Quantile regression applied to the estimation of Value at Risk. An application to the Spanish market

Inés Ortega Palacios

Trabajo de fin de Máster
Universidad de Zaragoza

Directores del trabajo: J. Tomás Alcalá Nalvaiz y Manuel Salvador Figueras

Septiembre de 2018

Prologue

The general aim of this work is to give a description of the estimation of Value at Risk (VaR) using Quantile Regression models. In order to be able to measure the performance of these models, different classical approaches for the estimation of VaR have been considered. These models will be used to predict the VaR for the series of returns of the Euro-Dollar Exchange Rate and three companies from the IBEX35: Banco Santander, Endesa and Indra.

The first studies in Value at Risk estimation were written by Markowitz (see [34]) and Roy (see [41]), who intended their VaR measures for theoretical portfolio optimization work. However, VaR studies became a main object of interest in 1980, when the Securities and Exchange Commission (SEC) enacted the first regulatory measures related to Value at Risk conditions for institutions. Global standards are nowadays published by the Basel Committee on Banking Supervision (BCBS). In the Basel Accords, indications for the measurement of VaR and forecasts are given and financial institutions lean on these requirements to calculate financial risk, capital changes and contingency plans.

The Value at Risk can be described as the worst expected loss over a given horizon under normal market conditions at a given confidence level. A historical introduction to Value at Risk and main definitions of VaR, Returns and VaR estimation are given in the first chapter.

Methods for estimating VaR can be divided in two main groups: parametric and non-parametric models. Classical methods from both groups are presented in the second chapter, in which explicit formulas of Value at Risk forecasts are given for each model. In order to measure their accuracy, classical performance tests and cost evaluation functions are given. Performance tests measure the dependence of excesses over the VaR forecast and their independence and cost functions evaluate the scale of these excesses.

The quantile regression approach to forecast Value at Risk is presented in the third chapter. Firstly, an introduction to Quantile Regression and its estimation of parameters is given and later, the Conditional Autoregressive Value at Risk (CAViaR) model is introduced. This model, described by Engle and Manganelli [21], tries to fit the VaR by focusing on modelling the quantiles of the distribution of returns instead of the entire distribution. The Dynamic Quantile (DQ) test proposed in the same article, measure whether the excesses over the VaR can be fitted to a linear quantile regression model. Thus, a performance test based on Quantile Regression is given.

In order to estimate the Value at Risk of the return series described above, estimation of distributions of all the models, VaR forecasts and performance evaluations are presented in the last chapter. The statistical programming language R is the main tool used in this work.

Resumen

La medición y modelización del riesgo en las instituciones financieras se ha convertido en una importante área de desarrollo en el sector de la banca. El riesgo de mercado se define como el riesgo asumido en una inversión provocado por las fluctucaciones de los valores de mercado. Sistemas regulatorios como el Comité de Basilea, y las instituciones financieras han definido diferentes medidas para este riesgo.

El Valor de Riesgo (VaR, por sus siglas en inglés), se define como la pérdida máxima de un inversor que se puede producir en un periodo de tiempo dado y con una cierta probabilidad. Es decir, el VaR es el valor mínimo que casi seguramente no traspasará la cartera de un inversor.

Los primeros resultados relacionados con la medición formal de riesgo se atribuyen a Markowitz (ver [34]) y Roy (ver [41]), que fueron pioneros en la investigación de la teoría de carteras y definieron los primeros indicadores del VaR.

Los primeros documentos regulatorios que recogen el Valor de Riesgo, fueron publicados por la Comisión de Seguridad e Intercambio (SEC), en los que se definieron unos requisitos por los cuales las instituciones financieras debían medir el riesgo que se tomaban en las operaciones con un 95 % de confianza a 30 días.

En 1988, el Comité de Basilea (BCBS), formado por estados miembros del G10 y Luxemburgo publicó Basilea I, un documento regulatorio en el que se exponían recomendaciones y reglas no obligatorias para la supervisión bancaria. En la actualidad, 27 países forman parte de este Comité y estos estÁ¡ndares globales para la regulación y previsión del riesgo para posibilitar la creación de planes de contingencia bancarios, han derivado en la publicación de Basilea II (2004) y Basilea III (2016).

Sea $\{r_t\}_{t=1}^n$ la serie temporales de retornos de un portfolio y sea n el número de elementos de la muestra. El Valor de Riesgo en el tiempo t con un nivel de confianza $1-\alpha$ y horizonte temporal $h \geq 1$, $VaR_{t,\alpha}(h)$, es el valor que satisface:

$$\wp(r_{t+h} \leq VaR_{t,\alpha}(h)|\Omega_t) = \alpha$$

donde Ω_t es la información de la serie en t.

El primer enfoque que se considera en este trabajo para el cálculo del Valor de Riesgo es la utilización de métodos clásicos. Los primeros métodos que se presentan son los más utilizados, por su simplicidad y rápida computación: el método histórico y el método Delta o Vairnaza-Covarianza. Dada una serie de retornos, $\{r_t\}_{t=1}^n$, el VaR del método histórico con probabilidad α en $n+1$ es el α-cuantil de la serie de retornos. Para el caso del método Delta, el VaR en t con horizonte h se define como:

$$VaR_{t,\alpha}(h) = \bar{r}_{t+h|t} + \Phi(\alpha) \cdot \bar{\sigma}_{t+h|t}$$

donde $\bar{r}_{t+h|t}$ and $\bar{\sigma}^2_{t+h|t}$ son la media y la varianza muestral de los retornos de la muestra con información hasta t y Φ es la distribución normal estándar.

Otra aproximación al problema de la estimación del Valor de Riesgo son los modelos GARCH. Estos modelos consideran que la varianza de uns serie de retornos no es constante, por lo tanto, modelan la heterocedasticidad de la serie. Se presentan diferentes tipos de modelos GARCH (IGARCH, GJR-GARCH, EWMA, etc.). Cada uno de ellos posee propiedades particulares y, dependiendo de la naturaleza de la serie, pueden resultar buenos modelos a la hora de estimar el VaR.

El último enfoque clásico son los modelos basados en teoría de valores extremos. Se presentan dos metodologías diferentes para abordar el problema: el análisis de máximos (o mínimos) y los excesos sobre umbral.

En la primera de ellas, dada una muestra de retornos de longitud n, se escoge una constante q, denominada longitud de bloque, que determina el número de retornos en cada subgrupo de la muestra, de manera que se tiene un total de n/q subgrupos. La muestra de valores utilizada para la estimación de la función de distribución de máximos (mínimos), denominada distribución de valores extremos generalizada (GEV, por sus siglas en inglés), es el conjunto de máximos (mínimos, respectivamente) de cada uno de los subconjuntos. Por lo tanto, el longitud muestral es n/q. En el caso de que n/q no sea entero, se eliminan las correspondientes primeras observaciones de la muestra.

El método de excesos sobre umbral considera como valores extremos aquellos retornos que excedan un límite (o umbral) fijo. La distribución de los excesos se denomina Pareto Generalizada. Una dificultad que puede encontrarse en este modelo es la elección del umbral. En el caso de que el umbral no sea adecuado, puede suceder que no haya suficiencia muestral para el ajuste del modelo, o bien se consideren como extremos valores que realemnte no lo son. En series financieras, es habitual encontrar la selección del umbral a partir de los cuartiles de la muestra. Es decir, dado un porcentaje de población p, se toma como umbral el p-cuantil de la serie de retornos, de manera que se consideran como extremos un $1-p$ por ciento de la población.

Como método alternativo a los modelos clásicos de estimación del Valor de Riesgo, el tercer capítulo ofrece una introducción y visión global a la regresión cuantil. Dada X variable aleatoria, y F su función de distribución, el cuantil de orden τ de X, q_τ se define como el valor que satisface

$$\wp(X \leq q_\tau) \geq \tau$$

y

$$\wp(X \geq q_\tau) \geq 1-\tau$$

En el caso de una muestra, $x_1, \ldots, x_n$, los cuantiles de la muestra pueden encontrarse como la solución de un problema de optimización. El valor óptimo del problema, $\widehat{x}$, responde a la la ecuación:

$$\int \rho_\tau(x-\widehat{x})\, dF_n(x) = n^{-1} \sum_{i=1}^{n} \rho_\tau(x_i - \widehat{x})$$

donde ρ_τ es la función de pérdida descrita por

$$\rho_\tau(u) = u(\tau - I(u<0))$$

Por lo tanto, el problema de encontrar el cuantil de una distribución dada una muestra, se puede reducir a un problema de optimización lineal. Dada una muestra $\{(X_1,Y_1), \ldots, (X_n,Y_n)\}$, el regresor cuantil $\widehat{\beta}_\tau$ de orden τ se define como

$$\widehat{\beta}_\tau = \min_{\beta \in \mathbb{R}^d} \sum_{i=1}^{n} \rho_\tau\left(Y_i - X_i^T \beta\right)$$

El principal método de estimación de este problema es planteado en el capítulo. La definición formal del problema de optimización como un problema de función objetivo con restricciones permite analizar tanto este problema como su dual y resolverlo utilizando diferentes métodos: algoritmo Simplex, método del punto interior, métodos de planos de corte, etc.

La regresión cuantil es un método robusto, es decir, no es sensible a datos atípicos o a ciertas asumciones relacionadas con la población objetivo, al contrario que con lo que ocurre con la regresión lineal clásica. Esto es debido a la naturaleza de la función objetivo del problema de minimización de la regresión cuantil.

El modelo de estimación del Valor de Riesgo planteado en el tercer capítulo toma el nombre de modelo de Valor de Riesgo Condicional Autoregresivo (CAViaR, por sus siglas en inglés). Desarrollado

por Engle y Manganelli (see [21]), este modelo utiliza la regresión cuantil para estimar el VaR de una muestra, es decir, en vez de estimar la función de distribución para calcular el VaR a partir del cuantil de la distribución, estima el cuantil de la distribución. Además, el modelo está planteado en base a la relación autoregresiva de las series de retorno.

