A comparison between Value at Risk forecasting methodologies

Department of Economics and Finance

Course of Empirical Finance

Supervisor: Prof. Santucci De Magistris Paolo

Co-supervisor: Prof. Carlini Federico Carlo Eugenio

Candidate: Imperi Galli Alessio

ID 717631

Academic Year 2021/2022

Abstract

This thesis compares the forecast accuracy of the GARCH and CaViAR class models for calculating VaR. The results show that the Caviar models outperform the GARCH models, particularly for the most extreme quantiles. For both classes of models, asymmetrical specifications achieve more accurate results. The analysis will be conducted by estimating the models using log-likelihood functions for GARCH models and Regression Quantiles for CaViAR. Forecasts will be backtested using different approaches. In the first part, the theoretical framework will be outlined, briefly addressing VaR's history, uses, and critics. Subsequently, the models used from a theoretical point of view and the empirical results will be presented.

Contents

Introduction

On 19 October 1987, after a growth of more than 44% in the first months of year, the Dow Jones Industrial Average suffered a record loss of 22.6%. The increasing occurrence of crises at the global level pushed countries worldwide to seek standard prevention measures. The use of financial instruments capable of triggering the exchange of securities on the occurrence of certain conditions made it clear that new risk measures needed to be developed. For a financial institution, estimating the risk of a position is a crucial issue. One possible way to measure risk is to consider the size of the loss that security may incur if its price falls. The Value at Risk is the greatest amount an investor can lose with a given probability over a certain period. Since its introduction as part of the 1988 Basel Accords, which mandated the use of VaR as the basis for calculating banks' capital requirements, VaR has become the financial world's principal measure of risk. As witnessed by the intense debate over methods to forecast the VaR, the accuracy of VaR estimates still remains a major issue in finance. This issue becomes even more relevant during crisis, when investors and banks may be more likely to resort to reserves to offset losses. To ensure the solvency of the financial system, regulators provided some ground rules for the minimum capital that financial institutions are obliged to hold as reserves to face possible loss scenarios. VaR offered several opportunities:

- it reflects current market conditions, which are contained in the P&L distribution;
- it can be computed for options, swaps, and other derivatives, even if it is simpler for stocks and bonds;
- it can be computed for a portfolio made by different instruments, even if it is not sub-additive.

However, VaR has also some drawbacks, both due to its most popular calculation approaches (i.e., assumption of normality) and intrinsically related to its definition. This thesis aims to compare the accuracy of one-day ahead forecasts of the VaR. To carry out this analysis, seven models of the classes GARCH and CaViaR, in different specifications, will be considered. The comparison of the accuracy of the forecasts will be conducted employing some of the most popular backtests. The analysis will assess the differences in accuracy of the various models on different assets in the 3000

daily observations period from 02/01/2009 to 01/12/2020. The sample includes the economic crisis generated by the spread of the Covid-19 outbreak, which witnessed a substantial increase in volatility. The analysis will be conducted on four assets: SP500, WTI, Euribor3m, and EurUsd. The comparison results show that CaViaR models provide more accurate forecasts than GARCH models. In particular, for both classes of models, the asymmetric models (i.e., taking into account the leverage effect) outperform the others.

Chapter 1

Risk Management and Value at Risk

1.1 Literature Review

Risk measurement is a significant issue for financial institutions. The need to develop a measure of risk has become increasingly apparent with the prominent role of finance within the economy. To proceed with the analysis, it is first necessary to introduce the notion of risk. Raiffa and Luce (1957)'s definition of risk distinguishes between risk, uncertainty, and certainty. If the decision-maker knows the probabilities associated with the outcome of the decision, then he is in the presence of risk. In the case in which such probabilities are not known, the situation is uncertain. Finally, if the decision will inevitably lead to a determined known outcome, one is in a situation of certainty. Although risk management within companies has always been a subject of study, it was not until the 20th century that it became a proper discipline. The first academic books on Risk Management were published by Mehr and Hedges (1963) and Williams and Heins (1964) and mainly dealt with pure Risk Management. With the spread of the studies of Markowitz, Lintner (1965), Treynor (1961), Sharpe (1966), and Mossin (1966), which later merged into the Capital Asset Pricing Model, Risk Management focused on developing more sophisticated mathematical tools to assess risk. As risk management became quantitatively oriented, formal risk definitions became widespread. In Kaplan and Garrick (1981)'s definition, risk can be defined as a "set of triplets": scenario, probability, and consequences. These answer the questions: what is likely to happen? How likely is it to happen? What are the consequences?
To answer these questions, it is possible to develop a list of possible outcomes or scenarios, described in Table 1.1

Table 1.1: Scenario List

Scenario	Likelihood	Consequence
S_1	p_1	x_1
S_2	p_2	x_2
.	.	.
.	.	.
.	.	.
S_N	p_N	x_N

Where i-th raw can be intented as

$$R = \{\langle s_i, p_i, x_i \rangle\} \tag{1.1}$$

R denotes the risk represented by the scenario s_i, its probability p_i, and the associated outcomes x_i.

This paper will focus on a particular type of risk, called market risk, which is defined by the European Banking Authority as:

> "The risk of losses in on and off-balance sheet positions arising from adverse movements in market prices".

In the 1970s and 1980s, due to the growing popularity of derivatives, Financial Risk Management had become a priority for financial institutions, increasingly exposed to price and interest rate fluctuations. As highlighted by Dionne (2013) risk management included in its area of focus the effect of financial decisions on the overall value of a company or portfolio.

Despite the intense scientific debate around risk management, there were still no shared and uniformly respected practices within the entire financial circuit capable of preventing or containing possible systemic financial crises. The market crash of 1987, which originated in Asia and then spread throughout the world, also known as "Black Monday", had one of its roots in the steady increase in volatility recorded in the months preceding October 1987. A new financial product called portfolio insurance had spread rapidly, involving extensive use of derivatives and options. These contributed to accelerating drastically the crash's pace due to the selling mechanisms triggered by the initial losses. On October 19th 1987, after several days of losses and extreme selling pressure on stocks, the Dow Jones Industrial Average reported a loss of 22.6%.

The spread of the crisis throughout the world emphasized the need to take shared measures between the various countries to cope with phenomena of such magnitude. In the following years, the world's major countries harmonized their provisions and intensified cooperation activities. The Basel agreements were certainly the primary efforts in this direction. The first agreement, included in the Basel 1 1988 for banking regulation, was limited to credit risk. It required each bank to hold a reserve equal to 8% (Cook ratio) of the value of securities to mitigate the credit risk of its portfolio. Alongside this provision, internal models for calculating VaR were developed as an additional element of risk assessment of the bank's financial position. According to JP Morgan's RiskMetrics Technical Document of 1996:

> "Value at Risk is a measure of the maximum potential change in value of a portfolio of financial instruments over a pre-set horizon. VaR answers the question: how much can I lose with α% probability over a given time horizon."

To clarify this definition it is possible to give an example: a 95% one-day Var is the amount that would be lost the next day with a probability of 95%. To provide a formal definition of the VaR, it is necessary to introduce its components first. Let

$$\Delta P_t(\tau) = P(t+\tau) - P(t) \tag{1.2}$$

denote the change in the value of a portfolio between time t and time $t+\tau$.
Let $\alpha \in (0,1)$ denote the probability level, then formally the VaR is defined as

$$\begin{gathered} Pr[\Delta P(\tau) < -VaR] = 1-\alpha \text{ or} \\ VaR(\alpha) = \inf\{v|Pr(r_t \leq v) = \alpha\}. \end{gathered} \tag{1.3}$$

Let ΔP have a cumulative distribution function F

$$F_{\Delta P}(x) = Pr(\Delta P \leq x). \tag{1.4}$$

Assuming that $F_{\Delta P}$ admits a density, the VaR is the α-th quantile associated with the distribution of the portfolio returns. It is then possible to write:

$$1-\alpha = \int_{-\infty}^{VaR_\alpha} f_{\Delta P(x)}\, dx. \tag{1.5}$$

In this first version of the Basel agreement, the VaR was calculated as the sum of the VaRs relating to asset return risk, interest rate risk, exchange risk, and commodity place risk. Computing VaR in this way did not consider the effect of diversification to mitigate risk. The document also specified some further rules for the calculation of VaR:

- time horizon was at ten market days or two weeks;
- confidence level was fixed at 99%;
- historical data used to produce estimates had to go back at least one year, with an update of the parameters every three months.

Considering the arbitrariness of setting the reserves at 8% and the absence of mechanisms designed to encourage diversification as a strategy for risk mitigation, VaR was quickly preferred to the Cook Ratio. Over the years, the simplicity of VaR, which expresses in a single number in monetary or percentage terms the likely loss of a portfolio, has made it the international standard for measuring financial risk.
Over the past 18 years, there have been three regimes for calculating market risk capital requirements: Basel 2, Basel 2.5, and Basel 3. The evolution of the international banking regulation is summarized in Fig 1.1. However according to Gneiting (2011), to understand the usefulness and popularity of VaR, it is helpful to consider these regimes briefly.
In 1996, the Basel 2 accords allowed banks to use their own "internal" models to calculate regulatory capital. These models included 99% VaR with a 10-day horizon. Once the latter was calculated and "validated" through some backtesting procedures, including the Traffic Light test introduced by the Commission, the regulatory capital was calculated as follows:

$$CA_t^{B2} = max(VaR_{t-1}, m_c\overline{VaR}_{t-1}), \tag{1.6}$$

where m_c is a multiplicative factor which is assigned according to the "traffic light" system. VaR_{t-1} is the VaR of the previous day and $\overline{VaR}_{t-1}$ is the average VaR of the previous 60 days. The multiplication factor is at least 3 but can take on greater values as the inaccuracy of the model (measured by the backtests) increases. After the outbreak of the 2008 crisis, the standards imposed by Basel 2 appeared inadequate to measure systemic risk. This resulted in maintaining inadequate levels of capital buffers to withstand a crisis. Therefore, Basel 2.5 introduced the Stressed VaR, which was a VaR calculated during 12 months of financial stress. The calculation of capital requirements was therefore modified:

$$CA_t^{B2.5} = max(VaR_{t-1}, m_c\overline{VaR}_{t-1}) + max(SVaR_{t-1}, m_s\overline{SVaR}_{t-1}). \tag{1.7}$$

The first term is the same from Basel 2. $SVaR_{t-1}$ is the Stressed VaR of the previous day, and $\overline{SVaR}_{t-1}$ is the average Stressed VaR for the previous 60 days. Given that Stressed VaR is at least equal to VaR and assuming that the two multipliers are equal, this change resulted in at least a doubling of the capital buffers under Basel 2. Finally, under Basel 3, it was decided to use Expected Shortfall as the measure to base capital requirements because it considers both the likelihood of loss and its value. The ES is more difficult to estimate due to the lack of the property of elicitability, i.e., the property whereby a risk measure is detectable as the solution to the problem of minimizing an expected loss function. Anyway, the ES is both consistent and coherent, while the VaR is neither. To understand why the VaR is not coherent, it is first necessary to introduce briefly the notion of risk measure.

