given time period. The Kupiec Test, proposed in 1995, evaluates the accuracy of VaR models. It focuses on unconditional coverage, assessing if VaR is violated more or less than $\alpha \times 100\%$ of the time. Kupiec's test, called the proportion of failures (POF) test, counts the number of VaR violations over a period. If violations significantly differ from $\alpha \times 100\%$, the risk model's accuracy is questioned. The test uses a sample of T observations and calculates a test statistic. Let's define a tool that will be useful for this test, the hit function: $H_{t+1}(\alpha) = 1$ if $x_{t+1} \leq -VaR_t(\alpha)$ and $H_{t+1}(\alpha) = 0$ if $x_{t+1} > -VaR_t(\alpha)$. Then the Kupiec test statistic test is the following:

$$K_{pof} = 2\log\left(\left(\frac{1-\overline{H}(\alpha)}{1-\alpha}\right)^{T-H(\alpha)}\left(\frac{\overline{H}(\alpha)}{\alpha}\right)^{H(\alpha)}\right)$$

where $\overline{H}(\alpha)$ is the ratio of the hitting function and the time period: $\overline{H}(\alpha) = \dfrac{H(\alpha)}{T}$ and $H(\alpha) = \sum_{t=1}^{T} H_t(\alpha)$.

The test statistic derived from the Kupiec test indicates the adequacy of the VaR measure. If the proportion of violations matches α, the backtest statistic is zero, suggesting no inadequacy in the VaR measure. However, as the proportion deviates from α, the backtest statistic increases, indicating potential understatement or overstatement of the portfolio's risk level. The Kupiec test is closely linked to market risk capital requirements. As the number of VaR violations increases, signaling a higher degree of risk underestimation or overestimation, the market risk capital multiplier also grows. This multiplier is a regulatory tool, ensuring sufficient capital reserves in light of the observed VaR violations. If the VaR model's accuracy is compromised, corrective measures are required. While the Kupiec test is a well-known example of a VaR backtest, it has its limitations. Firstly it may exhibit low power in detecting VaR measures that systematically underreport risk, especially with smaller sample sizes like one year. This can lead to false reassurances about the accuracy of VaR models. Another limitation of these tests is that they solely concentrate on the unconditional coverage property and may not detect VaR measures with dependent violations. When there is dependence in VaR violations, it suggests that the risk management practices are inadequate and the portfolio fails to respond appropriately to changing market conditions.

4.1.2 The Conditional Testing Backtest (Christoffersen)

The conditional backtesting method, proposed by Christoffersen as a complement to the Kupiec test, is a valuable approach to improving the evaluation of risk prediction models. By incorporating this method into our analysis, we can gain a more complete understanding of model performance. It allows us to evaluate specific predictions, such as the accuracy of event frequency and the independence assumption. As a result, we can make more informed risk management decisions and refine our strategies more effectively.

Suppose we have a risk forecasting model that predicts independent and identically distributed (i.i.d.) exceedances. The Chritoffersen approach involves isolating and testing each specific prediction separately. The first prediction examined is whether the model generates the correct frequency of exceedances, which is the method done in the above subsection. Another prediction that conditional backtesting examines is the independence of exceedances. This prediction suggests that exceedances should not cluster together over time but instead occur randomly. In simpler terms, it means that the occurrence of exceedances should not show any patterns or trends. If there is evidence of clustering, it indicates a potential problem with the model's assumption. If there is evidence of clustering, it indicates a potential issue with the model's specification, even if it passes the prediction of correct unconditional coverage. By conducting separate tests for each prediction, conditional backtesting allows us to gain insights into the model's performance and identify any potential weaknesses or deficiencies. This approach helps us ensure that the model provides accurate frequency predictions and adequately accounts for the independence of exceedances. We already have defined the basic frequency test, so turning to the independence prediction, let T_{kj} be the number of days where state j occurs after state k occurred the previous day, where the states refer to exceedances/non-exceedances. Let p_{kj} be the probability of state j on any given day, given that the previous day's state was k. Let just observe that the statistic test of Kupiec (which is by the way in a likelihood form) is distributed as a $\chi^2(1)$. The statistic test under the hypothesis of independence takes the form of:

$$C_{ind} = 2\ln\left[\frac{(1-\hat{p}_{01})^{T_{00}}\hat{p}_{01}^{T_{01}}(1-\hat{p}_{11})^{T_{10}}\hat{p}_{11}^{T_{11}}}{(1-\hat{p}_2)^{T_{00}+T_{11}}\hat{p}_2^{T_{01}+T_{11}}}\right]$$