Sea β un vector de parámetros desconocidos y sea $q_t(\beta,\alpha) \equiv q_t(\mathbf{x}_{t-1},\beta)$ el tiempo t α-cuantil de la distribución de retornos en $t-1$. El método general CAViaR se define de la siguiente forma:

$$q_t(\beta,\alpha) = \beta_0 + \sum_{i=1}^{s} \beta_i q_{t-i}(\beta,\alpha) + \sum_{j=s+1}^{r+s+1} \beta_j l(\mathbf{x}_{t-j})$$

donde$p = s+r+1$ es la dimensión de β y l es la función que relaciona el cuantil del tiempo t con un número finito de lags.

Finalmente, con el propósito de analizar la eficacia de este modelo frente a los métodos clásicos, se estima el Valor de Riesgo para la serie de cambio de divisa Euro-Dolar, y las series de retorno de los valores de cierre en bolsa de tres empresas del IBEX 35: Banco Santander, Endesa e Indra.

Las estimaciones realizas se comparan utilizando tests de rendimiento, como el test de Kupiec, el test de Christoffersen o el Ratio de excesos, y funciones de pérdida como la función de Caporin. Además, se presenta el test de Cuantiles Dinámicos (DQ), basado en la regresión cuantil.

Contents

Chapter 1

Value at Risk: An introduction

Risk measurement in financial institutions has become an important area of research and modelling for banks. This measurement ensures the solvency and the good functioning of financial institutions. Since the instability of a single institution can affect the stability of the whole banking system, a regulatory framework has been developed by the Basel Committee on Banking Supervision. These rules shall be followed by all banking institutions of the European Union.

The Market Risk can be defined as the risk taken from losses produced by changes in markets' rates such as interest or stock returns. This risk is one of the most important types of risk. The Basel Committee establishes a number of rules in order to measure this market risk. Different measures such as the Value at Risk and the Expected Shortfall can be used. The confidence level for forecasting should be at least of 99% and the period of historical observations (sample) used to forecast should be set at a minimum of one year. In addition, EU institutions are suggested to test their forecasts with a backtesting process.

As Prof. Tsay well explain in [44], Value at Risk can be defined as the maximal loss of a financial position during a given time period for a given probability, that is, it is the loss that the investor is fairly sure will not be exceeded in the case that the portfolio is held over some period of time. Thus, VaR can be used to elaborate a financial contingency plan. This results to be very useful as EU supervisors suggest financial institution to have a funding plan in case of a market crash.

1.1. Introduction

1.1.1. Motivation

Several large financial losses in this previous decades have led to the growth of extreme modelling researching. But, how can risk be described in the first place?

The concept of risk has been historically debated from a lot of points of view. The existence of different definitions of risk is a proof of this complexity. Taking risk can mean having the possibility of suffering damage. It also can mean the possibility of loosing something or being at disadvantage against something.

If the risk is measurable, it can also be modelled. This is known as *Risk Management*. In a financial activity, a Market Risk appears. This kind of risk is the one which is going to be forecast and analysed in this dissertation.

The Market Risk can be defined as the probability of having potential losses due to a negative evolution of asset prices in which the trader has decided to invest. Three different risks can be found: Interest rate risk, exchange rate risk and stock price risk. In order to minimise the impact of the market risk in a trader's portfolio, different measurements and models can be identified and controlled.

1.1.2. History of VaR

The first studies related to measure risk are attributed to Francis Edgeworth in 1888. He made important contributions to the statistical theory, using historical data as the basis to model future risky events.

In 1952, Harry Markowitz (see [34]), who was awarded the Nobel Prize in Economics in 1990 for his pioneering research in the area of portfolio theory, and independently three months later Arthur D. Roy (see [41]) proposed VaR indicators, which were surprisingly similar. In their researches, both of them tried to optimise the profit of a portfolio at a given level of risk.

During the 70th to 80th years many new financial products were created. This represented a new challenge for modelling of risk. The new financial products did not have historical data, through which the risk could be approximated. One option was to find approximately similar financial products and try to derive the risk from them. For example, the abolition of the monetary system in 1971 has resulted in a foreign exchange forward market. These new challenges stimulated the development of an easily understandable and reliable indicator of risk.

The first regulatory measures that evoke Value at Risk, though, were initiated in 1980, when the Securities and Exchange Commission (SEC) tied the capital requirements of financial service firms to the losses that would be incurred, with 95% confidence over a thirty-day interval, in different security classes; historical returns were used to compute these potential losses.

Although the first publications about predecessors of VaR date back to the 20th century, the credit for the use of current VaR attributed mainly to US investment bank JP Morgan. This company was the first one that introduced "Risk Metrics" in order to describe the measure of risk using financial data. The origin of the name "Value at risk" is unknown. In the 90s the Value at Risk was also known as "dollars at risk", "capital at risk", "income at risk", "earnings at risk" or "money at risk".

In 1994, J. P. Morgan published a technical document of this system. This was followed by the mass acquisition of the system by many institutions. VaR was popularized as the risk measure of choice among investment banks looking to be able to measure their portfolio risk for the benefit of banking regulators.

The SEC requirements were not the only initiatives created in order to control financial risk. In 1974, members of the G10 and Luxembourg founded the Basel Committee on Banking Supervision (BCBS). At the present time, 27 countries, including Spain (full member of the committee since 2001), and three observers are members of this committee. The committee formulates general supervision rules and makes recommendations for banking supervision. Even though recommendations are not compulsory, all the members (countries and authorities) tend to implement the suggestions made by the committee. These global standards have been included in the Basel Capital Accords: Basel I (1988), Basel II (2004) and Basel III (2016).

According to the Basel Committee, the VaR methodology could be used by financial institutions to calculate capital changes with respect to their financial risk. Its intention was to provide a tool to calculate a certain backup money in order to be able to affront non-payments and losses.

However, institutions soon realised that the VaR was useful not only to calculate previsions suggested by the Basel committee but also to manage other risk measures. Nowadays, the calculation of VaR has been set up in lots of different aspects and business activities.

1.2. VaR definition and elements

According to Philippe Jorion (see [28]), "VaR measures the worst expected loss over a given horizon under normal market conditions at a given level of confidence. For instance, a bank might say that the daily VaR of its trading portfolio is $1 million at the 99 percent confidence level. In other words, under normal market conditions, only one percent of the time, the daily loss will exceed $1 million."

Notice that there can be two different views that can explain the VaR concept. On the one hand, VaR can be seen as a measure of loss associated with an extreme event under normal market conditions. On

the other hand, VaR is the minimal loss under extreme market conditions. Both approaches lead to the same measure of VaR.

Value at Risk has two basic parameters:

- The significance level: it represents the probability for which the VaR is expected.
- The risk horizon: it is the period of time over which the Value at Risk (potential loss) is calculated.

The confidence level can be chosen by the trader, but also by an external agent. For example, in the case of banks: Under Basel Accords, if a bank uses internal VaR models, the risk should be measured at a determined confidence level (in Basel III, the confidence level was fixed at 0.99). Both the confidence level and the Risk horizon are given by the Bank for International Settlements of Basel. Its purpose is to ensure that financial institutions hold enough capital on account to meet obligations and absorb unexpected losses.

The different methodologies developed to calculate the VaR can be presented in four different groups (see [3]):

- Parametric methods: GARCH and Risk-Metrics.
- Non-parametric methods: historical simulations.
- Monte Carlo simulations.
- Semi-parametric methods: extreme value theory and CAViaR.

1.2.1. VaR elements

Definition 1.1 (Value at Risk)**.** *Let $\{r_t\}_{t=1}^n$ be the time series of portfolio returns and n denote the sample size. The Value at Risk of the financial position at time t with confidence level $1-\alpha$ for a given horizon h, $VaR_{t,\alpha}(h)$, is the value that satisfies*

$$\wp(r_{t+h} \leq VaR_{t,\alpha}(h)|\Omega_t) = \alpha$$

where Ω_t is the information available at time t.

From the definition, one can tell that $VaR_{t,\alpha}(h)$ is the α-quantile of the conditional distribution $r_{t+h}|\Omega_t$. Thus, the potential loss at time $t+h$ is less or equal to $VaR_{t,\alpha}(h)$ with probability α.

All VaR measurement approaches use a similar scheme:

1. Selection of basic parameters:
   - Time horizon (h): Period of time at which the VaR is calculated.
   - Confidence level ($1-\alpha$): it is the probability of incurring to a loss which is at least lower than the estimated Value at Risk.
   - Time units (t): in general, time is measured in days.
   - Data window: it is the number of time units which are used to estimate the VaR at a certain time t. It determines Ω_t, the information available for the given window. That is, the data window indicates the set of observations known at time t, Ω_t.
2. Selection of relevant market factors.
3. Risk mapping and VaR calculation.
4. Measurement of error and backtesting.

1.2.2. Returns

As the definition of VaR suggests, it is the portfolio return in the worst scenario. Therefore, the definition of the return of an asset is necessary in this occasion. The portfolio returns are the differences between portfolio values. That is, $\Delta P = P_{t+1} - P_t$, where P_t and P_{t+1} are the values of the portfolio at times t and $t+1$, respectively.

In general, the portfolio return is described by the rate of the return. Two different approaches to this rate can be done: arithmetic and geometric (see [7]).