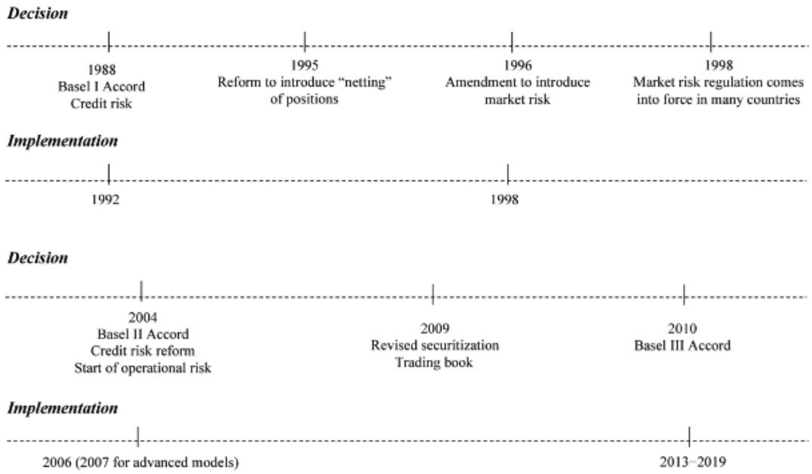


Figure 1.1: Evolution of International Bank Regulation

Let L be a space of random variables. A risk measure is a decreasing function that links the future risky position $X \in L$ to the minimal amount $\rho(X)$ that should be collected to cover the risk X. A risk measure ρ over a random variable X, is said to be coherent if it satisfies the following conditions:

- Homogeneity: given $h \geq 0$

$$\rho(h, X) = h\rho(X); \tag{1.8}$$

- Sub-additivity: given X_1, X_2;

$$\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2); \tag{1.9}$$

- Monotonicity:

$$X_1 \leq X_2 \Rightarrow \rho(X_1) \leq \rho(X_2); \tag{1.10}$$

- Translation invariance: given $a \in \mathbb{R}$

$$\rho(X - a) = \rho(X) - a; \tag{1.11}$$

The VaR is not a coherent risk measure, since it can be shown that sub-additivity is not always guaranteed. As demostrated by Ibragimov (2005), in presence of log-concave distributions, this condition is satisfied . Following the example of Daníelsson et al. (2013), it is possible to see why the VaR is not sub-additive. Let X_1, X_2 be two assets that are usually normally distributed, but that can be subject to the shocks:

$$X_i = \epsilon_i + \eta_i,\ \epsilon_i \sim IID\mathcal{N}(0, 1)$$

$$\eta_i = \begin{cases} 0 & \text{with prob. } 0.991 \\ -10 & \text{with prob. } 0.009 \end{cases}$$

X_1 and X_2 have the same distribution and are independent. If the shock η occurs the 1% VaR for X_1 and X_2 is 3.1, otherwise it would be 2.3. Since the probability of getting $\eta = -10$ for a X_1 or X_2 is higher than 1%, a portfolio formed by $(X_1 + X_2)$ has an higher VaR than that composed of $2X_1$. In this case then, the sub-additivity is not respected. The crucial importance of sub-additivity in portfolio optimization problems is due to its relationship with convexity. As shown by Acerbi and Tasche (2002), convexity follows from sub-additivity and positive homogeneity, and is necessary to ensure that the risk minimization process guarantees a unique and well-diversified solution. In addition to the lack of sub-additivity, VaR has other pitfalls: its widespread use can result in the overvaluation of its limited implications. First of all, it is necessary to remind that VaR is a measure of risk estimation, regardless of the method used to calculate it. It is just one of the useful tools to have an idea of the riskiness of a portfolio. Furthermore, Abad et al. (2014) have shown that the methods used for VaR estimation differ considerably from one another, causing a substantial difference in precision. As highlighted by Jorion (1996), to be interpreted correctly, the VaR must also be presented with its own confidence interval: the width of the interval provides information about the accuracy of the estimate, defining its reliability.

Although VaR is a rather simple concept, the methods used to calculate it can be very complex. The traditional models for calculating VaR can be divided into two approaches: the Delta-Normal method and the Historical Simulation method. The latter, together with all the other models that will be mentioned, will be analyzed from a theoretical point of view in the following sections.

The Delta-Normal method, also called Variance-Covariance Method, is a simple and not computationally expensive method that can be applied to portfolios with jointly normally distributed positions which can be represented by their delta exposures. Since linear combinations of normal positions that are jointly normally distributed are themself normally distributed, it is possible to compute the VaR in a simple way. Only market values of the positions, their respective weights in the portfolio, and their volatility estimates are needed to compute the Delta-Normal estimates. However, this method has several drawbacks. A first problem is the existence of fat tails in the distribution of returns on most financial assets as pointed out by Duffie and Pan (1997). These fat tails are of particular concern precisely because VaR seeks to understand the behavior of the tails. Secondly, the financial returns often show long-term memory and volatility clustering. According to Cont (2007) volatility clustering can be defined as the property often observed in time series that tends to cluster large price changes together, resulting in persistence of the amplitudes of price changes. Long term memory reflects the long run dependency between stock market return.Without taking those elements into account, VaR estimates can become rather inaccurate. Furthermore, this method cannot be applied to non-linear instruments such as options.

The Historical Simulation method does not make any assumption about the distribution of risk factors. The method consists of estimating the distribution of returns using the empirical distribution and then directly calculating the quantile. This approach solves the problem of fat tails while remaining simple to implement. Although this method is one of the most popular among banks, it has been shown by Abad and Benito (2013) that the estimates produced in this way are not very accurate. The choice of the length of the time window is subjective and one can run the risk of including irrelevant events for future periods, or conversely omitting relevant events just outside the selected range. Furthermore, this method assumes that the distribution of returns remains constant over time, although the findings in this regard are not univocal. To overcome the criticalities of the above mentioned approaches, more complex models were developed to improve the reliability of the above-mentioned VaR approaches. One of the most important contributions in this area has been the class of GARCH model developed by Bollerslev (1986) as a generalization of the ARCH model introduced four years earlier by Engle (1982). Given its ability to handle volatility clustering and its great predictive power, recently confirmed also by Hansen and Lunde (2005), the GARCH model has become very popular. However, it has some drawbacks too: first, GARCH cannot deal with the asymmetry of shocks. In GARCH models, volatility is a function of the magnitude of the lagged residuals but not of their sign. However, some studies have shown that adverse shocks can cause more significant increases in volatility than positive shocks. This is attributed to the leverage effect, caused by the worsening of the debt-to-equity ratio resulting from lower yields. The leverage effect is the negative correlation between past returns and future volatility. According to the work of Jean-Philippe et al. (2008), this effect can be interpreted as a "specific market panic phenomenon". Various variants of this model have been developed to overcome these drawbacks while remaining within the GARCH structure. The EGARCH by Nelson (1991) and GJR-GARCH model by Glosten et al. (1993) are particularly relevant as they take into account the asymmetry of return shocks are .

A different method of calculating VaR than that proposed by GARCH models is the CaViaR Engle and Manganelli (2004).This model, is based on a direct estimate of the quantile, without specifying any assumptions about the distribution of the time series. CaviaR has several specifications, but Şener et al. (2012) have shown that those specifications that take into account the leverage effect provide better estimates than symmetric models. Gerlach et al. (2011) showed that at a confidence level of 0.01, CaViaR performs better than several models in the GARCH family.

Chapter 2

Methodology

This thesis will consider several models to estimate the VaR. Firstly, the traditional methods, Historical Simulation, and Delta Normal method will be presented, which will serve as a benchmark for more sophisticated models. Subsequently, VaR estimation will be performed using the models of the GARCH class, and finally with the different specifications of the CaViaR model. In this chapter, models will be presented discussing their main assumptions and implications. VaR estimates will be constructed at 95% and 99% confidence levels. Each sample will be divided into an in-sample period in which the initial parameter estimates are produced, and an out-of-sample period which provides forecasts.

2.1 Model specification

2.1.1 Delta Normal Method

As previously mentioned, the Delta Normal method relies on the hypothesis that the market risk factors (and therefore also the Profit & Loss distribution of the portfolios) follow a multivariate normal distribution. Using the properties of the normal distribution, it is then possible to determine the quantile corresponding to the desired level α. For example, consider a VaR at 99%: according to the tables of the standard normal, the corresponding quantile is:

$$VaR_{99\%} = -\mu - 2.33\sigma$$

Where 2.33 represents the 99th percentile of the standard normal, μ the mean of the distribution.
It is then clear that the estimation of the standard deviation is a key issue for this method. The standard deviation of a portfolio depends on the standard deviations

of the securities that compose it and their respective correlations

$$\sigma_P = \sqrt{\sum_i \sum_j w_i w_j \rho_{ij} \sigma_i \sigma_j}.$$

Estimating correlations can be very time-consuming and expensive if the portfolio consists of many securities. With a portfolio of n securities, the number of parameters to estimate would be $\frac{n(n+1)}{2}$. It is possible to decompose the portfolio into the k major risk factors to which it is exposed and then recreate as simpler simulated portfolio as a linear combination of them. As shown by Linsmeier and Pearson (2000) this process will speed up the computations, avoiding estimating many parameters. Formally then, a normal distribution of portfolio returns ΔX, with mean μ and variance σ^2:

$$\Delta X \sim \mathcal{N}(\mu, \sigma^2),$$

and its probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right).$$

Then following the previously given definition of VaR with a probability level α, it is possible to write

$$1-\alpha = \int_{-\infty}^{VaR_\alpha} f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx.$$

Considering the cumulative distribution function F

$$\alpha = P(X \leq VaR^\alpha),$$

It is possible to obtain $Z \sim (0,1)$ standardazing X. It is now possible to write:

$$\alpha = P\left[Z \leq \left(\frac{VaR^\alpha - \mu}{\sigma}\right)\right]$$

Finally, with some simple algebraic manipulation, the following result is obtained:

$$VaR^\alpha = \mu + z^\alpha \sigma_P,$$

where z^α is the α-th percentile of a normal standard distribution.
As previously specified, this model is based on the normality assumption which is difficult to apply to time series of returns. Therefore, the derived estimates must be considered little reliable in much of the empirical cases.

2.1.2 Historical Simulation

According to Pérignon and Smith (2010), the Historical Simulation method is one of the most widely used for calculating VaR . Its use is mainly related to its two major advantages. First, it is a simple method to implement because it does not require optimization or parameter estimation. Second, this method is model-free, since Historical Simulation is not based on any parametric model, but only considers historical realizations. Consider the series of asset returns X in a portfolio with returns:

$$R = \sum_{i=1}^{n} w_i X_i,$$

and the information contained therein relating to the series of m events that have previously occurred. Consider a hypothetical portfolio built using the historical returns of the underlying securities, weighted by the current portfolio weights:

$$\{R_{t+1-\tau}\}_{\tau=1}^{m} \equiv \left\{ \sum_{i=1}^{n} h_i R_{t+1-\tau} \right\}_{\tau=1}^{m},$$

where τ is the time horizon of the returns and h_i are the historical weights of the portfolio. The return constructed in this way simulates the returns of a hypothetical portfolio using current positions. This method assumes that the distribution of the variable at time t+1 is well approximated by its historical distribution. The Value at Risk in this case is simply the percentile associated with the α level of the historical distribution.

The Filtered Historical Simulation method was developed by Barone-Adesi and Giannopoulos (2001) to improve this approach while retaining its framework, and it has been shown that it is able to provide more reliable estimates.