This test statistic also follows a $\chi^2(1)$ distribution. Note that we can estimate the probabilities as follows:

$$\hat{p}_{01} = \frac{T_{01}}{T_{00}+T_{01}}, \hat{p}_{11} = \frac{T_{11}}{T_1+T_{11}}, \hat{p}_2 = \frac{T_{01}+T_{11}}{T_{00}+T_{01}+T_{10}+T_{11}}$$

Thus under the combined hypothesis of correct coverage and independence the test statistic LR_{cc} is given by:

$$C_C = K_{pof} + C_{ind}$$

and follows then a $\chi^2(2)$. For more information on this method, see the paper from Zhang, Y. [9].

4.2 Comparative Back-testing

In this section we aim to explain how to compare the efficiency of the different models of forecasting. We will see in the next sections the experimental results. We start with the forecasting of VaR. Let's recall that we proved that VaR is elicitable through a scoring function. The scoring function we will use to compare the different models is $S(x,y)) = (\mathbb{1}_{x \geq y} - \alpha)(x - y)$. What we do here is that we take y as the observed VaR and x as the forecasted VaR. So let suppose we forecasted K value of VaR for each of our Parametric model, Filtered Historical Simulation model and Monte Carlo Model. Thus for each model we will get K different values of $S(x,y)$. Then what can be done is to take the empirical mean of the scoring function of these three models and rank them. Remembering that elicitability of VaR means that the actual VaR is the only value minimizing the scoring function, we can claim that the lower the empircical mean of the scoring function, the better.

It has been pretty easy to rank the methods used to forecast VaR as VaR is elicitable, but it is a much more complex problem to do so for ES, as it has been proved that it was not elicitable. But an important result that was stated in this paper is that ES is jointly elicitable with VaR, so we will see how to use this result in order to compare the models. The idea is that, as we can not rank the models using only the ES simulated, we want to associate the VaR to build some pairs (VaR, ES) and then rank them using a scoring function. It is possible to demonstrate that any scoring function for the pair $(\mathrm{VaR}_\alpha(Y), \mathrm{ES}_\alpha(Y))$ can be expressed in the following form:

$$\begin{aligned} S(x_1, x_2, y) = {} & (\mathbb{1}_{y \leq x_1} - \alpha) f_1(x_1) - \mathbb{1}_{y \leq x_1} f_1(y) - \mathscr{F}_2(x_2) \\ & + f(y) + f_2(x_2)(x_2 - x_1 + \frac{1}{\alpha} \mathbb{1}_{y \leq x_1}(x_1 - y)) \end{aligned}$$

where the functions f_1, f_2, $\mathscr{F}_2$, and f need to be carefully selected. The first two terms are closely associated with the scoring function linked to VaR, while the remaining part of the expression cannot be separated into a sum of components that solely depend on x_1 (the estimated value of VaR) or x_2 (the estimated value of ES). This highlights the fact that it is impossible to derive a scoring function for ES alone; knowledge of VaR is indispensable. We can adopt a specific scoring function S_1, with $f_1(x) = x$, $f_2(x) = \frac{e^x}{1+e^x}$, $\mathscr{F}_2(x) = \log(1 + e^x)$, and $f(x) = 0$, that will be used in the next section. To see in depth, look at the paper from Dendoncker, V. and Lebègue, A.: [2].

Chapter 5

Numerical experiment: Portfolio risk analysis

Now let us explore a data analysis of the CAC40 index that will be the only stock in our portfolio. Our data set contains the daily closed price of the index from 1990-03-01 to 2023-04-03 which gives us 8370 observations. We compute and add to the data the negative log-returns given by the following formula:

$$L_t = -\log\left(\frac{C_t}{C_t - 1}\right), \quad t = 2, \ldots, 8370 \text{ and}$$

where $(C_t)_{t\in\mathbb{N}}$ is the closed price of the index on day t. Thus we have a total of 8369 negative log-returns. We use a moving window of size $m_w = 669$ and $n_{test} = 7700$ for our forecasting and the backtesting. The choice of the size of the moving window allows to get a wide range of observations to get more accurate forecast. It corresponds to roughly 2.5 years of trading data.