The arithmetic rate at time t, $R_{a,(t)}$, can be described as the portfolio return divided by the original portfolio value:

$$R_{a,(t)} = \frac{P_t - P_{t-1}}{P_{t-1}}$$

The geometric return is defined as the logarithm of the price ratio, that is, the geometric return at time t, $R_{g,(t)}$, is of the form:

$$R_{g,(t)} = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

Notice that, $R_{g,(t)} = \ln\left(1 + R_{a,(t)}\right)$. If the time horizon of the working data is short (e.g. one day), the arithmetic rate of returns is near to 0. Therefore, by Taylor expansion:

$$R_{g,(t)} = \ln\left(1 + R_{a,(t)}\right) = R_{a,(t)} - \frac{R_{a,(t)}^2}{2} + \frac{R_{a,(t)}^3}{3} + \cdots \approx R_{a,(t)}$$

Thus, both geometric and arithmetic returns shall be treated the same way and the arithmetic return is used.

1.2.3. Expected Shortfall

The Expected Shortfall (ES) is known as tail Value at Risk or Conditional Value at Risk (CVaR). The ES is defined as the expected loss of financial position after an extreme event.

Using Definition 1.1, the Expected Shortfall at time t for a time horizon h can be expressed as

$$ES_{t,\alpha}(h) = \mathbb{E}\left(r_{t+h} | \Omega_t \text{ s.t } r_{t+h} \leq VaR_{t,\alpha}(h)\right)$$

From the definition above, it can be seen that ES is the expected loss of the conditional distribution of returns given that the return exceeds its VaR.

1.3. Advantages and limitations of VaR

As Andreas Krause well explain in [32]: "The VaR measure has an obvious benefit because it is easily and intuitively understood by non-specialists. Thus, it can therefore be well communicated within a company as well as between the company and regulators, investors, or other stakeholders. Furthermore, it can address all types of risks in a single framework, which not only allows the aggregation of risks but also further facilitates communication."

It is true that VaR facilitates its use as a basis for performance-related pay and the decisions about capital requirements and businesses for individual traders, divisions and companies. It also can help to decide which type of risk should be reduced as an strategy.

But the truth is that VaR has some weaknesses.

Definition 1.2. *[45, Section 1] A risk measure δ is said to be coherent if for any two loss random variables X and Y, δ satisfies:*

- *Subadditivity:* $\delta(X+Y) \leq \delta(X) + \delta(Y)$.
- *Monotonicity: If* $X \leq Y$ *a.s., then* $\delta(X) \leq \delta(Y)$.

- *Positive homogeneity: For* $c > 0$, $\delta(cX) = c\delta(X)$.
- *Translation invariance: For* $c > 0$, $\delta(c+X) = c+\delta(X)$.

It can be shown that VaR is coherent under normal distribution. However, it is not coherent in general. Let X and Y be two normal distributed random variable. Without loss of generality, the mean of the two variables are 0. Thus, $X \sim N(0,\sigma_1^2)$ and $Y \sim N(0,\sigma_2^2)$. The VaR at a given confidence level $1-\alpha$ and time horizon h at time t are $z_\alpha\sigma_1$ and $z_\alpha\sigma_2$, respectively. Because of normal distribution properties, the subadditivity property holds. It can be seen that,

$$\sigma_{X+Y}^2 = Var(X+Y) = Var(X)+Var(Y)+2Cov(X,Y) \leq (\sigma_1+\sigma_2)^2$$

Therefore, $\sigma_{X+Y} \leq \sigma_1+\sigma_2$ and this implies that VaR of the random variable $X+Y$ satisfies the subadditivity property. However, this property does not hold in general. Several counterexamples can be found in [29]. The fact that Value at Risk does not satisfy the sudadditivity property has important effects from a financial point of view. It implies that the Value at Risk does not favour portfolio diversification, that is, investing in different assets and securities instead of in the joint portfolio, which is the handbook for finance.

The Expected Shortfall is a coherent risk measure, even though it is derived from the Value at Risk, which is not.

The Value at Risk is calculated in the conventional way for an individual security assuming elliptical distributions such as normal, lognormal, etc. In the case of considering two or more securities, that is, a portfolio, the correlation between them is a linear relationship between the variables and the Value at Risk of the portfolio is the sum of VaR of the securities.

However, if non-elliptical distributions are considered, the Aggregate VaR for the portfolio can be greater than the sum of VaR of each individual security in the portfolio due to correlations.

Also, Value at Risk estimation is subject to large errors, the estimate is downward biased, and these shortcomings can be exploited by individuals within the company as well as the company as a whole. Plus, the cost effectiveness assumptions or the history of data may not be appropriate for the time series. Depending on the the time horizon, the estimation of VaR might be completely different and VaR can be underestimated (small window) or overestimated (large window). However, these disadvantages do not imply that VaR is not a useful tool in risk measurement.

VaR has proved to be a very useful and widely accepted measure, despite its weaknesses. But it is exactly these shortcomings that limit the extent to which VaR can be used. The VaR estimate should not be taken as a precise number, but it provides an indication as to how much risk is involved. It also aids in detecting any trends in the behaviour of individuals, divisions, or the company as a whole.

Even though that not a single measure can replace the knowledge and experience of managers and those using it should be aware of its limits, the benefits of the simplicity of VaR cannot be underestimated. This field is suffering constant changes and new methods and ways to model errors are being developed.

Currently research is under way to improve the estimation procedure for VaR so as to overcome some of the above-mentioned problems. Alternatives to VaR have been developed that employ a similar idea, but provide a better risk measure. One proposed measure is expected shortfall (ES). It thus includes more information on the distribution of losses than the VaR and can be shown to be a coherent risk measure.

However, an inconvenient of these measures is that they model risk of securities in a separated way. At present, due to the high interconnection level of markets, researchers are working on developing the so-called systemic risk measures. These measures try to model events which affect to several securities, producing a contagious effect.

In conclusion, the appropriate use of VaR with full awareness of its limitations can improve decision making in companies.

1.4. Approaches to estimation

From Definition 1.1, one can tell that there are two different methods to estimate the VaR (see [2]):

1. Inverting the distribution function of financial returns.
2. Inverting the distribution of innovations.

Thus, VaR with confidence level $1-\alpha$ at time $t+h$ with time horizon h can be estimated as:

$$VaR_{t,\alpha}(h) = F_t^{-1}(\alpha) = \mu_{t+h|t} + \sigma_{t+h|t} G_t^{-1}(\alpha) \tag{1.1}$$

where F_t is the conditional distribution function of financial returns, G_t is the conditional distribution of innovations, $\mu_{t+h|t} = E[r_{t+h}|\Omega_t]$ and $\sigma_{t+h|t} = Deviation[r_{t+h}|\Omega_t]$.

In the next chapter, different methods to estimate VaR are presented and it can be seen that the two approaches are used depending on the method.

Chapter 2

VaR classical estimation and testing methods

Many people have along the years tried to reach to a financial model which could be capable to gather different types of behaviour of financial series of assets' information. The importance of the measure and knowledge of the VaR has caused the development of several methods of estimation. Some of them are presented in this chapter.

Two classical models are firstly proposed: the historical model and the Variance-Covariance method, also known as the Delta method. These two models show simple approaches to VaR calculations.

The GARCH model is proposed in the third place. This model is developed in order to be able to model volatility of returns. A simple model is firstly shown, the ARCH model, and later the GARCH model and its extensions are explained. The MLE method is used in order to estimate the parameters of the model.

Finally, extreme value models are presented.The VaR method explains extreme events in data with a certain probability and a certain value. The Extreme Value approach, which estimates shocks of markets and crisis as extreme events, can be modelled as a Generalized Extreme Value distribution and by the Peaks over Threshold method.

Some validation methods are introduced in the last section of the chapter. These tests allow any trader to know the effectiveness of the VaR estimation and to find the model that best fits the returns of a series of asset values.

2.1. Historical method

The historical method is a widely used model to estimate VaR because of its simplicity. This method is based on the belief that the behaviour of returns of an asset in the future will follow the same distribution as returns in the past.

Unlike many parametric models such as GARCH models or the Delta method (fully explained in the following sections), this method does not have to make assumptions about the parametric form of the distribution of the asset returns and the VaR is directly inferred from the sample.

However, the main limitation of the model is the varying VaR due to choice of the sample size and the frequency of data. To avoid unstable VaR when the model is re-estimated day after day, a considerable amount of historical data is required.

According to C. Alexander (see [4]), the steps which shall be followed in order to implement the historical VaR are the following:

1. Obtain a sufficiently long period of historical data: the choice of the number of observations which will be used in the model can be complicated. On the one hand, if the window is too short, some particular events of the past might not be taken into account. An underestimation of the VaR is not at all desirable.

On the other hand, a large window might lead to an overestimation of the VaR and institutions and companies could be saving more money than the real needed quantity. This practise leads to losses at the companies because all the money they are saving is not being reinvested.

2. Adjust the observed portfolio returns to reflect the current market conditions: in order to obtain the VaR given a sample, the returns are given and disposed in ascending order.

3. Fit the empirical distribution of returns.

4. Derive the VaR for the relevant significance level and risk horizon: for a given significance level α (the confidence level is equal to $1-\alpha$), the VaR corresponds to the α-quantile of the sample returns.

2.2. Variance-Covariance model

The Variance-Covariance method, also known as the Delta-Normal method, consists in assuming that returns of an asset are i.i.d. and have a normal distribution.

This method is very simple as, in order estimate the VaR, the expected return and the standard deviation of the sample are the only required elements. The VaR of a sample of returns with confidence level $1-\alpha$ at time t for a time horizon h is given by:

$$VaR_{t,\alpha}(h) = \bar{r}_{t+h|t} + \Phi^{-1}(\alpha) \cdot \bar{\sigma}_{t+h|t}$$

where $\bar{r}_{t+h|t}$ and $\bar{\sigma}^2_{t+h|t}$ are the sample mean and variance for the information available at time t (calculated using the sample given by the chosen window length), respectively, and Φ is the standard normal distribution function.

Although this method seems useful, the reality is that in many occasions the assumption of normality is not satisfied by the sample. If the sample presents a high frequency, the distribution might be heavy-tailed or light-tailed. In this cases the hypothesis of normal distribution of data is rejected. Therefore, similar methods using the t-Student distribution for heavy-tailed distributions and the Generalized Exponential distribution (GED) have been developed as well.

Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. That is, data sets with high kurtosis tend to have heavy tails, or outliers. Data sets with low kurtosis tend to have light tails, or lack of outliers. A uniform distribution would be the extreme case.

In the case of the t-Student distribution, the degree of freedom parameter is related to sample kurtosis coefficient t as the kurtosis coefficient is calculated as the fourth moment of a distribution divided by the square of the second moment.

The kurtosis coefficient , k, shall be used to know the behaviour of the tails of the sample and indicates the degrees of freedom that shall be used in order to fit a given sample. Let v be the number of degrees of freedom of the t-student distribution. The kurtosis coefficient can be calculated as:

$$k = \frac{\mu^4}{\mu^2} = 3 + \frac{6}{v-4}$$

The normal distribution is often used as a benchmark. As $k = 3$ for a normal distribution, distribution such that $k > 3$ are called leptokurtic, that is, these distributions have fatter tails than Normal distribution.

2.3. GARCH Model

The usage of econometric models which are able to model volatility of the return of an asset is an important factor to improve approaches to value at risk models, understand the time-varying behaviour of the assets and improving the accuracy of forecasts.

Although AutoRegressive Moving Average (ARMA) time series models are very useful for modelling the trend and seasonality of series and their usage is widespread, they assume a constant volatility (see [44]). ARMA models are used to model the conditional expectation of an observation by using the previous observations. They do this by modelling the observations as a linear function of the trend, the seasonal component and a white noise term. Therefore, the conditional variance of the observations is constant.

In order to solve this issue, the AutoRegressive Conditional Heterocedasticity (ARCH) and, in a more general way, the Generalized AutoRegressive Conditional Heterocedasticity (GARCH), models are presented in this section.

2.3.1. Set-up

Let r_t be the return of a financial asset at time t. The variance of r_t is constant if it satisfies that $Var(r_t|X_{1,t},\dots,X_{p,t}) = \sigma^2$ for some variables $X_{1,t},\dots,X_{p,t}$. The general form of the regression of r_t using past returns $r_{t-1},\dots,r_{t-p}$ is given by

$$r_t = f(r_{t-1},\dots,r_{t-p}) + \varepsilon_t \tag{2.1}$$

where $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \sigma^2$. In general, the function f is linear.

Notice that the model can be modified so that it allows a non-constant conditional variance. In this case, Equation (2.1) can be written as

$$r_t = f(r_{t-1},\dots,r_{t-p}) + \sigma(r_{t-1},\dots,r_{t-p})\,\varepsilon_t$$

These models are called "variance function models".

2.3.2. ARCH(q) Models

The AutoRegressive Conditional Heterocedasticity model is a special case of variance function models (see [42, Section 12.3]). Proposed by R. F. Engle in 1982 (see [19]), these model are mean 0, serially uncorrelated processes with nonconstant variances conditional to the past, but present constant unconditional variances.

Definition 2.1. *A process $\varepsilon_1, \varepsilon_2, \dots$ is called a Gaussian white noise with unit variance if it satisfies that the process is independent $N(0,1)$ and therefore,*

$$E(\varepsilon_t|\varepsilon_{t-1},\dots) = 0$$

and

$$Var(\varepsilon_t|\varepsilon_{t-1},\dots) = 1$$

Using this definition, the ARCH(q) model can be defined as follows.

Definition 2.2 (ARCH(q) model)**.** *Let ε_t be a Gaussian white noise such that its variance is equal to 1. Then, a_t is an ARCH(q) process if*

$$a_t = \varepsilon_t \sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2}$$

The process can be also written as $a_t = \varepsilon_t \sigma_t$, where σ_t is the conditional standard deviation of a_t given past values.

It can be seen that the basic idea of the ARCH(q) model is that the mean-corrected asset return is uncorrelated, but dependent, and that dependence is described by a quadratic function of its lagged values.

To understand the ARCH models, the ARCH(1) model is studied in detail. Thus, the ARCH(1) model is given by

$$a_t = \sigma_t \varepsilon_t \ , \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2, \tag{2.2}$$

where $\alpha_0 \geq 0$ and $\alpha_1 \geq 0$. The conditional mean of a_t is clearly equal to 0 since ε_t is an independent process. Knowing that $E(\varepsilon_t^2) = Var(\varepsilon_t) = 1$, and using (2.2), the conditional variance of a_t is of the form,

$$Var\left(a_t|a_{t-1},\dots\right) = E\left\{\left(\alpha_0+\alpha_1 a_{t-1}^2\right)\varepsilon_t^2|a_{t-1},\dots\right\} = \left(\alpha_0+\alpha_1 a_{t-1}^2\right)E\left\{\varepsilon_t^2|a_{t-1},\dots\right\} = \alpha_0+\alpha_1 a_{t-1}^2$$

Notice that ARCH (q) processes are uncorrelated and have constant mean and a constant unconditional variance, but their conditional variance is non constant.

ARCH(q) postulates that the conditional variance is a linear function of the past q squared innovations,

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 = a_0 + \alpha(L) a_{t-1}^2$$

Let $v_t = a_t^2 - \sigma_t^2$, the ARCH(q) model can be written as

$$a_t^2 = a_0 + \alpha(L) a_{t-1}^2 + v_t$$

Since $\mathbb{E}_{t-1}(v_t) = 0$, the model corresponds directly to an AR(q) model for the squared innovations.

2.3.3. Weaknesses of ARCH models

Although ARCH models present numerous advantages, they also have some weaknesses:

- The model assumes positive and negative stocks to have the same effects on volatility, as it depends on the square of previous values so the sign is not taken into account. This is clearly not true in financial series, as the response of assets is different depending on the sign of the shock.
- ARCH coefficients' values are quite restrictive. For example, the conditional variance of the ARCH(1) model is finite if $\alpha_1 \in [0,1]$.
- In general, ARCH models respond slowly to sudden shocks to the return of a series, so they tend to overpredict the volatility of series.

2.3.4. General ARCH(p,q) Model

In 1986, T. Bollerslev (see [9]) introduced the Generalized Autoregressive Conditional Heterocedastic Model, which is based on the work of Engle in ARCH models, with a more flexible lag structure.

Definition 2.3. *Generalized Autoregressive Conditional Heterocedastic Model with parameters p and q, GARCH (p,q), can be represented as*

$$a_t = \varepsilon_t \sigma_t$$

where ε_t is the white noise term and

$$\sigma_t = \sqrt{\alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2}$$

defines the conditional variance.

GARCH models include ARCH models as a special case. It has been tested that GARCH processes present heavy tails. This means that they tend to have more extreme values than expected from a normal distribution. Therefore, this type of processes is useful both to model conditional heterocedasticity and heavy-tailed series.

Notice that the GARCH(p,q) is defined as

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 = \omega + \alpha(L) a_{t-1}^2 + \beta(L)\omega_{t-1}^2$$

Let $v_t = a_t^2 - \sigma_t^2$. The model can by rewritten as

$$a_t^2 = \omega + [\alpha(L) + \beta(L)] a_{t-1}^2 - \beta(L) v_{t-1} + v_t$$

which defines an ARMA[max(p,q),p] model for the squared innovations, a_t^2.

2.3.5. Parameter estimation

In order to obtain estimations of parameters of a general ARCH model, the maximum likelihood estimation method is applied.

Let $\{x_1, \ldots, x_n\}$ be a sample of n independent and identically distributed observations, with distribution function $f(x)$. Let θ be an unknown vector of parameters. The likelihood distribution function θ given $\{x_1, \ldots, x_n\}$ is of the form

$$L(\theta | x_1, \ldots, x_n) = f(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta)$$

The log-likelihood function is defined as $l(\theta) = \log(L(\theta))$ From the maximum likelihood problem, it is known that

$$\hat{\theta} = argmax_{\theta \in \Theta} L(\theta)$$

Using Newton's optimization method, it is easy to obtain approximated values of θ at iteration k, θ_k, by applying the following equation:

$$\theta_{k+1} = \theta_k - J^{-1}(\theta_k) \nabla L(\theta_k)$$

where L is the Fisher Information matrix and

$$\nabla L = \frac{\partial L}{\partial \theta}$$

Let a_t be a GARCH model as in Definition 2.3, with normal distribution with mean 0 and conditional variance equal to σ_t. The log-likelihood function of parameter vector $\theta = (\alpha_0, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)^T$ is of the form (see [48, Section 2.3]),

$$l(\theta) = \sum_{t=q+1}^{n} \left(-\frac{ln(2\pi)}{2} - \frac{ln(\sigma_t^2)}{2} - \frac{a_t^2}{2\sigma_t^2} \right)$$

2.3.6. The GARCH Model in Finance

Some financial models such as CAPM or the Black-Scholes model make the assumption that the conditional variance is constant. It has been proven that this kind of hypothesis usually lead to serious errors.

For this reason, the use of ARCH and GARCH processes in modelling errors of series has become a widespread research topic (see [10]), because while returns are usually uncorrelated, absolute returns are not. That is, large changes in time series tend to be followed by large change, and small changes tend to be followed by small changes. This phenomenon is denoted by volatility clustering.

Therefore, ARCH and GARCH models aim to describe this type of behaviour, as their main concept is that volatility is dependent upon the past. Plus, GARCH models are also useful to model series with heavy tails and an extreme value approach to the model can be done.

In [20], an example of risk measurement that could be the input to a variety of economic decisions is presented. The analysis of ARCH and GARCH models and their many extensions provides a statistical stage on which many theories of asset pricing and portfolio analysis can be exhibited and tested.