2.1.3 GARCH models

The ARCH model class was developed by Engle (1982) and later extended by Bollerslev (1986) with the introduction of the GARCH class. Let

- $\Theta_0 \in \Theta$ the element of interest belonging to the set of parameters Θ;

- $r_t(\Theta_0)$ be the stochastic process, with conditional mean

$$\mu_t(\Theta_0) := E_{t-1}(r_t); \tag{2.1}$$

- the past return innovations

$$\varepsilon_{t-1} = r_{t-1} - \mu_{t-1}; \tag{2.2}$$

Then ε_t follows an ARCH/GARCH model if

$$\sigma^2(\Theta_0) = Var_{t-1}[\varepsilon_t(\Theta_0)] = E_{t-1}[\varepsilon_t^2(\Theta_0)], \tag{2.3}$$

with $E_{t-1}[\varepsilon_t(\Theta_0)] = 0$.
Considering now the process

$$\varepsilon_t(\Theta_0) = z_t(\Theta_0)\sigma_t(\Theta_0) \tag{2.4}$$

$$E_{t-1}[z_t(\Theta_0)] = 0 \tag{2.5}$$

$$Var_{t-1}[z_t(\Theta_0)] = 1. \tag{2.6}$$

If $z_t(\Theta_0) \sim \mathcal{N}(0,1)$, it is possible to write

$$E_{t-1}[\varepsilon_t^2] = E_{t-1}[z_t^2(\Theta_0)]E_{t-1}[\sigma_t^2(\Theta_0)] = \sigma_t^2. \tag{2.7}$$

Having defined its elements, the model GARCH(1,1) is given by

$$\sigma_t^2 = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2,$$

where ω is the term referring to the long term variance; $\alpha_1\varepsilon_{t-1}^2$ is the the paramiter driving the impact of past innovations and $\beta_1\sigma_{t-1}^2$ is the autoregressive component. The GARCH(1,1) sufficient condition for strict stationarity is $\alpha_1 + \beta_1 < 1$. The unconditional variance $Var(r_t) = E(\varepsilon_t^2) = E(\sigma_t^2)$ under stationarity can be rewritten as follows:

$$\begin{aligned} E(\sigma_t^2) = \omega + \alpha_1 E(\varepsilon_{t-1}^2) + \beta_1 E(\sigma_{t-1}^2) \\ E(\sigma_{t-1}^2) = E(\varepsilon_{t-1}^2) = E(\sigma_t^2) \\ Var(r_t) = \frac{\omega}{1-\alpha_1-\beta_1}. \end{aligned} \tag{2.8}$$

The GARCH(1,1) model is among the most widely used, as it has only three parameters to estimate and has shown good predictive capabilities. It is possible to generalized the model to a GARCH(p,q), which parameterizes the variance based on q lags of squared errors and p lags of conditional variance:

$$\sigma_t^2 = \omega + \alpha_{\mathbf{1}}(\mathbf{L})\varepsilon_t + \beta(\mathbf{L})\sigma_t^2, \tag{2.9}$$

where $(\alpha_{\mathbf{1}}\mathbf{L}) = \alpha_1\mathbf{L} + \alpha_2\mathbf{L}^2 + ... + \alpha_p\mathbf{L}^p$ and $\beta(\mathbf{L}) = \beta\mathbf{L} + \beta\mathbf{L}^2 + ... + \beta\mathbf{L}^q$.
Since the normality assumption of z_t did not seem very realistic, to consider heavier tails, some modification to the initial model was proposed. Bollerslev (1987) proposed

a modification where z_t was distributed as a t-student or as a Generalized Error Distribution (GED).
In addition, although the GARCH is able to account for volatility clustering and leptokurtosis phenomena, it did not account for the leverage effect. The development of several variants of GARCH dealt with these problems.
Nelson (1991) proposed a version of GARCH that solves the problem of non-negativity of the parameters and considers the leverage effect: the EGARCH. This model can be specified with the following equation:

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^{q} \beta_i \log(\sigma_{t-i}^2) + \sum_{j=1}^{p} \alpha_i[\phi z_{t-1} + \psi(|z_{t-1}| - E|z_{t-1}|)]. \tag{2.10}$$

With this new specification, volatility depends on the magnitude and direction of shocks. The term $(|z_{t-1}| - E|z_{t-1}|)$ accounts the magnitude effect and is weighted by ψ. The direction of the z_t shocks affects volatility through the parameters ϕ and ψ. For example, if $\psi = 0$ and $\phi < 0$, a negative innovation shock z_{t-1} result in an increase in volatility $\log(\sigma_t^2)$. Nelson assumed that z_t has a Generalized Error Distribution (GED). The latter is a vast family of distributions, including the normal and uniform distribution, through distributions with fatter tails.
Glosten et al. (1993) proposed a new version of the GARCH, which was able to take into account the asymmetry of the shocks in a more straightforward way. They specified the following GJR-GARCH model in the following way:

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \beta_j \sigma_t^2 + \sum_{i=1}^{p} (\alpha_i \varepsilon_{t-1}^2 + \gamma_i 1_{(-\infty;0)}(\varepsilon_t)\varepsilon_{t-1}^2). \tag{2.11}$$

If γ is positive, the increase in volatility will be greater than in the opposite case, since the indicator function will take on a null value. Therefore, if $\gamma > 0$, a leverage effect is observed. To control for outliers in the series and provide a better fit to the returns, Harvey and Chakravarty (2008) proposed the Beta-t-EGARCH, which was then brought back to the class of Generalized Autoregressive Score (GAS) models introduced by Creal et al. (2013). The GAS models use the score of the density function of the data to update the density function of the process under analysis. Among the various advantages that the use of the score involves, particularly relevant is the one related to the use of the information coming from the whole distribution f_t, while traditionally, only the first and second-order moments of the series under analysis are used. The Beta-t-EGARCH(1,1) model can be specified as follows: defining the time series

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \varepsilon_t &= \exp{}_t u_t, \end{aligned} \tag{2.12}$$

where $u_t \sim t(\nu) i.i.d.$. Let then define the following:

$$\lambda_t = \alpha_0 + \alpha_1 e_{t-1} + \beta_{1t-1} \tag{2.13}$$

$$e_t = (\nu + 1) b_t - 1 \tag{2.14}$$

$$b_t = \frac{(y_t - \mu_t)^2 / (\nu \exp 2\lambda_t)}{1 + (y_t - \mu_t)^2 / [\nu \exp 2\lambda_t]}. \tag{2.15}$$

Estimating GARCH models

The Maximum Likelihood method is a common way to estimate the parameters of the GARCH models. The process finds the parameter values most likely to return the actual data. A Likelihood function is first constructed and then maximized with respect to the model parameters to proceed with the estimation.
Consider a random variable y, whose realization is conditional on a set of unknown parameters $\mathbf{\Theta}$ and with a probability density function $f(y|\mathbf{\Theta})$. The product of n independent and identically distributed individual densities is called the likelihood function

$$f(y_1, ..., y_n|\Theta) = \prod_{i=1}^{n} f(y_i|\Theta) = \mathbf{L}(\mathbf{\Theta}|\mathbf{y}). \tag{2.16}$$

Since this specification may lead to numerical problems, usually the logarithmic transformation is applied:

$$log\, \mathbf{L}(\mathbf{\Theta}|\mathbf{y}) = \sum_{i=1}^{n} log f(y_i|\Theta) \tag{2.17}$$

This is the most common form, so usually the log-likelihood function is identified directly with $\mathbf{L}(\mathbf{\Theta}|\mathbf{y})$. Considering the example of a GARCH(1,1) with normal innovations, the log-likelihood function for the t-th observation would be:

$$log\, \mathbf{L}(\mathbf{\Theta}|\mathbf{y}) = -\frac{1}{2} log(2\pi) - \frac{1}{2} log(\sigma_t^2) - \frac{\varepsilon^2}{2\sigma_t^2}, \tag{2.18}$$

with $\varepsilon_t = y_t - \mu_t(\Theta)$.

Once the parameters of a GARCH model have been estimated, it is possible to conduct the diagnostic test based on the Ljung-Box statistic to check for the presence of serial correlation and ARCH effects in the standardised residuals. Let

$$\zeta_t \sim N(0, 1), \tag{2.19}$$

be a White Noise series. Its sample autocorrelation coefficients will be approximately distributed according to a normal

$$\hat{\rho}_s \sim \text{approx. } N(0, 1/T), \tag{2.20}$$

where T is the sample size and $\hat{\rho}_s$ is the autocorrelation obtained in the sample at lag s. Ljung and Box (1978) proposed a statistical test to check that all the coefficients $\hat{\rho}_n$ are simultaneously zero. The test considers squared coefficients to avoid eliding terms with opposite signs. The test statistic Q^* is calculated as follows:

$$Q^* = T(T+2)\sum_{k=1}^{n} \frac{\hat{\rho}_s}{T-k} \sim \chi^2_n. \tag{2.21}$$

This test is a modification of the Box and Pierce (1970) test, which is less accurate for small samples. Therefore, the null hypothesis of the Ljung-Box test is that all coefficients are simultaneously zero, against the alternative hypothesis that at least one of them is non-zero. In the case of GARCH models, the test verifies that the standardized residuals are not autocorrelated. Therefore, the absence of autocorrelation makes it possible to establish that the model has captured any serial correlation present in the time series.

2.1.4 Caviar Models

In 2004, Engle and Manganelli (2004) introduced a new class of models for estimating VaR to overcome some of the critical issues associated with previous approaches. Parametric approaches, such as the Delta Normal method or GARCH models, estimate VaR by estimating portfolio volatility. These methods need to assume a specific distribution of innovations and therefore the distribuition of their returns. These models are exposed to the risk of incorrect specification of the distribution of the Data Generating Process. The application of Extreme Value Theory in the financial field has shown that it can sometimes improve the accuracy of forecasts. As summarised by Rocco (2010), the accuracy and flexibility of estimates made with this approach are due to its consideration of the tails of the data, without considering the centre of the distribution. This fundamental feature is particularly useful when estimating risk measures that focus on the tails. However, even this approach proved somewhat inaccurate when not particularly "extreme" quantiles, such as 5%, are taken into analysis.
Rolling quantile methods typically assume that each return has the same probability of being observed within the specified time window. This method implicitly assumes that the distribution of returns is stationary over the time window, which is usually one year. The approach to quantile estimation that Engle and Manganelli propose involves directly modeling the quantile. The model includes an autoregressive component able to consider the phenomena of volatility clustering. The Conditional

Autoregressive Value at Risk can be introduced as follows.
Defining:

- α the level of probability associated to the VaR,
- $\{y_t\}_{t=1}^m$ the vector of portfolio returns,
- $\boldsymbol{\beta_\alpha}$ the p-dimentional vector of the unkonwn parameters,
- $\boldsymbol{x}_t$ a vector of observable variables,
- $f_t(\boldsymbol{\beta}) \equiv f_t(\boldsymbol{x_{t-1}}, \boldsymbol{\beta_\alpha})$ the time t α-quantile of the portfolio return distribution conditioned on information available at time $t-1$ (for simplicity $\boldsymbol{\beta_\alpha}$ will be indicated just as $\boldsymbol{\beta}$),

then the CAViaR can be described as follows:

$$f_t(\boldsymbol{\beta}) = \beta_0 + \sum_{i=1}^{q} \beta_i f_{t-i}(\boldsymbol{\beta}) + \sum_{j=1}^{r} \beta_i l(\boldsymbol{x}_{t-j}), \tag{2.22}$$

where the dimension of β is $p = r + q + 1$. l is a function of the lagged values of observables and links $f_t(\boldsymbol{\beta})$ with the information set. $\beta_i f_{t-i}(\boldsymbol{\beta})$ is the autoregressive component. Usually lagged returns are used as $\boldsymbol{x}_{t-j}$.
It is than possible to specify different specification of CAViaR:

- Adaptive:

  $$f_t(\beta_1) = f_t(\beta_{t-1}) + \beta_1 \left([1 + \exp G[y_{t-1} - f_{t-1}(\beta_1)]])]^{-1} - \alpha \right), \tag{2.23}$$

  where G is a positive finite number. Using this specification, if the VaR forecast is less than the realised loss, then the next forecast will increase and vice versa. This decreases the probability of no hits throughout the series.

- Symmetric:

  $$f_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 f_{t-1}(\boldsymbol{\beta}) + \beta_3 |y_{t-1}|$$

- Asymmetric:

  $$f_t(\boldsymbol{\beta}) = \beta_1 + \beta_2 f_{t-1}(\boldsymbol{\beta}) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^-$$

  where $(y)^+ = max(y, 0)$ and $(y)^- = -min(y, 0)$

Since the coefficients of the lagged VaR are not constraint to be one, the Symmetric and Asymmetric specifications are mean revearting. In this way, when VaR forecasts differ too much from their average, they will be "reset" to their long term average.

Estimating Caviar models

Let $\{y_t\}_{t=1}^m$ be the vector of observed portfolio returns generated by the model

$$y_t = \boldsymbol{x_t'}\boldsymbol{\beta^0} + \varepsilon_{\alpha t}, \quad Quant_\alpha(\varepsilon_{\alpha t}|\boldsymbol{x_t}) = 0 \tag{2.24}$$

where $Quant_\alpha(\varepsilon_{\alpha t}|\boldsymbol{x_t})$ is the α-quantile of $\varepsilon_{\alpha t}$ conditional on $\boldsymbol{x_t}$. Considering $f_t(\boldsymbol{\beta}) \equiv \boldsymbol{x_t}\boldsymbol{\beta}$ is now possible to define the estimate associated with the level α as the value $\hat{\boldsymbol{\beta}}$ that solves:

$$\min_\beta \frac{1}{m}\sum_{t=1}^{m}[\alpha - I(y_t < f_t(\boldsymbol{\beta})][y_t - f_t(\boldsymbol{\beta})]. \tag{2.25}$$

This minimisation can be carried out using either OLS (ordinary least squares) or LAD (least absolute deviation), although the latter provides more robust results.