5.1 Exploratory data analysis of the CAC40 index

Prior to examining the VaR and CVaR values of the CAC40 index, we first delve into the whole data to offer a comprehensive explanation of the assumptions we made in section 3. Specifically, we analyze the negative log-returns on the entire data set to gain insights.

First of all, the daily returns of the index varies between -12.28% to $+11.18\%$ with a mean of 0.03% and a standard deviation of 1.37%. As explained previously and as shown on the graph in Fig. 4.1, the daily returns have more volatility in clusters.

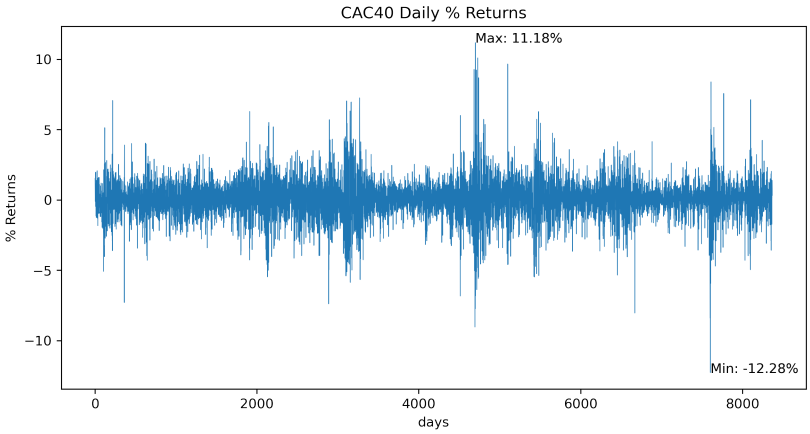

Fig. 5.1 CAC40 daily percentage returns.

Since we are working with the negative log-returns in this paper, we present below the corresponding time series.

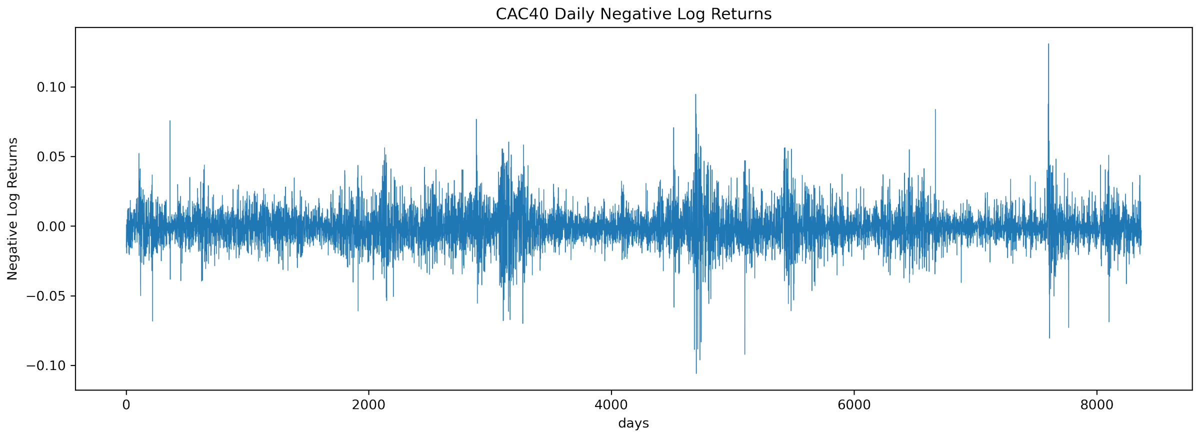

Fig. 5.2 CAC40 daily negative log-returns.

We assumed that the time series follows an AR(1)-GARCH(1,1) process, and the plot above together with tests on our side support this conclusion. Now, let us examine the distributions of the innovations (WN) of the negative log-returns. We made three assumptions regarding the latter. We considered modeling the innovations using either a normal, t-student, or skewed t-student distribution.