2.3.7. VaR estimation

In order to calculate the VaR using a GARCH model, the evolution of returns and their volatility can be presented as the following system:

$$\begin{aligned} r_t &= \mu + \varepsilon_t \\ \varepsilon_t|\Omega_{t-1} &\sim D(0,\sigma_t^2) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \end{aligned}$$

where D is the chosen distribution with mean 0 and variance σ_t^2. Normally, D is the Normal or the t-Student distribution.

Notice that the process is covariance stationary, that is, has unconditional finite variance, if $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$. In this case, the unconditional variance is

$$\sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}$$

Therefore, the estimation of Value at Risk at time t for a given time horizon h at a confidence level $1-\alpha$ is given by

$$VaR_{t,\alpha}(h) = \mu_{t+h|t} + D^{-1}(\alpha) \cdot \sigma_{t+h|t}$$

where $\mu_{t+h|t}$ and $\sigma_{t+h|t}$ are the mean and deviation of the sample within the information at time t and D is the distribution function of errors. It shall be notice that, in GARCH models, the mean is considered to be constant.

2.3.8. Extensions and modifications of GARCH

The are several modifications of the GARCH model that are commonly applied in finance (see [2] for more examples). Essentially these models change the way to calculate the volatility, that is, σ_t. The most widely used models are shown in this section.

IGARCH

Integrated GARCH models are useful to model series that present persistent changes in volatility. A GARCH(p,q) model is called an IGARCH model if it satisfies

$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i = 1$$

The conditional variance properties of the IGARCH model are not very attractive from the empirical point of view due to the very slow phasing out of the shock impact upon the conditional variance.

The Exponentially Weighted Moving Average (EWMA) is a special case of the IGARCH family. This model was firstly used by RiskMetrics, when this company introduced the concept of VaR as a risk measure. It shall be noticed that the EWMA is a GARCH(1,1) model in which $\alpha_1 + \beta_1 = 1$ and α_0 is equal to 0.

Thus, the model calculates the return variance using the following recursive formula:

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

GJR-GARCH

Despite of being uncorrelated, the errors of the series ε_t do not need to be independent necessarily. Thus, the Glosten-Jagannathan-Runkle (GJR) GARCH model, proposed by Glosten, Jagannathan and Runkle in 1993 (see [23]) assume a specific parametric form of this conditional heterocedasticity.

This model is considered as it is capable of explain the leverage effect of a series, that is, the asymmetry of the errors of the series of returns. The leverage effect is caused by the fact that the negative returns have a different influence on future values of the series than positive returns, that is, the influence of positive and negative returns is not symmetric.

For any time t, the return variance is calculated as follows:

$$\sigma_t^2 = \alpha_0 + (\alpha_1 + \gamma\, I_{t-1})\, \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

where I_{t-1} is the indicator function:

$$I_{t-1} = \begin{cases} 0 & if \quad r_{t-1} \geq \mu \\ 1 & if \quad r_{t-1} < \mu \end{cases}$$

E-GARCH

The Exponential-GARCH process, proposed by D. Nelson in 1991 (see [36]), is used to model the leverage effect.The leverage effect explains an investor's return in terms of its return on capital employed and debt. It occurs when an asset's return become more volatile as its price decreases.

The E-GARCH (p,q) models the return variance, σ_t, as follows:

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} \left(\alpha_i \left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| + \lambda_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right) + \sum_{j=1}^{p} \beta_j \ln(\sigma_{t-j}^2)$$

In contrast to the GARCH model, the E-GARCH model does not need any restrictions due to the logarithmic transformation.

The parameters λ_i allow for the asymmetric effect. If $\lambda_1 = 0$ then a positive surprise, $\varepsilon_t > 0$, has the same effect on volatility as a negative surprise, $\varepsilon_t < 0$. The presence of leverage effect can be investigated by testing the hypothesis that $\lambda_1 < 0$ (see [5]).

APARCH

The APARCH model, proposed by Ding, Granger and Engle in 1993 (see [18]), can be seen as a generalization of the GARCH model. GARCH models try to model the variance of returns and gives no special treatment to positive or negative shocks. As in the E-GARCH models, the leverage effect, that is, the asymmetric effect of negative shocks in the market should have an impact in the estimation of $sigma_t$.

Thus, the Asymmetric Power ARCH model (APARCH) is presented. This method tries to model the δ-power of the standard deviation of returns of an asset. Therefore, the δ-power of σ_t is modelled by the following expression:

$$\sigma_t^{\delta} = \alpha_0 + \sum_{i=1}^{q} \alpha_i \left(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i} \right)^{\delta} + \sum_{i=1}^{p} \beta_j \sigma_{t-j}^{\delta}$$

where γ determines the sign of the asymmetric effect.

Note 2.4. *It should be notice that these models estimate the variance of returns. Therefore, in order to find the corresponding VaR, using Equation* (1.1)*, the VaR with confidence level* $1-\alpha$ *and time horizon* h *at time* t *is estimated as:*

$$VaR_{t,\alpha}(h) = \mu + D^{-1}(\alpha) \cdot \sigma_{t+h|t}$$

where μ is the mean of the distribution, $\sigma_{t+h|t}$ is estimated using the previous methods with information until time t and $D^{-1}(\alpha)$ is the α-quantile of the distribution (usually normal or t-student distribution).

In addition, as $\mu \sim 0$, it can be assumed that the return of the asset at time t is approximately equivalent to ε_t.

2.4. ARMA-GARCH Models

Let $\{r_t\}_t$ be a time series of asset returns. Generally, this type of series can be decomposed as follows:

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t \\ \varepsilon_t|\Omega_{t-1} &\sim D(0, \sigma_t^2) \end{aligned}$$

where $\mu_t = \mathbb{E}[r_t|\Omega_{t-1}]$ and $\varepsilon_t^2 = Variance[r_t|\Omega_{t-1}]$.

It has been seen that GARCH methods can be useful in order to model the heterocedasticity of the series and to study the relationship between the variability of returns and their lags.

The class of AutoRegressive-Moving-Average (ARMA) models is the most broadly utilized method for modelling the mean of the series of returns.

Definition 2.5 ([47]). *A process $\{r_t\}_t$ is called an ARMA(p,q) process if there exist real coefficients $c, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ where P and Q are integers, such that*

$$r_t - \sum_{i=1}^{P} \phi_i r_{t-i} = c + \varepsilon_t + \sum_{j=1}^{Q} \theta_j \varepsilon_{t-j} \quad \forall t \in \mathbb{Z}$$

where ε_t is the white noise $(0, \sigma^2)$

2.4.1. ARMA(p,q)-GARCH(r,s) model

The ARMA-GARCH model is a combination of both ARMA and GARCH model in which a linear ARMAmodel is used for modelling the mean behaviour and a GARCH model is used for modelling the variance of the residuals from the ARMA model.

Therefore, the general form of an ARMA(q,r)-GARCH(r,s) model is:

$$\begin{aligned} &r_t - \textstyle\sum_{i=1}^{P} \phi_i r_{t-i} = c + \varepsilon_t + \sum_{j=1}^{Q} \theta_j \varepsilon_{t-j} \\ &\varepsilon_t|\Omega_{t-1} \sim D(0, \sigma_t^2) \\ &\sigma_t^2 = \alpha_0 + \textstyle\sum_{i=1}^{s} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{r} \beta_j \sigma_{t-j}^2 \end{aligned}$$

Under this model, both the conditional mean and variance variance of r_t depend on the previous time steps.

In practice, in order to choose the right ARMA orders, the Autocorrelation function (ACF) for the AR order and the Partial Autocorrelation function (PACF) for the MA order should me analysed. They show the significant correlations between lags in the series. Thus, P and Q should be chosen according to the significant values of both functions.

In order to fit a series of returns, ARMA-GARCH models offer a wide range of options because, as it has been seen in previous sections, there exist several types of GARCH models to fit the heterocedasticity of the series and a good fit of ARMA models can ensure the stationarity of residuals.

2.4.2. Estimation of VaR

As it has been seen in Note 2.4, the Value at Risk depends on the estimation of the mean and the variance (or standard deviation) of the series of returns and on the corresponding distribution of the residuals.

In the case of the ARMA-GARCH models, both the mean and the variance of the series vary with time. Therefore, for a given time t , a time horizon h and confidence level $1-\alpha$, the Value at Risk is given by:

$$VaR_{t,\alpha}(h) = \mu_{t+h|t} + D^{-1}(\alpha) \cdot \sigma_{t+h|t}$$

where $\mu_{t+h|t}$ and $\sigma_{t+h|t}$ are estimated by the ARMA-GARCH model given in Section (2.4.1) with information until time t and D is the distribution function of the residuals of the series.

2.5. Extreme value models

In this section, a non-parametric method is presented: the approach based on Extreme Value theory.

Because of the definition of VaR, it is important to understand the behaviour of extreme events and tail distributions. Assuming that losses are serially independent with common CDF, F, an Extreme Value Theory (EVT) application can be made. A review on EVT is presented (see [38]) and the application to VaR estimation is discussed in this section using two methods: the first of which is based on the extreme value distributions of the Gumbel, Fréchet or Weibull distributions (GEV), while the second is based on the generalized Pareto distribution (GPD) and the Peaks Over Threshold (POT) approach.

2.5.1. Block Maxima Method Characterization

Let $r_1, \cdots, r_n$ be iid finantial returns and let

$$R_{n,n} = \max(r_1, \cdots, r_n) \tag{2.3}$$

Let F be the distribution function of $r_1, \cdots, r_n$. The distribution function of $R_{n,n}$ can be calculated exactly for all values of n using the independence of the random variables as follows:

$$\wp(R_{n,n} \leq x) = \wp(r_1 \leq x, \cdots, r_n \leq x) = \wp(r_1 \leq x) \times \cdots \times \wp(r_n \leq x) = (F(x))^n$$

Although this may seem useful, there is a problem when calculating this if the distribution function F is unknown.