2.1.5 Backtesting VaR

Assessing the predictive capabilities of VaR models is a key element in risk management. The Basel Accords included the need to test models to understand their reliability. Backtesting refers to evaluating a financial risk model using future profit and loss realizations forecasts. As pointed out by Christoffersen (2008), risk management teams typically perform this type of evaluation, or bank regulators. In VaR models, backtesting is performed to ensure that the model can provide sufficiently accurate estimates. Although several approaches exist, the most popular ones traditionally are those developed by Kupiec (1995) and Christoffersen (1998).
Consider a model $VaR_t(\alpha)$, where α is the probability level the VaR prediction return is reasonably assumed to be exceeded. Given y_t the series on which VaR is calculated, define the hit sequence of VaR_t, the violations:

$$I_t = \begin{cases} 1, & \text{if } y_t < VaR_t(\alpha), \\ 0, & \text{otherwise} \end{cases} \tag{2.26}$$

This definition does not consider the magnitude of the violation, so it will be used in terms of frequency. Kupiec (1995) tests for the following hypothesis:

$$\begin{cases} H_0 \; I_t \sim IID \text{ Bernoulli}(\alpha) \\ H_1 \; I_t \sim IID \text{ Bernoulli}(\pi) \end{cases} \tag{2.27}$$

In other words, it is possible to rewrite the null hypothesis as $H_0 : E[I_t] \equiv \pi = \alpha$. To estimate the expected value of the hit sequence, one can consider the sample average: $\hat{\pi} = \frac{1}{T}\sum_{t=1}^{T} I_t = \frac{T_1}{T}$, where T_1 is the number of 1s in the sample. If the VaR model passes the Kupiec's test, it produces hits on average with probability α. If the null hypothesis was right, the expected value of $\hat{\pi}$ would be:

$$E[\hat{\pi}] = \frac{1}{T}\sum_{t=1}^{m} I_t = \alpha. \tag{2.28}$$

A simple test can then be implemented to check if the difference between π and α is statistically significant:

$$MT = \sqrt{T}\frac{\pi - \alpha}{\sqrt{Var(I_t)}} \sim \mathcal{N}(0,1), \tag{2.29}$$

Where the variance $Var(I_t)$ can be estimated as the sample variance of the hit sequence.

It is also possible to implement the test using a Likelihood ratio test. To do so, it is first necessary to write the likelihood function of an IID Bernoulli(π) sequence:

$$L(\pi) = \prod_{t=1}^{T}(1-\pi)^{1-I_{t+1}}\pi^{I_{t+1}} = (1-\pi)^{T_0}\pi^{T_1}, \tag{2.30}$$

where T_0 and T_1 are respectively the number of zeros and ones in the the sample. $\hat{\pi}$ can be estimated as stated before, as the observed violations on the correct valuation of the model, $\hat{\pi} = \frac{T_1}{T}$. Replacing π in equation 2.30, with $\hat{\pi}$, the Likelihood function is maximised:

$$L(\hat{\pi}) = \left(1 - \frac{T_1}{T}\right)^{T_0}\left(\frac{T_1}{T}\right)^{T_1} \tag{2.31}$$

If the null hypothesis is correct, $\pi = \alpha$, where α, the likelihood would be:

$$L(\alpha) = \prod_{t=1}^{T}(1-\alpha)^{1-I_{t+1}}\alpha^{I_{t+1}} = (1-\alpha)^{T_0}\alpha^{T_1} \tag{2.32}$$

The Likelihood ratio test statistics can be written as follows:

$$LR_{uc} = -2ln\left[\frac{L(\alpha)}{L(\hat{\pi})}\right] \sim \chi_1^2 \tag{2.33}$$

If the model is correctly specified, violations should occur randomly. However, due to the phenomenon of volatility clustering, changes in the volatility of large magnitude tend to occur one after the other. For example, as stated by Pritsker (2006) the presence of volatility clustering, therefore, violations will not happen randomly but will be correlated with each other.

It is possible to test the independence of the violations through a test based on the Likelihood function. To proceed, it is necessary to assume that the sequence of hits can be described by a first-order Markov process, with the following transition probability matrix:

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix},$$

where the probability that in period $t+1$ a hit occurs given that at the current period t this has not occurred (so the VaR estimate is lower than the actual realization) is equal to π_{01}. Similarly, the probability that there are two consecutive violations is given as $\pi_{11} = Pr(I_t = 1 \bigcap I_{t+1} = 1)$. The other elements of the matrix represent the probabilities of the complements.
Once this process is defined and the T observations of the sample are considered, the Likelihood function for this Markov process can be written as follows:

$$L(\Pi_1) = (1-\pi_{01})^{T_{00}}(1-\pi_{11})^{T_{10}}\pi_{01}^{T_{01}}\pi_{11}^{T_{11}},$$

where $T_{i,j}$, with $i, j = 0, 1$ is the number of observations j followed by observations i. Maximising the function with respect to π_{01} and π_{11}, then $\hat{\pi}_{01}$ and $\hat{\pi}_{11}$ are obtained:

$$\begin{aligned} \hat{\pi}_{01} &= \frac{T_{01}}{T_{01}+T_{11}} \\ \hat{\pi}_{11} &= \frac{T_{11}}{T_{10}+T_{11}}. \end{aligned} \tag{2.34}$$

The probabilities for the other two events can be obtained simply as follows:

$$\begin{aligned} \hat{\pi}_{00} &= 1-\hat{\pi_{01}} \\ \hat{\pi}_{10} &= 1-\hat{\pi_{11}}. \end{aligned} \tag{2.35}$$

If the model is correctly specified, all probabilities $\pi_{ij}\ with\ i, j = 1, 1$ must be equal. One can consider the hypothesis $\pi_{01} = \pi_{11}$ using the likelihood ratio test:

$$LR_{ind} = -2ln\left[\frac{L(\hat{\pi_{01}})}{L(\hat{\Pi_1})}\right] \sim \chi_1^2.$$

Combining the tests of independence and the tests of unconditional coverage yields the test of conditional coverage: $LR_{cc} = LR_{ind} + LR_{uc}$, which asymptotically $LR_{cc} \sim \chi^2(2)$.

To evaluate the validity of the VaR calculation models, the Basel Committee in 1996 proposed an intuitive system for backtesting VaR called Traffic Light.
Consider the VaR forecast series $VaR(\alpha)_i$, with $i = 1, \ldots, N$ such that:

$$VaR(\alpha) := inf\{y \in \mathbb{R} : F_L(z) \leq \alpha\},$$

Where $F_L(y)$ is the cumulative distribution function of the variable L evaluated at y. Recalling the definition of VaR violation I_t 2.26, we define the total number of violations as:

$$X^N_{VaR}(\alpha) := \sum_{i=1}^{N}(I_t).$$

To evaluate the goodness of the forecasts, the Traffic Light test considers the difference with the theoretical number of breaches $E[X^N_{VaR}(\alpha)] = N\alpha$ and the actual number of violations. A correctly specified model should have a number of breaches very close to the theoretical one, with deviations due to the randomness of sampling . Having defined α and N, the cumulative probability of obtaining a number of breaches equal to or less than x, is defined as:

$$\Psi^{\alpha,N}_{VaR}(x) := \mathcal{P}[X^N_{VaR}(\alpha) \leq x].$$

This probability is compared with that of a binomial with parameters N and $1-\alpha$. In particular, the model under test is assigned a colour according to the following criterion:

- Green: $\Psi^{\alpha,N}_{VaR}(x) \leq 95\%$;
- Yellow: $95\% \leq \Psi^{\alpha,N}_{VaR}(x) \leq 99\%$;
- Red: $\Psi^{\alpha,N}_{VaR}(x) \geq 99\%$;

Thus, a model falls into the green zone if the cumulative probability of obtaining the number of breaches x is equal to or less than 95%. It falls into the yellow area if the probability of getting those breaches is between 95% and 99%, and finally, it belongs to the red zone if the cumulative probability exceeds 99%. These three colors identify a hierarchical criterion to assess the adequacy of a bank's risk model. Intuitively, green does not cast doubt on the accuracy of the model. Yellow expresses an inconclusive doubt. Finally, red indicates that the model is inadequate and should be modified.

In their paper Engle and Manganelli (2004) presented the Dynamic Quantile test to verify the independence of VaR violations and the violation rate of the exceedances. Within this work, the test was implemented on R, via the GAS package Ardia et al. (2019). Consider the violation indicator function I_t, where $t = 1, \dots, N$, defined as in equation 2.26. In the paper the function Hit^α_t is constructed by subtracting its mean from the indicator:

$$Hit^\alpha_t = I_t - \alpha.$$

If the tested model is correctly specified, the process Hit^α_t with $t = 1, \dots, N$ will have mean 0 and serially uncorrelated. These two hypotheses together constitute the null hypothesis of the test, which is implemented by the following regression:

$$Hit^\alpha_t = \delta_0 + \sum_{l=1}^{L} \delta_l Hit^\alpha_{t-l} + \delta_{L+1} VaR_{lt-1}(\alpha) + \epsilon_t, \tag{2.36}$$

where L represents the number of Lags. Wald's test is performed on this regression, with null hypotheses predicting joint nullity of all coefficients δ.
In the paper of González-Rivera et al. (2004), a VaR-based loss function based on the work of Koenker and Bassett (1978) is proposed. The authors propose function defined as follows:

$$Q = P^{-1} \sum_{t=1}^{N} (\alpha - I_{t+1})(y_{t+1} - VaR_{t+1}^{\alpha}), \tag{2.37}$$

Where I_t is defined in eq. 2.26. The function is asymmetric because the observations in which the returns exceed the forecasted VaR have more weight. Between two models the one with the smaller Q-statistic is preferred. The loss function allows the performance of the models to be compared, although to get a clearer idea of the correct specification of the models, it is better to look at the whole set of tests proposed.

Chapter 3

Data

3.1 Data collection

Data from four time series will be used to compute VaR estimates. Each data series is taken from a different asset class, to understand if there are differences in their behaviours. The assets chosen for the analysis are the following:

- Stocks: SP500
- Commodity: WTI
- Forex: EURUSD
- Interest Rate: Euribor 3M

These assets were chosen because they are highly liquid, widley traded and identify key market risk factors. Each time series goes from 2nd January 2009 to 1st December 2021, with 3000 daily observations. Data were extracted from Bloomberg.
To deal with returns instead of prices, the following log-transformation was applied:

$$r_t = ln\left(\frac{P_t}{P_{t-1}}\right). \tag{3.1}$$

In this formula, r_t is the daily return, P_t is the price at day t and P_{t-1} is the price at the preceding day. Due to this transformation, each time series loses one observation. The negative value in the WTI series on 20th April 2020 was replaced with the average of the two preceding and following days. This manipulation is necessary to avoid a negative argument in the logarithmic transformation. In this respect, a negative price indicates that producers are willing to pay buyers to take their production. Negative prices are due both to a technical component of the market, linked to the dynamics of the futures, and to the fact that producers' storage capacities had reached their limit. Sellers needed to rent oil tankers in order to manage the supply surplus. For model estimates and forecasts, Euribor3m interest rates were converted into prices using the formula

$$P_{Euribor,t} = 100 - r_t,$$

Where r_t is the interest rate extracted from Bloomberg.