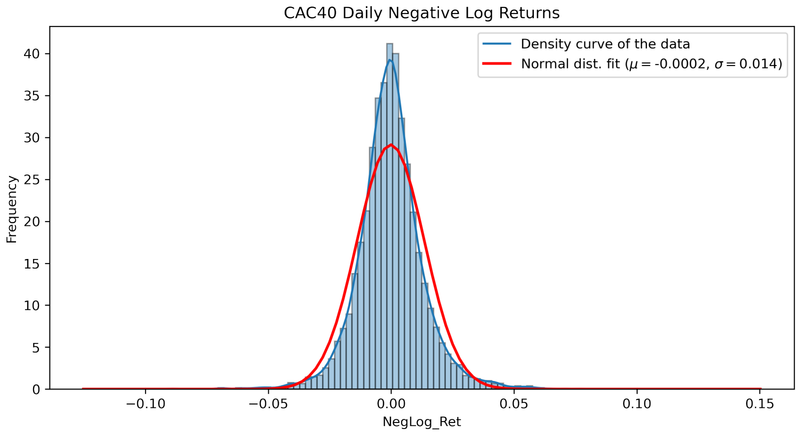

Fig. 5.3 Histogram of negative log-returns vs fitted normal distribution.

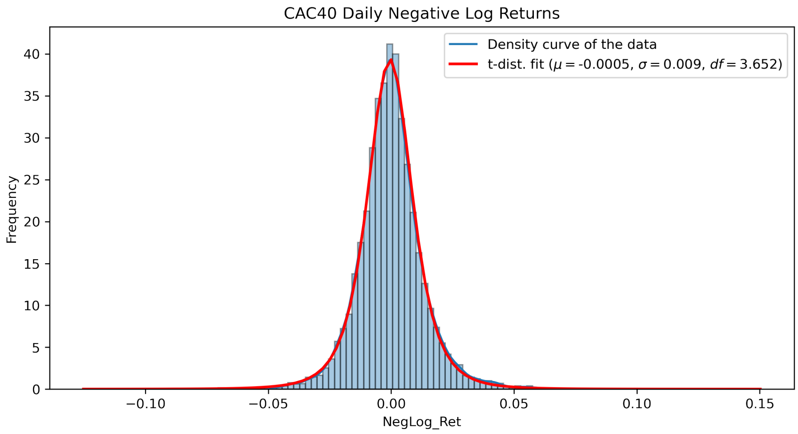

Fig. 5.4 Histogram of negative log-returns vs fitted t-student distribution..

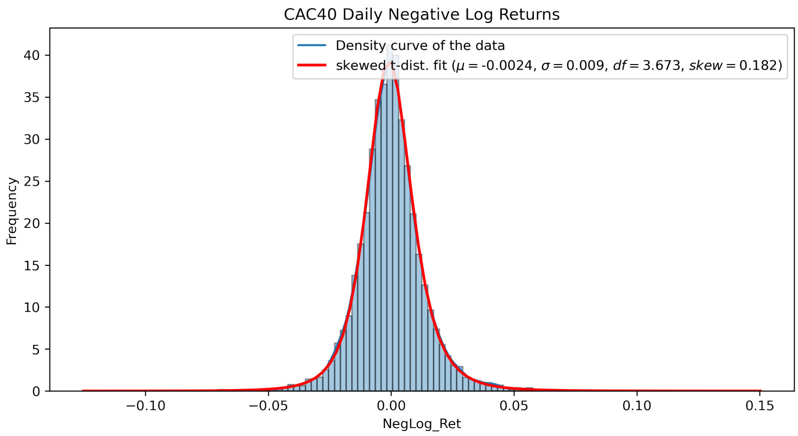

Fig. 5.5 Histogram of negative log-returns vs fitted skewed t-student distribution..

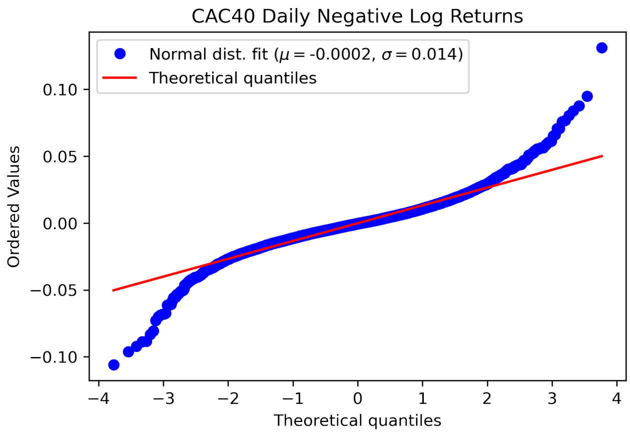

Fig. 5.6 Histogram of negative log-returns vs fitted normal distribution.