When F is estimated, it is clear that there can be little discrepancies due to the lack of data. In this case, some calculations are needed to obtain an estimation of F^n, but as n increases to infinity, $R_{n,n}$ becomes degenerated to a point mass.

Extreme theory seeks norming constants $(\sigma_n > 0)$ and (μ_n) and a nondegenerate G such that

$$\wp\left(\frac{R_{n,n} - \mu_n}{\sigma_n} \leq x\right) \longrightarrow G \tag{2.4}$$

where G is an extreme value distribution function and F is said to be in the domain of attraction of G.

This allows to study the behaviour of the normalized function, G, instead of $R_{n,n}$, which will be useful in order to explain the general extreme value theory.

2.5.2. GEV distribution

The following theorem contains the most important result about extreme value distribution functions (see [25]). Fisher and Tippett started the research in 1928, and later, Gnedenko formalized it in 1948. It formalizes the limiting distribution of Equation (2.4).

Theorem 2.6 (Fisher - Tippett - Gnedenko theorem)**.** *[16, Section 3.1.2] If there exist sequences of constants* $(\sigma_n > 0)$ *and* (μ_n) *such that*

$$\lim_{n\to\infty} \wp\left(\frac{R_{n,n}-\mu_n}{\sigma_n} \leq x\right) = G(x)$$

where G is a non-degenerate distribution function, then G belongs to one of the following families

I: $G(x) = \exp\left(-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right),\ -\infty < x < \infty$ *(Gumbel distribution)*

II: $G(x) = \begin{cases} \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right) & x > -1/\xi \\ 0 & otherwise \end{cases}$ *(Fréchet distribution)*

III: $G(x) = \begin{cases} \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right) & x < -1/\xi \\ 1 & otherwise \end{cases}$ *(Weibull distribution)*

for a scale parameter $\sigma > 0$*, a location parameter* μ *and a shape parameter* ξ*.*

This theorem shows that the only possible limiting distribution for G given $R_{n,n}$ and sequences $(\sigma_n > 0)$ and (μ_n), is one of these three types. In some way, this theorem gives an extreme value analogue of the central limit theorem, as G is the normalized distribution of $R_{n,n}$.

Notice that the three distribution can be written as a single Generalized Extreme Value Distribution where the Fréchet family corresponds to $\xi > 0$, the Gumbel family with $\xi = 0$ and the Weibull family with $\xi < 0$.

Therefore, the algebraic expression for the generalized distribution can be expressed as follows:

$$G_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\left(-1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi} & \xi \neq 0 \ \text{ and } \ (1+\xi(x-\mu)/\sigma) > 0 \\ \exp\left(-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right) & \xi = 0 \end{cases}$$

Given n observations, the Value at Risk is calculated depending on the value of ξ as follows:

$$\widehat{VaR}_\alpha = \begin{cases} \mu_n - \frac{\widehat{\sigma}_n}{\widehat{\xi}_n}\left(1-(-n\ln(\alpha))^{-\widehat{\xi}_n}\right) & if\ \widehat{\xi}_n \neq 0 \\ \widehat{\mu}_n - \widehat{\sigma}_n \ln(-n\ln(\alpha)) & if\ \widehat{\xi}_n = 0 \end{cases}$$

and the estimation of the parameters of the GEV models is obtained using maximum likelihood estimation methods (see [38, Section 1.2.]).

In most situations, in order to obtain the VaR at certain time t for a given sample of historical returns, the length of the window of data of returns with which the VaR is calculated is selected in such a way that the length matches a year interval and n is the number of observations within that year period (width of the window).

Note 2.7. *It shall be noticed that the generalized extreme value model for Block Maxima (or minima) can be approached in two ways: firstly, given the return level estimation, the return period can be calculated, while in the second, the return level can be estimated for a given return period.*

If a daily return period is chosen, the sample used to fit the GEV distribution's values are the maximum daily observation of the considered data frame. Other periods such as weeks, half months or months can be considered. If weekly periods are chosen, each observation of the Block Maxima is the maximum return of each week (5 working days). That is, given a length of block of returns, b, each observation of the block maxima series is the maximum value of the returns within b days of observations.

The return level for a given distribution $G_{\xi,\mu,\sigma}$ of Block Maxima is defined as:

$$R_n^k = G_{\xi,\mu,\sigma}^{-1}\left(1-\frac{1}{k}\right)$$

The return level R_n^k is explained as the value expected to be exceeded in one out of k periods of length n. Therefore, this is a conservative measure of Value at Risk.

In practice, in order to obtain the Value at Risk for high confidence levels, the Block Maxima of the minus returns is considered.

2.5.3. Peaks over thresholds

The method of peaks over threshold (POT) is based on the fact that excesses over a fixed value u have a generalized Pareto distribution and the occurrence of these excesses is a Poisson process. As a matter of fact, the GEV distribution for maxima can be obtained in terms of Poisson processes.

Let $r_1, r_2, \cdots$ be a sequence of independent returns with common marginal distribution function, F. It can be intuitive to say that r_i is an extreme event if it exceeds some fixed threshold, u (e.g. shock of market or crisis). A theoretical description for this behaviour can be given by the following conditional probability.

Definition 2.8 (Excess distribution function)**.** *Let X be a random variable with distribution function F. For a fixed u,*

$$F_u(x) = \wp\{X-u \leq x | X > u\},\ x \geq 0 \tag{2.5}$$

is the excess distribution function of the random variable X over the threshold u.

Using this definition we have the following equality,

$$\wp\{X > u+x | X > u\} = \frac{1-F(u+x)}{1-F(u)} = 1-F_u(x) \tag{2.6}$$

If the distribution function F was known there would be no problem in calculating that probability and obtaining a formula for the distribution. As this is usually not the case, under the same conditions where the GEV distribution function can be used as an approximation to the distribution function for maxima of long sequences, an explicit expression of (2.5) can be obtained by substituting the distribution given in Theorem 2.6 for the distribution function for maxima, $F_{n,n}$.

The main Pareto distribution result is given in the following theorem.

Theorem 2.9 (Pickands (1975) [39], Balkema and de Haan (1974) [6])**.** *Let $r_1, r_2, \cdots$ be a sequence of returns with common distribution function F, and let*

$$R_{n,n} = max\{r_1, \cdots, r_n\}.$$

Denote an arbitrary element in the sequence as R, and suppose that F satisfies Theorem 2.6. Then for large enough u, the distribution function of $(R-u)$, conditional on $R > u$, is approximately

$$H_{\xi,\sigma}(x) = 1-\left(1+\frac{\xi x}{\sigma_u}\right)^{-1/\xi} \tag{2.7}$$

defined on $\{x : x > 0,\ (1+\xi x/\sigma_u > 0)\}$ where $\sigma_u = \sigma + \xi(u-\mu)$.

This distribution is called the generalized Pareto distribution (GPD) where σ_u is the scale parameter and ξ is the shape parameter.

Assuming that the distribution of excesses losses over the threshold u is a Generalised Pareto Distribution, let n be the total number of observations and N_u the number of excesses over the threshold u. Using (2.6), the distribution of the returns can by given as follows:

$$F(x) = [1 - F(u)]\, F_u(x) + F(u) \tag{2.8}$$

In order to construct a tail estimator from (2.8), the estimation of $F(u)$ must be given. Thus, using the empirical approximation to the estimation of $F(u)$, $(u - N_u)/u$, the tail estimation is of the form:

$$F(x) = 1 - \frac{N_u}{n}\left[1 + \frac{\widehat{\xi}}{\widehat{\sigma_u}}(x-u)\right]^{-1/\widehat{\xi}} \tag{2.9}$$

Given a probability α, the VaR with probability α (confidence level $1-\alpha$) can be obtained by inverting the tail estimator given in Equation (2.9) as follows:

$$\widehat{VaR}(\alpha) = u + \frac{\widehat{\sigma_u}}{\widehat{\xi}}\left[\left[\frac{n}{N_u}(1-\alpha)\right]^{-\widehat{\xi}} - 1\right] \tag{2.10}$$

In a similar way, Expected shortfall is fitted for $\xi < 1$ using Equation (2.10) , obtaining (see [43]):

$$\widehat{ES}_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 \widehat{VaR}(x)\, dx = \frac{\widehat{VaR}(\alpha)}{1-\widehat{\xi}} + \frac{\widehat{\sigma}_u - \widehat{\xi}u}{1-\widehat{\xi}}$$

It shall be noticed that the choice of the threshold is essential in order to obtain the Value at Risk for a given sample. In practice, for a given set of financial returns, the threshold is selected as the p-quantile of the sample, where $1-p$ is the proportion of sample which the investor decides should not be considered as extreme.

There are other techniques in order to choose the adequate threshold such as the analysis of the linearity of the mean excess function. If the selected threshold is too high, there is a lack of data and the fitted Pareto distribution is not adequate for the given sample. On the other hand, if the threshold is too low, some excesses might not be real extreme values.

As for GEV distribution, in practice, the GP distribution is usually fitted to the minus series of returns for high confidence levels. The Value at Risk in these cases is the -VaR obtained from the obtained Pareto distribution.

In POT method, all the parameters of the excess distribution are fitted using Maximum Likelihood Estimation and confidence intervals are calculated using profile likelihood (see [22] for more details). Once the parameters are fitted, VaR can be calculated.

These methods are known as Unconditional Extreme Value Theory methods. Given the conditional heteroscedasticity characteristic of most financial data, a new methodology to estimate the VaR can be proposed. This method combines the Extreme Value Theory with volatility models and it is known as the Conditional Extreme Value Theory.