3.2 Descriptive statistics

As previously specified, financial returns often do not come from normal distributions. The statistical tests of Kolmogorov-Smirnov Massey Jr (1951), Anderson-Darling Anderson and Darling (1954), and Cramer-von Mises Cramr (1928) were performed to test for normality.
Kolmogorov-Smirnov Test tests the theoretical distribution and the data distribution have the same cumulative distribution function. The statistic used for the test is the maximum difference in absolute value between the empirical cumulative distribution function $\hat{F}(x)$ and the theoretical one $G(x)$:

$$D^* = \max_x(|\hat{F}(x) - G(x)|).$$

The two-sided test tests the null hypothesis that the empirical and Hypothesised distribution are equal. If this difference is sufficiently large (i.e. exceeds the critical threshold determined by the distribution of the test- statistic), the null hypothesis is rejected.
Another way that can be used to assess whether or not a sample comes from a normal distribution is testing with quadratic statistics. Let $F(x)$ be the empirical distribution and $G(x)$ the theoretical distribution to be tested, and let $x_1 \leq x_2 \leq x_3 \cdots \leq x_n$ be the ordered observations and $u_i = G(x_i)$. Quadratic statistics measure the distance between F(x) and G(x) using the following formula:

$$Q^2 = n \int_{-\infty}^{+\infty} (F(x) - G(x))^2 \Psi(x)\, dF(x). \tag{3.2}$$

The AD test is imposed $\Psi(x) = \frac{1}{F(x)(1-F(x))}$, so Q^2 takes the form:

$$A^2 = n \int_{-\infty}^{+\infty} \frac{(F(x) - G(x))^2}{F(x)(1 - F(x))}\, dF(x), \tag{3.3}$$

And its test statistic can be approximated from 3.3 as follows:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \left[ln(u_i) + ln(1 - u_{n-j+1})\right].$$

	SP500		WTI		Euribor3M		EurUsd	
	Statistic	**P-Value**	**Statistic**	**P-Value**	**Statistic**	**P-Value**	**Statistic**	**P-Value**
Kolmogorov-Smirnov	0.4799	0.0000	0.4636	0.0000	0.3108	0.0000	0.4908	0.0000
Anderson-Darling	76.260	0.0005	193.345	0.0000	133.02	0.0000	123.196	0.0005
Cramer-von Mises	13.905	0.0010	146.111	0.0010	22.102	0.0010	20.703	0.0010

Table 3.1: This table shows the results of the main goodness of fit tests for each time series. The tests statistics are followed by their p-values.

The function is chosen to give more weight to the tails. In fact, given its formulation, $\Psi(x)$ takes on larger values when F(x) is close to 0 or 1, and vice versa for values relative to 0.5. This choice of $\Psi(x)$ is suitable for comparing distributions assumed to disagree in the tails. Still, it loses power when they have differences close to the median of F(x). This test is one of the most powerful ones to check whether the tails of the theoretical and empirical distribution coincide. However, both tails are assigned the same weight, which should be taken into account if one wants to obtain information on only one of the tails. A modification of the AD test was proposed by Sinclair et al. (1990) to assign different weights to the two tails of the distributions. The Cramer-von Mises test instead assigns a value of 1 to Ψ, and its test statistics is computed as follows:

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i-1}{2n} - G(x_i) \right]^2 .$$

This test is a variant of the Anderson-Darling test, which does not assign higher weights to the tails of the distributions than to the less extreme values. As with the Anderson-Darling test, the null hypothesis is that the theoretical and empirical distributions are equal. A high value of the W^2 statistic means that the null hypothesis cannot be accepted.

The results of the three good-of-fitness tests just presented are proposed in the Table 3.1. They show all reject the null hypothesis, confirming that data are not normally distribuited. Figure 2.1 shows the comparison of the empirical distribution of the SP500 data with the normal distribution. The time series of the SP500 displays volatility clustering during the last period of the sample. The presence of this phenomenon should favour the use of those GARCH models designed to deal with it. Descriptive statistics of the four time series are displayed below.

All series show a mean and median very close to zero and with kurtosis and skewness values incompatible with a normal distribution. These data support the evidence that financial returns exhibit fatter tails than a normal and skewness in the distribution.

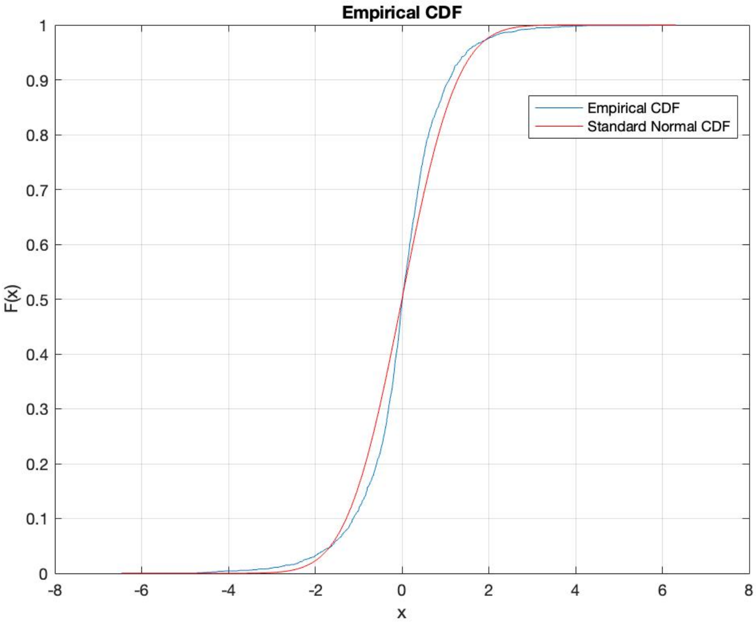


Figure 3.1: The figure shows the difference between the empirical cumulative distribution function and the normal cumulative distribution function for the SP500 series.

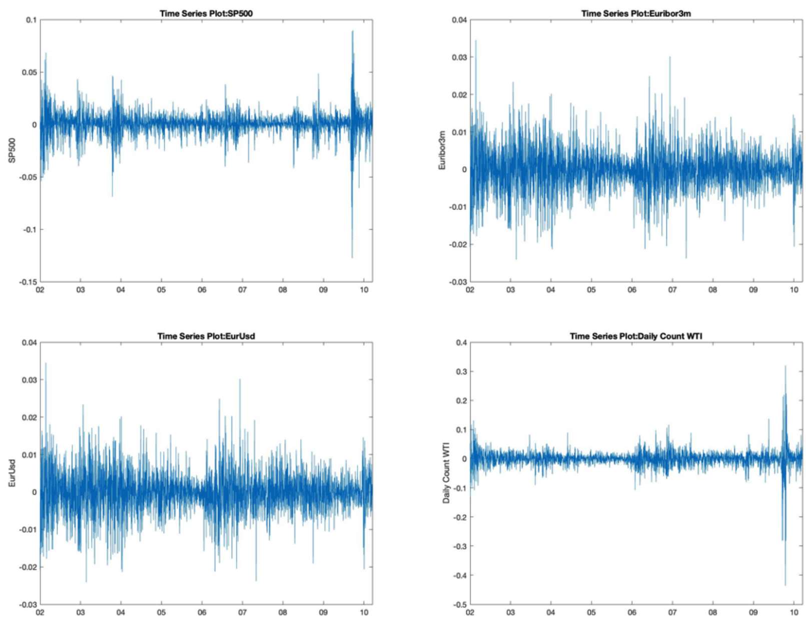


Figure 3.2: Log-returns of the four times series from 2nd January 2009 to 1st December 2021.

	SP500	WTI	Euribor3M	EurUsd
Observations	3000	3000	3000	3000
Minimum	-0.1277	-0.4361	-0.0006	-0.0241
Maximum	0.0897	0.3196	0.0009	0.0345
Mean	0.0001	0.0000	0.0001	0.0000
Median	0.0007	0.0000	0.0000	0.0000
Variance	0.0001	0.0000	0.0000	0.0000
Stdev	0.0117	0.0057	0.0000	0.0057
Skewness	-0.6805	-10.343	2.9753	0.0260
Kurtosis	160.503	415.813	48.073	48.599

Table 3.2: This table shows the descriptive statistics for each time series.

Chapter 4

Results

This chapter will present the main empirical results, comparing the accuracy of VaR forecasts. Two classes of models are proposed in the analysis: GARCH models and CAViaR models. For both classes, several variants are considered. Finally, the results will be compared with the main backtests of the literature to compare their characteristics.

4.1 GARCH models

The models considered are those described in the previous sections: GARCH(1,1), EGARCH(1,1), Beta-t-EGARCH(1,1), and GJR-GARCH(1,1). Only models with lags (1,1) are adopted to keep the confrontation easy to present. This model is the most widely used since models with p, q bigger than one usually provide very similar results to the simplest one. Brooks and Burke (2003) however, showed that the use of more sophisticated information criteria than the traditional BIC and AIC statistic could sometimes lead to improvements in the estimates.

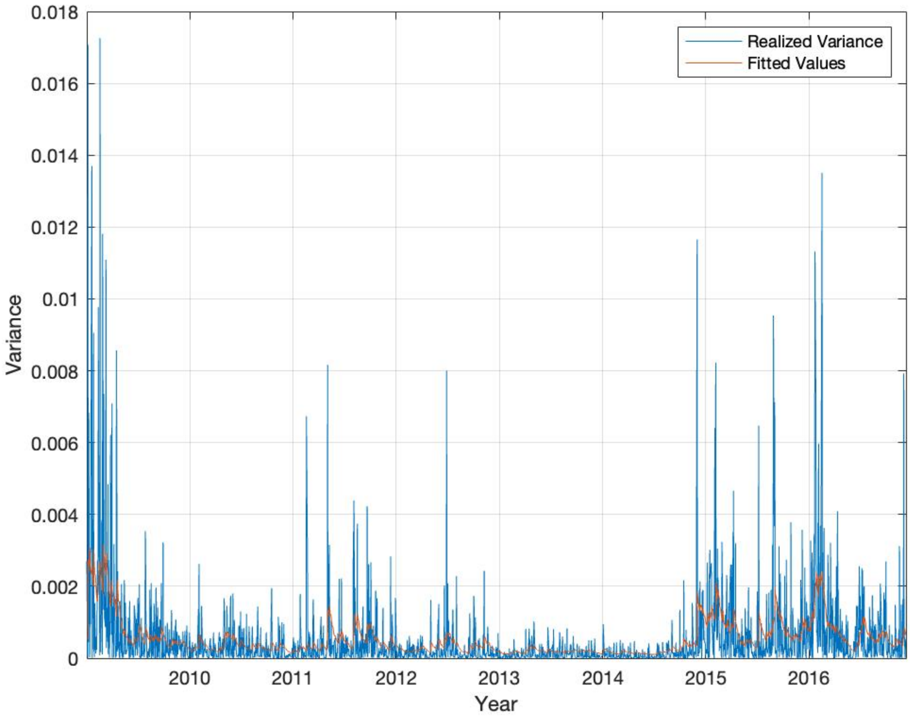
In evaluating the models, three types of z_t innovations were considered: normal, t-student, and Generalized Error Distribution (GED). Based on the above definition of GARCH models, it is evident that future returns and z_t innovations follow the same conditional distribution. Therefore, an incorrect specification of the distribution of innovations may lead to inaccurate estimates. Furthermore, considering that financial returns typically follow distributions with fat tails, it is appropriate to consider distributions that take this characteristic into account. The results of this analysis confirm that models with skewed and fat-tailed innovation distributions outperform those that adopt normal innovations. Only the estimates of models with GED innovations that provided the best results among the others will be presented for brevity.

The samples were divided into two parts: an in-sample part, to estimate the parameters, and an out-of-sample part to compute the forecasts to be tested later. The table below shows the in-sample parameter estimates and their respective standard errors.