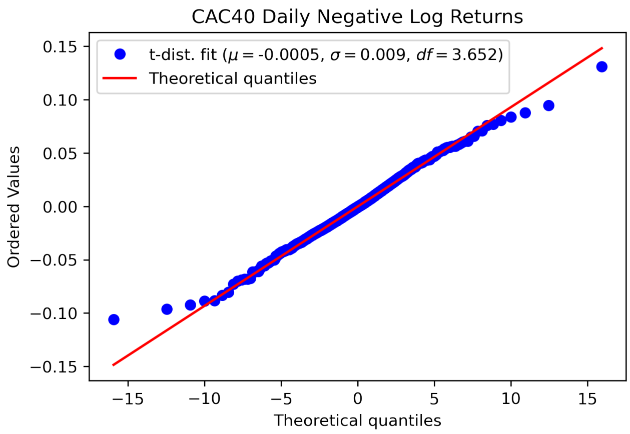

Fig. 5.7 Histogram of negative log-returns vs fitted t-student distribution..

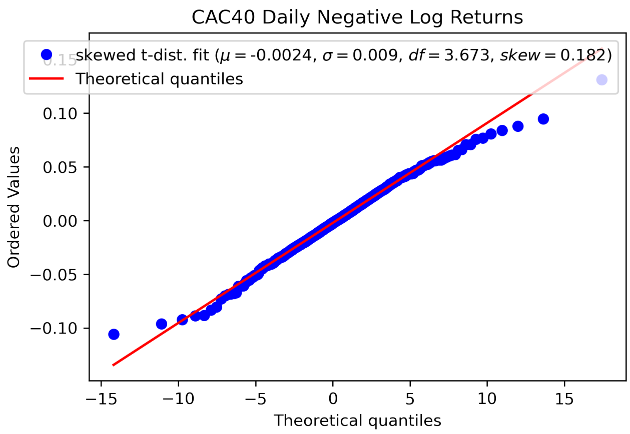

Fig. 5.8 Histogram of negative log-returns vs fitted skewed t-student distribution..

Kolmogorov-Smirnov (K-S) test results:

Table 5.1 Kolmogorov-Smirnov Test Results

Distribution	Statistic	p-value
Normal distribution	0.064	$< 10^{-3}$
t-distribution	0.009	0.507
Skewed t-distribution	0.007	0.795

- **Normal distribution:**

  The K-S test statistic for the normal distribution is 0.064, indicating the maximum vertical distance between the CDF of the sample and the theoretical normal distribution CDF. The extremely small p-value ($< 10^{-3}$) suggests strong evidence against the null hypothesis, indicating that the innovations do not follow a normal distribution.

- **t-distribution:**

  For the t-distribution, the K-S test statistic is 0.009. The relatively large p-value (0.507) suggests that there is insufficient evidence to reject the null hypothesis. Thus, we can conclude that the innovations are consistent with a t-distribution.

- **Skewed t-distribution:**

  The K-S test statistic for the skewed t-distribution is 0.007. The high p-value (0.795) indicates that there is no significant evidence to reject the null hypothesis. Thus, the sample is likely to follow a skewed t-distribution.

Note that we applied exactly the same process on the training data in order to be fair in the assumptions made. We do not present it here since it is a replication of what we presented just above but on the training set.

5.2 VaR and CVaR analysis

The VaR and ES forecast were estimated as explained above for the negative log-returns of the CAC40 index. We have collected the average value of the forecasted one-ahead period VaR and ES for each method of estimation. The parametric average forecasts are $\overline{VaR}^p$ and $\overline{CVaR}^p$ and the results for $\alpha = 0.99$ are given is the table 4.2 bellow.

We added the percentage of violation of VaR which reprensents the percentage of value of negative log-returns of the CAC40 index that exceed the forecasted VaR for each distribution.