The CEVT proposes GARCH models to estimate the current volatility and Extreme Value Theory to estimate the distributions tails of the GARCH model shocks. Therefore, the VaR can be estimated using Equation (1.1) knowing that, in this case, the errors of the series follows a Pareto distribution. Therefore, VaR at time t with confidence level $1-\alpha$ at a given time horizon t is given by:

$$VaR_{t,\alpha}(h) = \mu_{t+h|t} + \sigma_{t+h|t} \cdot H^{-1}_{\xi,\sigma}(\alpha)$$

where $\sigma^2_{t+h|t}$ is the variance of returns estimated by GARCH models and $H^{-1}_{\xi,\sigma}(\alpha)$ is the α-quantile of the Generalized Pareto Distribution calculated as in Equation (2.10).

2.6. Validation and testing

For a given sample of returns, $r_1, \ldots, r_n$, there exist different methodologies capable of evaluate the accuracy of the methods exposed in this chapter.

In order to check whether Value at Risk forecasts fulfil regulatory requirements (specified in the Basel Accords) and are adequate and effective, this section presents a wide range of methods and tests to select the best model for the given sample of returns (see [11]).

2.6.1. Performance tests

Several procedures based on statistical hypothesis testing have been proposed, and researchers usually select one or more tests to evaluate the accuracy of VaR models in order to compare them and select the best model for every return series VaR forecast values. Thus, the standard tests about the accuracy VaR models are presented down below.

In order to implement these tests, given a sample of returns $r_1, \ldots, r_n$ and a confidence level $1-\alpha$, an indicator function must be defined as follows:

$$I_t = \begin{cases} 1 & if \quad r_t < VaR(\alpha) \\ 0 & if \quad r_t \geq VaR(\alpha) \end{cases} \tag{2.11}$$

Thus, it shall be noticed that $I_t = 1$ when the Value at Risk at time t with a certain probability α is greater than the return at time t, that is, when the VaR prediction fails and the loss at time t is greater than the estimated maximum loss with a given probability α.

Excess Ratio

The Excess Ratio (ER) can be described as:

$$ER = \frac{\sum_t I_t}{n}$$

where n is the number of VaR forecasts and I_t is the indicator function described in Equation (2.11). One can tell that if $I_t = 1$ for a certain value t, an exception is given. This means that the modelled $VaR_{t,\alpha}$ has been exceeded by the return of the asset at time t. Thus, the Excess Ratio represents the percentage of model failure. In theory, the ER should be equal to the significance level α.

Traffic light test

The Basel Committee has created the traffic light test . This test is created based on the Excess Ratio value. If the probability of an exception is assumed to be constant, the number of exceptions is expressed as a random variable $X = \sum_t I_t$. This variable X follows a binomial distribution $B(n,\alpha)$, where n is the number of observations (see [1]).

In this test, the measure of VaR forecast quality is made based on the attribution of lights assigned on the basis exceeding the next threshold of ER. Thus, depending on the performance of the model, a color (level) describes the model and measures whether the quality of the model is adequate, the model needs supervision or the quality of the model is not good enough to be considered. The three lights are:

- Green: No problems with the forecast quality. The model considered is valid.
- Yellow: A supervision of the model shall be considered. The yellow zone begins at the point where the cumulative binomial distribution is greater or equal to 0.95.

  Let $B(n,\alpha)$ be the cumulative binomial distribution where n is the sample size and $1-\alpha$ is the VaR confidence level. The value, k_Y, at which the border between the green and yellow zone of the Basel traffic light test begins is given by

  $$k_Y = \sup_x \wp(B(n,\alpha) \leq x) < 0.95$$

- Red: The model generated the VaR forecasts of bad quality. The red zone begins at the point, k_R, where the cumulative binomial distribution is greater or equal to 0.9999, that is,

  $$k_R = \sup_x \wp(B(n,\alpha) \leq x) < 0.9999$$

Unconditional coverage

It is easy to see that an accurate VaR measure with probability α should produce an unconditional coverage, that is,

$$\widehat{\alpha} = \frac{\sum_t I_t}{n}$$

should be equal to α.

The Basel light test does not take into account the overestimation of the model, that is, the expected proportion of excesses corresponds to the α level of the VaR but the traffic light test assumes that a forecast with a lower proportion of exceedances is a good model.

Therefore Kupiec (see [33]) developed the unconditional coverage test, in which the deviation from both sides of the assumed number of excesses is taking into account.

This tests measures the significant departure of $\widehat{\alpha}$ from α. Kupiec showed that if the probability of an exception is constant, then the number of exceptions follows a binomial distribution, $B(n,\alpha)$. The test for the significance of the departure of $\widehat{\alpha}$ from α is carried out using the z-statistic which follows an asymptotic normal distribution:

$$Z = \frac{n\widehat{\alpha} - n\alpha}{\sqrt{n\alpha(1-\alpha)}}$$

It asymptotically follows a $N(0,1)$ distribution.

The test evaluates VaR forecasts by checking how a VaR forecast model performs over a period. It prevents underestimation of VaR and not only ensures that a financial institution carries significantly high capital, but it also reduces overestimation of VaR, which could lead to excessive conservatism.

The unconditional coverage test is justified with a null hypothesis, $H_0 : \widehat{\alpha} = \alpha$, and a likelihood ratio statistic satisfying:

$$LR_{uc} = 2\left[\log(\widehat{\alpha}^x(1-\widehat{\alpha})^{N-x}) - \log(\alpha^x(1-\alpha)^{N-x})\right] \sim \chi^2(1) \tag{2.12}$$

It shall be noticed that this test assumes that the exceptions are independent, which could be a false statement in some cases.

Conditional coverage

Christoffersen and Pelletier (see [15]) proposed a new test in order to solve the posed problem that unconditional coverage does not cover, that is, the assumption of independent exceptions. This new test is the conditional coverage test.

Thus, the null hypothesis of this test is the hypothesis of the unconditional coverage test ($H_0 : \widehat{\alpha} = \alpha$) and in addition, the independence of excesses of VaR forecasts.

Let N_{ij} the number of observations at state i that change to state j, where $i, j \in \{0,1\}$, where 0 means not an exceedance and 1 means exceedance.

Therefore, probabilities of changes can be defined as

$$\pi_{01} = \frac{N_{01}}{N_{00}+N_{01}} \; ; \; \pi_{11} = \frac{N_{11}}{N_{10}+N_{11}}$$

where π_{01} is the probability of exceedance provided the lack of exceedance in the previous period and π_{11} is the probability of exceedance provided the exceedance in the previous period.

It shall be noticed that N_{ij} can be expressed in terms of I_t as follows:

$$N_{00} = \sum_{t=2}^{n}[I_t = 0 | I_{t-1} = 0] \; ; \; N_{01} = \sum_{t=2}^{n}[I_t = 1 | I_{t-1} = 0]$$

$$N_{10} = \sum_{t=2}^{n}[I_t = 0 | I_{t-1} = 1] \; ; \; N_{11} = \sum_{t=2}^{n}[I_t = 1 | I_{t-1} = 1]$$

The likelihood ratio statistic corresponds to:

$$LR_{cc} = LR_{uc} + LR_{ind}$$

and it is asymptotically distributed $\chi^2(2)$.

The LR_{uc} statistic is the same one as in Equation (2.12). The LR_{ind} is the likelihood ratio statistic for the hypothesis of serial independence against first order Markov dependence. The idea behind the alternative hypothesis is that clustered exceptions is a signal of risk model misspecification.

The LR_{ind} statistic is of the form

$$LR_{ind} = 2\left(\log L_A - \log L_0\right) \sim \chi^2(1)$$

where

$$L_A = \left(1-\pi_{01}\right)^{N_{00}} \pi_{01}^{N_{01}} \left(1-\pi_{11}\right)^{N_{10}} \pi_{11}^{N_{11}}$$

and

$$L_0 = \left(1-\pi\right)^{N_{00}+N_{01}} \pi^{N_{01}+N_{11}}$$

Thus, the conditional coverage tests the null hypothesis about the independence of excesses, against the alternative hypothesis the the excesses are characterised by the first order Markov chain.

Notice that the LR_{cc} test only takes into account the first order autocorrelation of the sequence of exceptions.

2.6.2. Evaluation via cost functions

In the Basel committees, supervisors shows their concern not only to the number of exceptions exceeding an estimated VaR, but to the magnitude of these excesses.

Therefore, the loss functions examine a magnitude between the observed returns and the forecasted $VaR_{t,\alpha}$ for each time t. The basic loss function (BLF), lf, is of the form

$$lf_t = \begin{cases} \left(r_t - VaR_{t,\alpha}\right)^2 & if \quad r_t < VaR_{t,\alpha} \\ 0 & if \quad r_t \geq VaR_{t,\alpha} \end{cases}$$

By using this loss function, the VaR measure is penalized with the square distance. Therefore, the best model will be the one that has lower mean loss value, which is the mean of the penalty scores ($\sum_{t=1}^{n} lf_t/n$).

The BF focuses on the penalty of excesses. A cost function is called a firm function if the cost of no exceedances is taken into consideration in order to punish too high capital protection (contingency plans). A cost fuction that is often used in these cases is the Caporin firm's cost function. For each time t, the function is described as:

$$FC_t = \left|r_t - VaR_{t,\alpha}\right| \ , \ \forall r_t$$

The result of the function given a sample of n observations and their fitted VaR estimations is the average Caporin cost function, $FC = \sum_{i=1}^{n} FC_i/n$.

Another statistical loss function is the one given by González-Rivera et al. (see [24]). For a given probability value α, the loss function Q is the loss function used in quantile estimation and is given by:

$$Q = \frac{1}{n}\sum^{t}\left(\alpha - d_t^{\alpha}\right)\left(r_t - VaR_{t,\alpha}\right)$$

where $d_t^{\alpha} \equiv I\left(r_t < VaR_{t,\alpha}\right)$. As Gonzalez-Rivera stress in the article, this is an asymmetric loss function that penalizes more heavily with weight $(1-\alpha)$ the observations for which $r - VaR_{\alpha} < 0$ for high confidence levels whereas the observations for which $r - VaR_{\alpha} \geq 0$, are weighted by α. The best model is the one that minimize the value of Q.