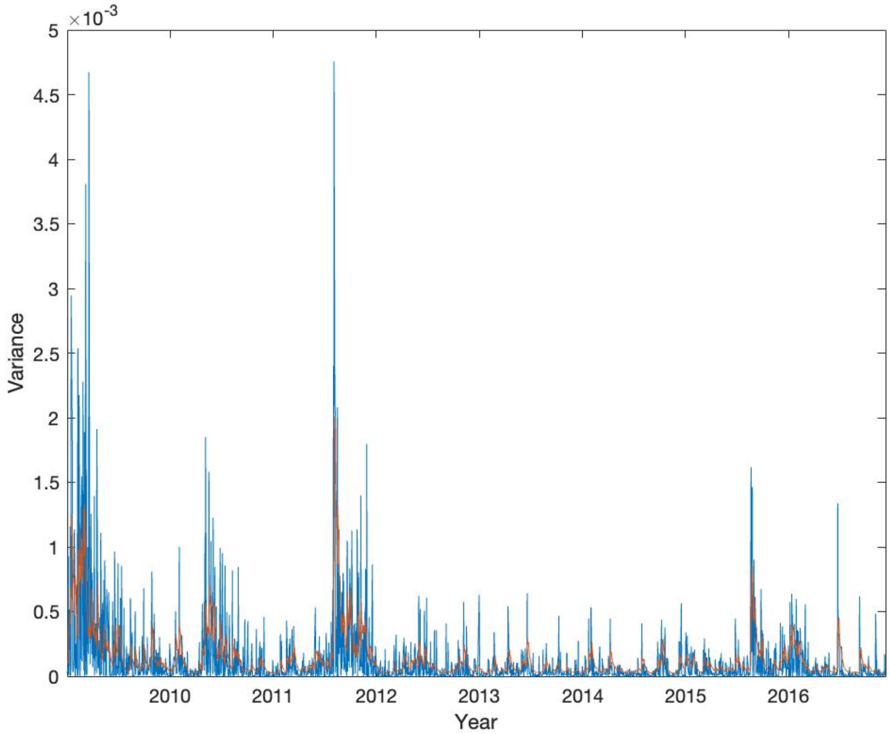
For the sake of brevity, only the model GJR-GARCH, with GED innovations, which led to the best forecasting performance, is shown. Models calculated with t-student innovations show very similar results. However, the number of VaR exceedings increase sensibly when normal innovations are used. Estimates are performed in Matlab, using the log-likelihood maximization method. The in-sample covers the days from 01/02/2009 to 10/20/2016 for 2000 observations. After estimating the parameters of the various GARCH models, the fitted values of the variance were reconstructed.

The table below shows that the all series presents a null intercept, while the leverage coefficients γ characteristics of the GJR model are expecially high in the SP500 and Euribor3m series. This data shows that the leverage phenomenon is present in the in-sample period. It confirms the evidence that the direction of the shock has an impact on the volatility estimates. The serial correlation between the standardized residuals is analyzed by performing the Ljung test. According to the null hypothesis that the standardized returns do not exhibit serial correlation for 20 lags. The inability to reject the null hypothesis for all series, except Euribor, indicates that in the in-sample period, all GARCH effects were captured by the model.

	GJR-GARCH(1.1)			
	Parameter	**Estimate**	**Std. Error**	**Ljung-Box Reject Null**
SP500	Constant: α_0	.000	.000	NO
	ARCH: α_1	.000	.019	
	GARCH: β	.839	.019	
	Leverage: γ	.029	.040	
WTI	Constant: α_0	.000	.000	NO
	ARCH: α_1	.012	.008	
	GARCH: β	.094	.008	
	Leverage: γ	.079	.014	
Euribor3m	Constant: α_0	.000	.000	YES
	ARCH: α_1	.541	.905	
	GARCH: β	.011	.220	
	Leverage: γ	.772	1.01	
Eurusd	Constant: α_0	.000	.000	NO
	ARCH: α_1	.017	.009	
	GARCH: β	.958	.007	
	Leverage: γ	.038	.001	



(a) WTI



(b) SP500

Figure 4.1: WTI and SP500 Fitted GJR-GARCH Variance and Realized Variance. The fitted series are computed from in-sample estimates on 2000 observations.

: This table shows the in-sample parameter estimates of GJR-GARCH model. The estimates are computed considering an in-sample period starting from 2nd January 2009 and containing 2000 observations.

The out-of-sample forecasts were then formulated, considering a rolling window of 100 days, used to calculate one-day-ahead volatility. The VaR was calculated from the product between the estimate of volatility and the quantile corresponding to the α value. The quantile was calculated considering the following formula:

$$VaR_{t+1} = F(\alpha)\sigma_{t+1}, \tag{4.1}$$

$F(\alpha)$ corresponds to the desired quantile of the distribution's density funtion. σ_{t+1} is the 1-day ahead forecast provided by the models. To provide backtesting, the realized variance was computed applying the following formula:

$$RV = \sum_{i=1}^{N} r_{t,i}^2,$$

where N is the number of observations. Table 4.2 shows that the model that has provided the best performance in terms of the difference between expected and observed violations is, for all models, the GJR-GARCH. While for the SP500, WTI, and Euribor series, the forecasts tend to underestimate VaR, the models tend to overestimate VaR for the EurUsd. In order to verify the validity of the forecasts, the models were subjected to backtesting. The Unconditional and Conditional Coverage tests agree almost every time in rejecting or not the correct specification of the model. The Null hypothesis of the Unconditional Coverage test is that the frequency of realized hits is equal to that expected value α. The SP500 is the only model where the test rejects the null hypothesis for both levels of α. In this sense, likely, the economic crisis following the spread of the Covid-19 epidemic has negatively affected the performance of the model. For an $\alpha = 5\%$, the GJR-GARCH provides estimates that do not reject the null hypothesis of correct specification of the models for both WTI and EurUsd. The proposed VaR models for the latter were able to provide forecasts that almost always failed to reject the null hypothesis of the correct specification. The hits for the other models, although rather small in number, are not independent from each other. This result is also observed by the DQ test on the models' Hit Variables. For both confidence levels, all models reject the null hypothesis of no correlation of the sequence of Hit variables. Even if the lowest values of the loss functions are obtained from the forecasts on the Euribor series, it is not possible to establish a univocal hierarchy of models based on the loss function values. The loss function values indicate that for GARCH models, forecasts with a confidence level of 1% obtain higher values than those calculated on forecasts at 5%. This failure in model specification might be traced back to the pandemic crisis, which generated a period of sudden and strong increases in volatility. A subsample of 250 observations from 05/12/19 to 01/12/2020 was considered to check for differences in terms of VaR exceeding. The in-sample period was set to 160 observations. The rolling window considered is 20 days. The results of the estimates are proposed in the Table 4.1 and show that the GJR-GARCH model produces forecasts that obtain the "pass" the Traffic Light test. In addition, all the series do not reject the Unconditional and Conditional Coverage tests. These results confirm the sensitivity of the period used to produce the estimates and the size of the rolling window. The decrease in sample size makes it more challenging to assess the model's predictive ability. Despite the decrease in sample size, however, the predictive power of VaR improves significantly when an in-sample period, including the onset of the crisis in 2020, is considered.

	Level	Unconditional Coverage Reject Null	Conditional Coverage Reject Null	Observed Exceed	Expected Exceed	Observed Exceed %	Traffic Light Test Colour
SP501	5%	**NO**	**NO**	4	4.4	4.9%	**GREEN**
WTI	5%	**NO**	**NO**	6	4.4	6.7%	**GREEN**
Euribor3m	5%	**NO**	**NO**	4	4.4	4.5%	**GREEN**
EurUsd	5%	**NO**	**NO**	5	4.4	5.6%	**GREEN**

Table 4.1: This table shows the VaR backtesting results of the four assets during the year 2020. The in-sample period consisted in 160 observations, out of a total sample size of 250 observations. The rolling window was reduced from the previous analysis to 20 days. The table presents the results of the Unconditional and Conditional Coverage tests with a confidence level of 95% and the output of the Traffic Light test.

4.2 CAViaR models

The table below shows the values of the in-sample estimates of the various specifications of the CaViar models. Estimates calculation and forecasts were performed following the work of Engle and Manganelli (2004), using R for the basic code, with the addition of some C++ functions. The parameters are estimated using the Regression Quantile proposed in the paper. The table provides some first insights for understanding the subsequent forecasts. First of all, the coefficients β_1 are practically zero (in the order of 10^{-7}). The series relative to the SP500, the WTI, and the EurUsd, present a markedly high value of the autoregressive coefficient. Such values indicate that the clustering phenomenon is present and accentuated in the tails. In particular, during crises, there are often exceptionally high volatility peaks, during which it is possible to observe the phenomena of volatility clustering. Abad and Benito (2013) shown that even if GARCH and CaViaR models take into account volatility clustering, sudden increases in volatility can reduce the

		Level	Unconditional Coverage Reject Null	Conditional Coverage Reject Null	DQ test Reject Null	Loss Function	Observed Exceed	Expected Exceed	Observed Exceed %
SP500	**GARCH**	1%	**YES**	**YES**	**YES**	0.02481	22	10	2.2%
		5%	**NO**	**YES**	**YES**	0.01524	64	50	6.4%
	E-GARCH	1%	**YES**	**YES**	**YES**	0.02284	21	10	2.1%
		5%	**YES**	**YES**	**YES**	0.01442	69	50	6.9%
	GJR-GARCH	1%	**YES**	**YES**	**YES**	0.02503	20	10	2.0%
		5%	**YES**	**YES**	**YES**	0.01553	63	50	6.3%
	Beta-t-EGARCH	1%	**YES**	**YES**	**YES**	0.02283	21	10	2.1%
		5%	**YES**	**YES**	**YES**	0.01393	69	50	6.9%
WTI	**GARCH**	1%	**YES**	**NO**	**YES**	0.06663	22	10	2.2%
		5%	**NO**	**NO**	**YES**	0.04235	64	50	6.4%
	E-GARCH	1%	**YES**	**YES**	**YES**	0.05895	23	10	2.3%
		5%	**YES**	**YES**	**YES**	0.03832	71	50	7.1%
	GJR-GARCH	1%	**YES**	**YES**	**YES**	0.06742	22	10	2%
		5%	**NO**	**NO**	**YES**	0.04336	63	50	6.3%
	Beta-t-EGARCH	1%	**YES**	**YES**	**YES**	0.05908	23	10	2.3%
		5%	**YES**	**YES**	**YES**	0.03758	73	50	7.3%
Euribor3m	**GARCH**	1%	**YES**	**YES**	**YES**	0.00007	14	10	1.4%
		5%	**YES**	**NO**	**YES**	0.00005	44	50	4.4%
	E-GARCH	1%	**YES**	**YES**	**YES**	0.00089	12	10	1.2%
		5%	**YES**	**YES**	**YES**	0.00003	49	50	4.9%
	GJR-GARCH	1%	**NO**	**NO**	**YES**	0.00086	11	10	1.1%
		5%	**YES**	**YES**	**YES**	0.00054	52	50	5.2%
	Beta-t-EGARCH	1%	**YES**	**YES**	**YES**	0.00012	8	10	0.8%
		5%	**YES**	**YES**	**YES**	0.00004	52	50	5.2%
EurUsd	**GARCH**	1%	**NO**	**NO**	**YES**	0.01101	9	10	0.9%
		5%	**NO**	**YES**	**YES**	0.00701	38	50	3.8%
	E-GARCH	1%	**NO**	**NO**	**YES**	0.01098	6	10	0.6%
		5%	**NO**	**NO**	**YES**	0.00703	39	50	3.9%
	GJR-GARCH	1%	**NO**	**NO**	**YES**	0.01072	8	10	0.8%
		5%	**NO**	**NO**	**YES**	0.00686	47	50	4.7%
	Beta-t-EGARCH	1%	**NO**	**NO**	**YES**	0.01093	6	10	0.6%
		5%	**NO**	**NO**	**YES**	0.00693	39	50	3.9%

Table 4.2: This table shows the VaR backtesting results of the four assets. Each series contains 999 forecasted VaR values per each CaViaR specification. The Null hypothesis of the Unconditional, Conditional coverage tests and of the DQ test represents the correct specification of the model.

performance of these models .