Table 5.2 Risk measure results for $\alpha = 0.99$

Distribution	$\overline{VaR}^p$	% of violation	$\overline{CVaR}^p$
Normal distribution	2.88	2.14 %	2.90
t-distribution	3.07	1.50 %	4.29
Skewed t-distribution	3.25	1.22 %	3.33

The results confirm the fact that the normal distribution does not take into account of the fat tails coming from the fact that financial returns show excess of kurtosis. We get more precise with student and skewed student distributions. However, they also have their problems. Like the normal, they fail to respect constraints on maximum possible losses, and can produce misleadingly high risk estimates as a result. This can be observed on the following plot:

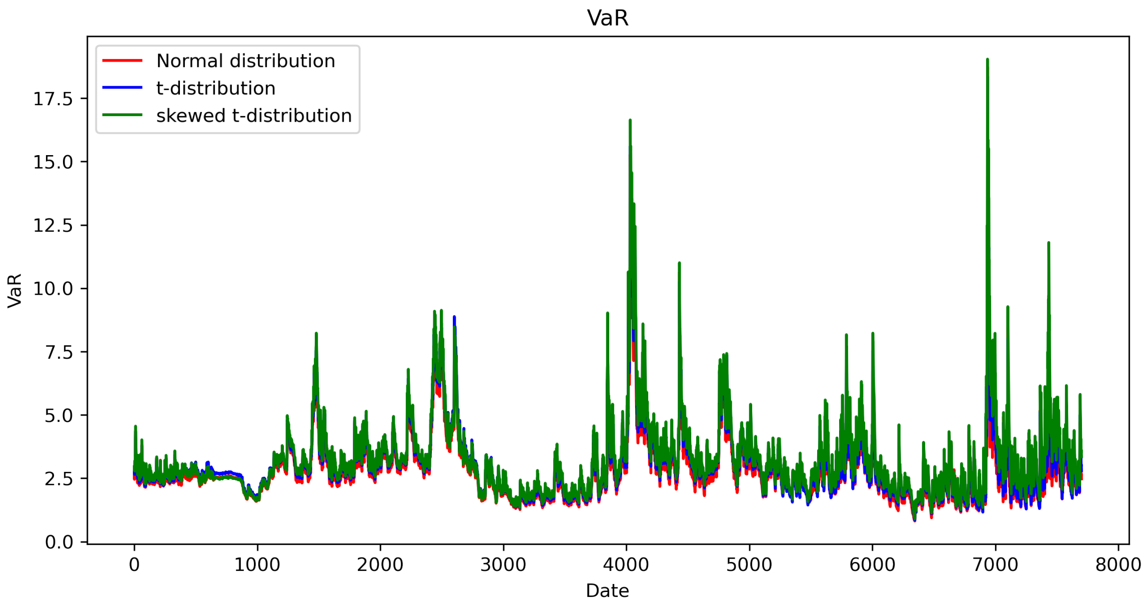

Fig. 5.9 VaR forecasted on the test set for the three assumptions on the WN: normal, student and skewed student.

Also, it is important to notice, if one wants to create a portfolio with different stocks where returns are t-distributed then it will suffer of the non-stability of it (i.e., the sum of two or more t-distributed random variables is not necessarily distributed as a t variable itself). However for the CAC40 index return, we observe that the models are badly able to hit nearly exactly the expected proportion of violations by matching the risk measure level given by $1-\alpha = 0.01$ in our case. In fact, the better is 1.22% of violation for the skewed student distribution which is, according to the previous EDA, the best assumption for the WN process.

5.2.1 Traditional Back-testing

We have applied the Kupiec and Christofferson tests on the results of the parametric and FHS method to test the correctness of the VaR model where we forecasted the VaR of the negative log-returns at a level $\alpha = 0.99$ with the different assumptions on the WN and got the follwing results:

Kupiec test: The null hypothesis H_0 testing if the VaR model is correct fails for the case where the WN are normally or student distributed and is accepted for the skewed student distribution for the parametrics models. That means that we have, at a level of 95% no reason to suppose that the VaR forecasted of the negative log-returns for normal or student distributed WN are good to predict an accurate VaR level but we can say that it is when the WN are skewed student distributed. This result is in line with what we observed in the EDA. In case we use FHS models, all of the tests accept that the model is correct.

Christoffersen test: For this test, we have the same results as for the Kupiec test but slightly greater. That means that the independence of VaR violation is accepted in every models. However, since the Christofferson test is the combination of the Kupiec static and the independce statistic, we have the same results in term of accepting or not the correctness of the model as for the ones in the Kupiec test.