		Parameter	Values
SP500	**Asymmetric**	β_1	0.00
		β_2	0.86
		β_3	0.06
		β_4	0.57
	Symmetric	β_1	0.00
		β_2	0.80
		β_3	0.51
	Adaptive	β_1	0.00
WTI	**Asymmetric**	β_1	0.01
		β_2	0.57
		β_3	0.51
		β_4	1.77
	Symmetric	β_1	0.01
		β_2	0.67
		β_3	0.97
	Adaptive	β_1	0.00
Euribor3m	**Asymmetric**	β_1	0.00
		β_2	0.88
		β_3	0.03
		β_4	0.43
	Symmetric	β_1	0.00
		β_2	0.74
		β_3	0.59
	Adaptive	β_1	0.00
EurUsd	**Asymmetric**	β_1	0.00
		β_2	0.94
		β_3	0.15
		β_4	0.13
	Symmetric	β_1	0.00
		β_2	0.94
		β_3	0.14
	Adaptive	β_1	0.00

Table 4.3: This table shows the in-sample parameter estimates of CaViaR model. The estimates are computed considering an in-sample period starting from 2nd January 2009 and containing 2000 observations.

Once the in-sample estimates were calculated, the one-day ahead forecasts of VaR were processed using a 100-day rolling window. The results have been represented in the graphs below. To provide a clearer comparison, only the CaViar estimates with α equal to 1% will be analyzed. From the visual analysis, the Asymmetric Caviar model follows the SP500's return pattern quite precisely. The breaches are mainly related to the last period when the economic crisis caused by the spread of the Covid-19 pandemic led to a sudden and prolonged increase in volatility. To assess the accuracy of the Caviar model estimates more accurately, the table below shows the backtests previously introduced. The models generally seem to perform well in forecasting, in agreement with the evidence found by Dionne (2013). None of the models is placed in the red zone of the Traffic Light test, showing that the overruns that are always moderately low. The estimates provided for Euribor, while delivering a relatively small percentage of overruns in the Symmetric Absolute Value and Asymmetric specifications, reject the null hypothesis of the Kupiec and Christoffersen tests. This divergence indicates that the breaches that occurred were not random but linked together by a dependency function. The DQ test do not reject the Null Hypothesis for the asymmetric specification, while for the adaptive the Null is always rejected. This means that the hits are serially uncorralated and have expected value equal to zero. The lowest values of loss function are provided by the Euribor, but the lovest value is provided by the SP500 the asymmetric specification. Another essential element that emerges from the estimates is the difference in performance between symmetric and asymmetric models: the latter produces the same or fewer breaches. This evidence confirms the results provided by the forecasts of the GARCH models.

		Level	Unconditional Coverage	Conditional Coverage	Traffic Light Test	DQ test	Loss Function	Observed Exceed	Expected Exceed	Observed Exceed %
			Reject Null	Reject Null	Colour	Reject Null				
	Adaptive	1%	NO	NO	GREEN	YES	5.91E-04	14	10	1.4%
SP500	Symmetric Absolute Value	1%	NO	NO	GREEN	YES	2.89E-04	14	10	1.4%
	Asymmetric	1%	NO	NO	GREEN	NO	2.73E-04	11	10	1.1%
	Adaptive	1%	NO	YES	YELLOW	YES	1.91E-03	15	10	1.5%
WTI	Symmetric Absolute Value	1%	NO	NO	YELLOW	NO	7.78E-04	19	10	1.9%
	Asymmetric	1%	NO	NO	YELLOW	NO	9.05E-04	18	10	1.8%
	Adaptive	1%	NO	YES	GREEN	NO	2.00E-06	9	10	0.9%
Euribor3m	Symmetric Absolute Value	1%	YES	YES	GREEN	NO	1.00E-06	8	10	0.8%
	Asymmetric	1%	YES	NO	GREEN	NO	2.00E-06	8	10	0.8%
	Adaptive	1%	NO	NO	GREEN	YES	1.33E-04	5	10	0.5%
Eurusd	Symmetric Absolute Value	1%	NO	NO	GREEN	NO	1.28E-04	7	10	0.7%
	Asymmetric	1%	NO	NO	GREEN	NO	1.28E-04	7	10	0.7%

Table 4.4: This table shows the VaR backtesting results of the four assets. Each series contains 999 forecasted VaR values per each CaViaR specification. The Null hypothesis of the Unconditional, Conditional coverage tests and of the DQ test represents the correct specification of the model.

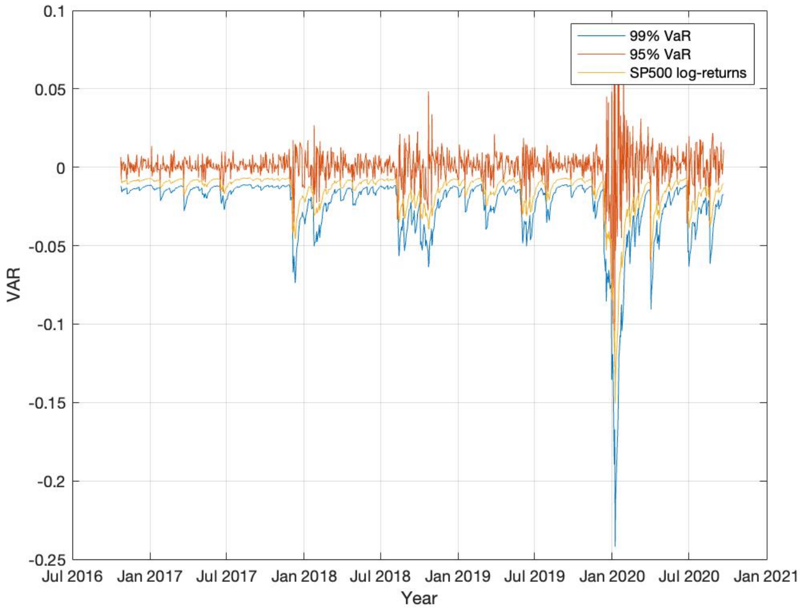


Figure 4.2: Realized returns of the SP500 and one-day ahead forcasted VaR with GJR-GARCH(1,1), with $\alpha = 0.01$ and $\alpha = 0.05$.

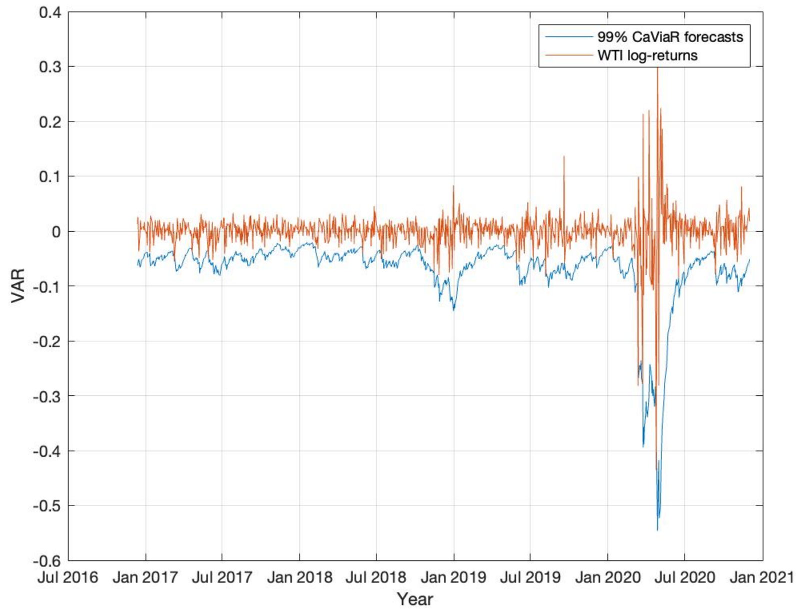


Figure 4.3: Realized returns of the WTI and one-day ahead Asymmetric Caviar forcasts with $\alpha = 0.01$.

Conclusive remarks

This thesis presented a comparison of different models for estimating VaR based on both a parametric and a semi-parametric approach. The time series showed that data did not originate from normal distributions since they presented particularly heavy tails. The models were described and analyzed from a theoretical and empirical point of view, highlighting their strengths and weaknesses. Two classes of models were considered: the GARCH class models, belonging to the parametric approach for calculating VaR, and the CaViaR semi-parametric models. The empirical analysis was conducted by dividing the samples into in-sample and out-of-sample periods, making estimates and one-day-ahead forecasts, respectively. A 100-day rolling window was used to compute the analysis. The empirical results confirm that the most effective GARCH estimates were obtained using GED and t-student innovations. The classical GARCH model showed performances very similar to the more sophisticated versions, demonstrating the good predictive capacity of even the most straightforward system. However, the best results in terms of the forecast were obtained by the GJR-GARCH model. This evidence underlines the presence of the leverage effect within the time series under examination. However, although the number of breaches in these models is relatively small, most of them reject the null hypotheses of the Christoffersen and Kupiec tests. The DQ test agrees with the unconditional and conditional tests. These results shows that the hits are not independent from each other. By restricting the sample to 2020 to include the economic crisis from Covid-19 in the in-sample period, the accuracy of the GARCH models increases significantly. This result may indicate that the breaches in the whole sample can be traced back to the sudden increase in volatility in 2020. Considering the good results obtained by the CaViaR models, it is also possible that the difference in performance (as measured by the backtests) can be explained by an incorrect specification of the data distribution. Although it can be concluded with reasonable certainty that the returns are not distributed according to a normal distribution, identifying the correct distribution is not so easy. The outbreak of the crisis in the first half of 2020 has also led to an evident change in the time series trend, to which the GARCH models have struggled to adapt. However, it should be noted that the asymmetric CaViaR model with only one lag uses four parameters to estimate the quantile. The presence of such a high number of parameters to estimate can be particularly computationally intensive, making the use of simple GARCH models more attractive. The CaViaR model was analyzed in its Symmetric Absolute Value, Asymmetric and Adaptive specifications. The results provided by the latter proved able to pass the traffic light test, recording a minimal number of surpluses. In terms of VaR estimates at 99%, the CaViaR models showed superior performance than GARCH. Also, the best performing specifications in this class of models are the asymmetric ones. Therefore, the importance of the leverage effect in a period of high volatility, such as the first semester of 2020, is evident.

Summary

Risk management within organizations has always been a subject of study, but it was only in the first half of the 20th century that it began to be structured into a real discipline. From the need to find ways to measure risk, to be able to exploit the opportunities and mitigate the dangers, Risk Management was born. This need became even more evident after the 1987 crisis. On 19 October 1987, after more than 44% growth in the first months of the year, the Dow Jones Industrial Average suffered a record loss of 22.6%. The increasing occurrence of crises at the global level pushed countries worldwide to seek standard prevention measures. Moreover, the use of financial instruments capable of triggering the exchange of securities on the occurrence of certain conditions made it clear that new risk measures needed to be developed. In particular, to ensure the financial system's stability, it was necessary to develop methods to estimate possible losses. In response to this need, countries worldwide sought shared measures to counter the onset of global crises. These efforts took the form of the Basilea I Accords in 1988, which set minimum standards in terms of capital financial that each bank had to maintain. In the first draft of the Accords, reserves had to be maintained above the Cook Ratio of 8%. However, banks were allowed to use internal models to assess the riskiness of their positions. Among these measures, Value at Risk (VaR) was the most relevant, and it became the basis for calculating the capital requirements of later versions of the Basel Accords. One way of assessing the risk of a position is to consider the maximum loss that can be realised. In 1996, JP Morgan's RiskMetrics Technical Document provided the following definition of VaR:

> "Value at Risk is a measure of the maximum potential change in value of a portfolio of financial instruments over a pre-set horizon. VaR answers the question: how much can I lose with α% probability over a given time horizon."

VaR is defined on the basis of two elements: the time horizon and the confidence level α. In the Basel I agreements, VaR had to be calculated according to the following criteria:

- time horizon was at ten market days or two weeks;
- confidence level was fixed at 99%;

- historical data used to produce estimates had to go back at least one year, with an update of the parameters every three months.