5.2.2 Comparative Back-testing

To back-test CVaR we used the method of comparative back-testing with the specific scoring function described in the section 4.2. To this aim we have estimated the joint (VaR, CVaR) at a level $\alpha = 0.975$ and back-test them accordingly. As a result we obtain a vector of scores for each method and assumption on the WN. We look at the average of those scores. For example, if we take the parametric method with normal distributed WN. We computed VaR and ES on the test set and for each day in the test set we compute a score given by the formula in the section 4.2. We then obtain a vector of scores and we report the scaled average $\frac{1}{1-\alpha}\overline{S}$ variable. See table 5.3 and 5.4 below. The results for (VaR, ES) at level $\alpha = 0.975$ suggest that the performances are better when we use a skewed student assumptions on the WN to fit the AR(1)-GARCH(1,1) model of the negative log-returns. We also notice that the parametric method is overall less performant than the FHS method.

Table 5.3 Comparative back-test Results for parametric method

Distribution	$\overline{S}$
Normal distribution	3.47
t-distribution	3.45
Skewed t-distribution	3.43

Table 5.4 Comparative back-test Results for FHS

Distribution	$\overline{S}$
Normal distribution	3.42
t-distribution	3.42
Skewed t-distribution	3.43

Chapter 6

Conclusion

Through this project we have analyzed the risk measures of Value at Risk (VaR) and Conditional Value at Risk (CVaR), both theoretically and through experiments. We looked at various forecasting methods, including parametric methods, historical simulation methods and Monte Carlo methods. Our aim was to understand their advantages and disadvantages, and to evaluate their performance using backtests. The theoretical part of the project involved understanding the mathematical properties of VaR and CVaR as risk measures. We saw that VaR is not consistent but comonotonically additive, law-invariant and elicitable, while CVaR is consistent, convex, a risk adverse measure, conditionnaly elicitable but not elicitable. Since elicitability allows us to rank different values of a risk measure using a score function, it is not possible to backtest only the CVaR. But we have seen that VaR and Cvar are jointly elicitable, so we can test CVaR using VaR. Then, we simulated negative log returns with our three different forecasting methods and applied VaR and CVaR measures. Each method had its strengths and weaknesses. Parametric is efficient and suitable for large datasets but this methods assume specific patterns in the data, which is not always accurate. Filtered historical simulations, in contrast to historical simulations, made it possible to capture extreme values while using historical data, by putting weights on certain data thanks to its semi-parametric nature. Monte Carlo simulations are flexible in capturing complex relationships and dependencies, but require many simulations to be reliable. To check the effectiveness of these methods, we tested them against real losses (Backtesting). This enabled us to assess their accuracy and reliability. By comparing estimated VaR and CVaR values with actual losses, we were able to assess the ability of these measures to predict and capture risk. Overall, this project aimed to understand and evaluate VaR and CVaR measures, the different forecasting methods and their practical implications in risk management. By studying their theoretical properties and testing them in simulations, we gained a better understanding of how they work, their strengths and limitations.

References

[1] A.J. McNeil, R. F. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton Series in Finance.

[2] Dendoncker, V. and Lebègue, A. (2019). *Comparative Backtesting of the Expected Shortfall*. Reacfin.

[3] Embrechts, P. and Wang, R. (2015). *Seven Proofs for the Subadditivity of Expected Shortfall*. doi 10.1515/demo-2015-0009.

[4] for International Settlements, B. (2017). *Basel Committee on Banking Supervision*. (www.bis.org).

[5] Nolde, N. and Ziegel, J. (2017). *Elicitability and backtesting: Perspectives for banking regulation*. The Annals of Applied Statistics.

[6] R. Kieselb, R. Ruhlickeb, G. S. J. Z. (2015). *Conceptualizing Robustness in Risk Management*. https://ssrn.com/abstract=2065723.

[7] S. Emmer, M. Kratz, D. T. (2015). *What is the best risk measure in practice? A comparison of standard measures*. arXiv:1312.1645v4 [q-fin.RM] 28 Apr 2015.

[8] S. Sarykalin, G. S. and Uryasev, S. (2008). *Value-at-Risk vs. Conditional Value-at-Risk in Risk Management and Optimization*. doi 10.1287/educ.1080.0052.

[9] Zhang, Y., . N. S. (2017). *A review of backtesting for value at risk*. Communications in Statistics - Theory and Methods, 1-24. https://doi.org/10.1080/03610926.2017.1361984.