As already mentioned, in 1996, with the Basel II agreements, the criteria for calculating capital requirements were redesigned. In particular, the following formula was defined:

$$CA_t^{B2} = max(VaR_{t-1}, m_c\overline{VaR}_{t-1}),$$

where $\overline{VaR}_{t-1}$ is an average of the VaR of the previous 60 days and m_c is the multiplication factor. This factor assumes a value of at least 3, but if the model used to calculate VaR obtains negative results in the backtest, then it increases. Thus, the amount of capital to be held increases if the criterion used to estimate VaR gives inaccurate results. To assess the accuracy of VaR estimates, the Traffic Light test was introduced. This test assigned a color to the models according to the accuracy with which VaR was estimated: green models were of no particular concern and could be used; those that were yellow needed to be analyzed further and tweaked; and those that were red were rejected. The 2008 crisis made it clear that the capital requirements outlined in 1996 were not adequate, so a new criterion for calculating reserves was introduced in the Basel 2.5 agreements. The formula was modified as follows:

$$CA_t^{B2.5} = max(VaR_{t-1}, m_c\overline{VaR}_{t-1}) + max(SVaR_{t-1}, m_s\overline{SVaR}_{t-1}),$$

where $\overline{SVaR}$represents the stressed VaR, i.e., the VaR calculated during a financial crisis. Thus, the capital to be held as reserves compared to Basel II was at least doubled. Having briefly outlined the history of VaR to understand its importance, this thesis proposes an empirical analysis conducted on four financial assets to assess the differences in accuracy of different models to estimate VaR. The four assets represent an essential part of market risk: the SP500, WTI, the three-month Euribor, and the Euro-Dollar exchange rate.
Many models for calculating VaR stem from the scientific debate, and these are continually evaluated to compare their accuracy. However, the Delta Normal Method and Historical Simulation are among the simplest models in terms of computational simplicity. In the second chapter, some of the methods used to calculate VaR are analyzed from a formal point of view. The Delta Normal Method assumes that returns are normally distributed to derive volatility estimates in a relatively simple way from a computational point of view. However, several studies have shown that this method often produces inaccurate estimates, as the assumption of normality of returns is unrealistic. The Historical Simulation method uses a different approach, approximating the distribution of returns based on the distribution of previous realizations. The main problem lies in the subjectivity of the time window one decides to use as a basis for estimation. A more sophisticated approach than the previous ones

is used by the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) class of models, introduced by Bollerslev (1986). This class of models estimates the conditional variance at time t through an outlier component and the variances in previous periods. This class of models has become widely used, as it can return quite reliable volatility estimates using a relatively low number of parameters. Several modifications of GARCH have been produced over time. In this paper, we will consider, together with GARCH, the E-GARCH Nelson (1991), GJR-GARCH Glosten et al. (1993), and T-Beta-EGARCH models Harvey and Chakravarty (2008). These variants manage to improve the performance of the classical GARCH, capturing phenomena such as the leverage effect or the presence of outliers.

The other class of models analyzed is CAViaR, introduced by Engle and Manganelli (2004). Unlike GARCH, which estimates volatility, CAViAR directly models the quantiles of distributions. In this work, the Adaptive, Asymmetric, and Symmetric specifications of CAViaR are analyzed.

The sample contains 3000 observations from Bloomberg from 2 January 2009 to 1 December 2021. Before proceeding with the VaR estimation, the descriptive statistics of the time series are presented. The data show an average very close to zero, in agreement with the literature, as well as a particularly pronounced kurtosis. The Kolmogorov-Smirnov (Massey Jr (1951)), Anderson-Darling (Anderson and Darling (1954)), and Cramer-von Mises (Darling (1957)) tests show that the data do not come from normal distributions. These results confirm the evidence in the scientific literature that financial returns come from distributions with fat tails.

In order to carry out the model forecasts, the samples are divided into an in-sample period and an out-of-sample period. During the in-sample period, which includes 2000 observations, data are estimated from which forecasts are calculated for the subsequent out-of-sample period. GARCH model estimates are made by maximizing log-likelihood functions, while CaViaR estimates are made using Quantile Regression. Once the forecasts are obtained, they will be subjected to several backtests to assess their accuracy. The proposed backtests use different approaches. The Unconditional (Kupiec (1995)) and Conditional Coverage (Christoffersen (1998)) tests, respectively assess that VaR violations occur independently, with a frequency similar to the $alpha$ confidence level. The Basel Committee in 1996 proposed an intuitive system for backtesting VaR called Traffic Light. To evaluate the goodness of the forecasts, the Traffic Light test considers the difference with the theoretical number of breaches and the actual number of violations. The models are classified within a hierarchical color system based on the number of breaches obtained. Intuitively, green does not cast doubt on the accuracy of the model; yellow expresses an inconclusive doubt; finally, red indicates that the model is inadequate and should be modified. A further proposed backtest is the DQ test introduced by Engle and Manganelli (2004). This method simultaneously tests that VaR violations are not serially correlated and occur with a frequency close to that of the α confidence level. Finally, the loss function of González-Rivera et al. (2004) is proposed, which allows ordering the accuracy of two or more models.

The empirical analysis is implemented in Matlab and R, with some functions in C++.

The empirical results confirm that the most effective GARCH estimates were obtained assuming that returns do not come from a normal distribution. The classical GARCH model showed performances very similar to the more sophisticated versions, demonstrating the good predictive capacity of even the most straightforward system. However, the best results in terms of the forecast were obtained by the GJR-GARCH model. This evidence underlines the presence of the leverage effect within the time series under examination. Some studies have shown that adverse shocks can cause more significant increases in volatility than positive shocks. This is attributed to the leverage effect, which can be defined as the negative correlation between past returns and future volatility. However, although the number of breaches in these models is relatively small, most of them reject the null hypotheses of the Christoffersen and Kupiec tests. The DQ test agrees with the Unconditional and Conditional tests. These results shows that the hits are not independent from each other. By restricting the sample to 2020 to include the economic crisis from Covid-19 in the in-sample period, the accuracy of the GARCH models increases significantly. This result may indicate that the breaches in the whole sample can be traced back to the sudden increase in volatility in 2020. Considering the good results obtained by the CaViaR models, it is also possible that the difference in performance (as measured by the backtests) can be explained by an incorrect specification of the data distribution. Although it can be concluded with reasonable certainty that the returns are not distributed according to a normal distribution, identifying the correct distribution is not so easy. The outbreak of the crisis in the first half of 2020 has also led to an evident change in the time series trend, to which the GARCH models have struggled to adapt. However, it should be noted that the asymmetric CaViaR model with only one lag uses four parameters to estimate the quantile. In terms of VaR estimates at 99%, the CaViaR models showed superior performance than GARCH. Also, the best performing specifications in this class of models are the asymmetric ones. Therefore, the importance of the leverage effect in a period of high volatility, such as the first semester of 2020, is evident.

Bibliography

P. Abad and S. Benito. A detailed comparison of value at risk estimates. *Mathematics and Computers in Simulation*, 94:258–276, 2013.

P. Abad, S. Benito, and C. López. A comprehensive review of value at risk methodologies. *The Spanish Review of Financial Economics*, 12:15–32, 2014.

C. Acerbi and D. Tasche. Expected shortfall: a natural coherent alternative to value at risk. *Economic notes*, 31(2):379–388, 2002.

T. W. Anderson and D. A. Darling. A test of goodness of fit. *Journal of the American statistical association*, 49(268):765–769, 1954.

D. Ardia, K. Boudt, and L. Catania. Generalized autoregressive score models in r: The gas package. *Journal of Statistical Software*, 88(6):1–28, 2019.

G. Barone-Adesi and K. Giannopoulos. Non parametric var techniques. myths and realities. *Economic Notes*, 30(2):167–181, 2001.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

T. Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, pages 542–547, 1987.

G. E. Box and D. A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526, 1970.

C. Brooks and S. P. Burke. Information criteria for garch model selection. *The European journal of finance*, 9(6):557–580, 2003.

P. Christoffersen. Backtesting. *Available at SSRN 2044825*, 2008.

P. F. Christoffersen. Evaluating interval forecasts. *International economic review*, pages 841–862, 1998.

R. Cont. Volatility clustering in financial markets: empirical facts and agent-based models. In *Long memory in economics*, pages 289–309. Springer, 2007.

H. Cramr. On the composition of elementary errors. *Scandinavian Actuarial Journal*, 1:141–180, 1928.

D. Creal, S. J. Koopman, and A. Lucas. Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795, 2013.

J. Daníelsson, B. N. Jorgensen, G. Samorodnitsky, M. Sarma, and C. G. de Vries. Fat tails, var and subadditivity. *Journal of econometrics*, 172(2):283–291, 2013.

D. A. Darling. The kolmogorov-smirnov, cramer-von mises tests. *The Annals of Mathematical Statistics*, 28(4):823–838, 1957.

G. Dionne. Risk management: History, definition, and critique. *Risk management and insurance review*, 16(2):147–166, 2013.

D. Duffie and J. Pan. An overview of value at risk. *Journal of derivatives*, 4(3):7–49, 1997.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.

R. F. Engle and S. Manganelli. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of business & economic statistics*, 22(4):367–381, 2004.

R. H. Gerlach, C. W. Chen, and N. Y. Chan. Bayesian time-varying quantile forecasting for value-at-risk in financial markets. *Journal of Business & Economic Statistics*, 29(4):481–492, 2011.

L. R. Glosten, R. Jagannathan, and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801, 1993.

T. Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

G. González-Rivera, T.-H. Lee, and S. Mishra. Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of forecasting*, 20(4):629–645, 2004.

P. R. Hansen and A. Lunde. A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*, 20(7):873–889, 2005.

A. C. Harvey and T. Chakravarty. Beta-t-(e) garch. 2008.

R. Ibragimov. *New majorization theory in economics and martingale convergence results in econometrics*. Yale University, 2005.

B. Jean-Philippe, M. Andrew, and P. Marc. The leverage effect in financial markets: retarded volatility and market panic. *February, Service de Physique de l'EtatCondens'e, Centre d'etudes de SaclayOrme des Merisiers*, 91191:109–111, 2008.

P. Jorion. Risk2: Measuring the risk in value at risk. *Financial analysts journal*, 52 (6):47–56, 1996.

S. Kaplan and B. J. Garrick. On the quantitative definition of risk. *Risk analysis*, 1 (1):11–27, 1981.

R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/1913643`.

P. Kupiec. Techniques for verifying the accuracy of risk measurement models. *The J. of Derivatives*, 3(2), 1995.

T. J. Linsmeier and N. D. Pearson. Value at risk. *Financial Analysts Journal*, 56(2): 47–67, 2000. doi: 10.2469/faj.v56.n2.2343. URL `https://doi.org/10.2469/faj.v56.n2.2343`.

J. Lintner. Security prices, risk, and maximal gains from diversification. *The journal of finance*, 20(4):587–615, 1965.

G. M. Ljung and G. E. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

H. Markowitz. Portfolio selection (1952) 7 j.

F. J. Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.

J. Mossin. Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, pages 768–783, 1966.

D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.

C. Pérignon and D. R. Smith. The level and quality of value-at-risk disclosure by commercial banks. *Journal of Banking Finance*, 34(2):362–377, 2010. ISSN 0378-4266. doi: https://doi.org/10.1016/j.jbankfin.2009.08.009. URL `https://www.sciencedirect.com/science/article/pii/S0378426609001940`.

M. Pritsker. The hidden dangers of historical simulation. *Journal of Banking & Finance*, 30(2):561–582, 2006.

H. Raiffa and R. D. Luce. Games and decisions: Introduction and critical survey. *New York and London*, 1957.

M. Rocco. Extreme value theory for finance: A survey. 2010.

E. Şener, S. Baronyan, and L. A. Mengütürk. Ranking the predictive performances of value-at-risk estimation methods. *International Journal of Forecasting*, 28(4): 849–873, 2012.

W. F. Sharpe. Mutual fund performance. *The Journal of business*, 39(1):119–138, 1966.

C. Sinclair, B. Spurr, and M. Ahmad. Modified anderson darling test. *Communications in Statistics-Theory and Methods*, 19(10):3677–3686, 1990.

J. L. Treynor. Market value, time, and risk. *Time, and Risk (August 8, 1961)*, 1961.

Acknowledgement

I would like to express my gratitude to Dr. Pierluigi Vallarino for all his help and support with this thesis. I would like to thank my family and friends for always being there for me in my time of need.