Therefore, for high confidence levels, the weight given to returns such that the Value at Risk prediction is lower than the return (that is, there is no exceedance at that point) is less than the weight given

to exceedances. Thus, the value of the loss function Q increases when more exceedances are produced. However, if two models predictions of Value at Risk have the same number of exceedances, the best model is the one for which the difference between the series of returns and the VaR predictions is lower, that is, the VaR predictions fit better to the series of returns.

It should be noticed that the loss function Q is not differentiable due to the indicator function. In empirical works, the prediction period should be smaller than the estimation period in order to obtain reliable results of the test in the case that parameter estimations are involved in the decided used model. In this case, the rolling method applied in the practical example, obtains predictions for one day using data from a year. Thus, asymptotic trustworthy results are given.

In order to obtain a differentiable loss function, a smoothed loss function $\tilde{Q}$ can be described. In this case, the indicator function, which causes the non differentiability, is replaced by a continuous differentiable function. It is defined as follows:

$$\tilde{Q} = \frac{1}{n}\sum^{t}(\alpha - m_{\delta}(r_t, VaR_{t,\alpha})) \cdot (r_t - VaR_{t,\alpha})$$

where $m_{\delta}(a,b) = [1 + exp\{\delta(a-b)\}]^{-1}$. The parameter δ controls the smoothness and a higher value of δ makes $\tilde{Q}$ closer to Q. If $\delta > 10$ the loss values of both functions are very similar. Different considerations can be made in order to fit the best model by varying the value of δ.

Chapter 3

Quantile regression: a new approach to asset pricing estimation

It is curious that the first trials in doing regression are so closely related to the notions of quantile regression. The first publications about the least-square method is dated 1805, while Boscovich's initial work on regression, and the first approximation to quantile regression, was published in the XVIII century. The object of interest was the study of the ellipticity of the earth.

Boscovich's proposal to the problem was later named the 'method of situation', and it is a bend of mean and median ideas. In 1888, Edgeworth revived these ideas, which had been neglect for over a century. He observed that if observations of a sample were discordant, the median could be superior to the mean, that is, the median had smaller asymptotic variance than the mean. This approach started to be developed considerably later, with the use of linear programming as an efficient computational method.

Regression quantile methods went through an enormous development in the early 80s, with the first results in Robust Statistics. The contributions of Huber and Hampel in Theory of Robustness in 1981 and 1986 respectively and, later, Koenker (see [30]) who is considered the author par excellence of quantile regression theory and has been developing his theory from the 70s until the present, are still in use nowadays.

3.1. Introduction to Quantile Regression

3.1.1. Definition of quantile

Let X be a random variable. The distribution function of X, F, can be defined as follows:

$$F(x) = \wp(X \leq x)$$

Definition 3.1 (Quantile)**.** *Let X be a random variable. For any $0 < \tau < 1$, the quantile of order τ, or τth-quantile, which will be named as q_τ, is the value that verifies:*

$$\wp(X \leq q_\tau) \geq \tau$$

and

$$\wp(X \geq q_\tau) \geq 1 - \tau$$

It should be mentioned that if X is a continuous random variable, then

$$\wp(X \leq q_\tau) = \wp(X < q_\tau) = \tau$$

Therefore, the quantile function of a random variable can be defined as the inverse of its distribution function.

Depending on the nature of X, two cases can be distinguished:

1. If the distribution function, $F : \mathbb{R} \longrightarrow (0,1)$, is continuous and strictly monotonous, the quantile distribution, F^{-1}, gives a value that satisfies:

$$\wp(X \leq x) = \tau$$

2. If F is a discrete distribution function, the inverse function is not well defined and therefore it shall be defined as follows:

$$Q(\tau) = F^{-1}(\tau) = \inf\{x \in \mathbb{R} | \tau \leq F(x)\}$$

3.1.2. The loss function

Given a random variable and its distribution function, F, the quantiles of the distribution can be calculated as the solutions of an optimisation problem. For any $0 < \tau < 1$, the "loss function", also called the "check function", is defined as:

$$\rho_\tau(u) = u\,(\tau - I(u < 0))$$

This function can be illustrated for a general value of q, obtaining:

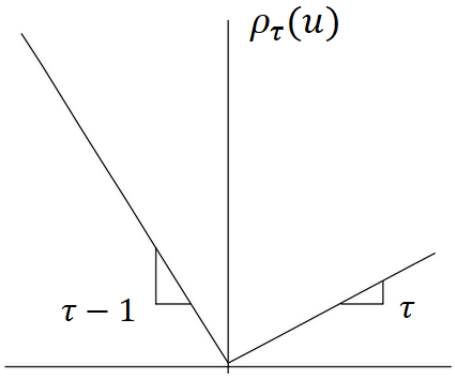


Figure 3.1: Quantile Regression Loss function [30]

The problem focuses on finding a value, $\widehat{x}$, such that it minimizes the expected loss. Thus, the function which shall be minimized with respect to $\widehat{x}$ is:

$$\mathbb{E}\,[\,\rho_\tau\,(X - \widehat{x})\,] = (\tau - 1)\int_{-\infty}^{\widehat{x}} (x - \widehat{x})\,dF(x) + \tau\int_{\widehat{x}}^{\infty} (x - \widehat{x})\,dF(x) \tag{3.1}$$

In order to find the optimum, Equation (3.1) is derived with respect to $\widehat{x}$, obtaining:

$$0 = (1 - \tau)\int_{-\infty}^{\widehat{x}} dF(x) - \tau\int_{\widehat{x}}^{\infty} dF(x) = F\,(\widehat{x}) - \tau$$

Thus, any element of $\{x | F(x) = \tau\}$ minimizes the expected loss. The smallest element of this set must be chosen as the solution of the optimisation problem.

3.1.3. Empirical quantiles

Let X be a random variable taking values $\{x_1, \ldots, x_n\}$. The empirical distribution function is of the form:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\,(X_i \leq x)$$

In this case, the quantiles can be found as solutions of optimisation problems as well. The optimum value will be the value of $\widehat{x}$ for which the minimum of the expected loss is found:

$$\int \rho_\tau (x-\widehat{x})\, dF_n(x) = n^{-1} \sum_{i=1}^{n} \rho_\tau (x_i - \widehat{x}) \tag{3.2}$$

Thus, the quantile of order τ of the sample is found, an will be denoted by, c_τ. Again, the optimisation problem may have multiple solutions, but there is no repercussion in practice.

3.2. Optimisation via Linear Programming

3.2.1. Regression quantiles

From the section above, it shall be noticed that the problem of finding the τth sample quantile has been transformed into finding the solution to a simple optimisation problem.

Recalling Equation (3.2), the linear optimisation problem of searching for the sample quantiles is given by the following objective function:

$$\min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} \rho_\tau (X_i - \beta) \tag{3.3}$$

In order to obtain estimation for quantiles in a model from a sample, the linear regression quantile shall be defined.

Definition 3.2. *Let Y be the response variable of the problem depending on d covariables, X, for which a sample $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ is known.*

The regression quantile $\widehat{\beta}_\tau$ of order τ is defined as:

$$\widehat{\beta}_\tau = \min_{\beta \in \mathbb{R}^d} \sum_{i=1}^{n} \rho_\tau \left(Y_i - X_i^T \beta\right) \tag{3.4}$$

It should be notice that, in Definition 3.1, the quantile q_τ was a number, that is, its dimension was equal to one. In the quantile regression model described in Definition 3.2, the regression quantile, β_τ, is a vector with dimension d.

3.2.2. The linear programming problem

The quantile regression problem can be extended to a linear programming (LP) problem if $2n$ artificial variables, $\{u_i, v_i : i = 1, \ldots, n\}$ are added to the problem. Koenker (see [30]) and Davino et al. (see [17]) give wide explanations and details of all the necessary calculations of this problem. A brief presentation is given in this section.

Let $u = [Y - X\beta(\tau)]_+, v = [X\beta(\tau) - Y]_+$ the two vectors of artificial variables representing the positive and negative parts of $[Y - X\beta(\tau)]$, where $Y = (Y_1, \ldots, Y_n)$ and $X = (X_1, \ldots, X_n)$ and τ is a generic conditional quantile. For simplicity of notation, β will be used to refer to $\beta(\tau)$

The optimisation problem can be reformulated as follows:

$$\min_{\beta} \left\{ 1_n^T u + 1_n^T v | Y = X\beta + u - v, \{u, v\} \in \mathbb{R}_+^n \right\}$$

where 1_n represents an n-vector of ones.

Furthermore, let

$$B = [X - X \cdot I - I]$$

and

$$\psi = \begin{bmatrix} [\beta]_+ \\ -[\beta]_- \\ u \\ v \end{bmatrix}$$

$$s = \begin{bmatrix} 0_n \\ 0_n \\ 1_n \\ 1_n \end{bmatrix}$$

where $[\beta]_+$ is the vector for which each component i takes value β_i if $\beta_i > 0$ and 0 in other case, and $[\beta]_- = [\beta]_+ - \beta$.

The reformulation presents a standard LP problem:

$$\begin{array}{rl} \min & s^T \psi \\ \text{subject to} & B\psi = Y \\ & \psi \geq 0 \end{array}$$

The dual formulation of the problem is given by:

$$\begin{array}{rl} \max & Y^T z \\ \text{subject to} & B^T z \leq s \end{array}$$

which can be simplified as

$$\max_z \left\{ Y^T z | X^T z = 0, z \in \{-1, 1\}^n \right\}$$

In order to solve the problem and find the optimal solution, different methods can be applied. Some of them are:

1. Simplex algorithm:

   This method, proposed by Dantzig in 1947, is the most popular method for solving linear programming problems. The simplex is based on iterations. In a linear programming problem, if the optimal solution exists, the optimal vector occurs in one of the vertices of the feasible set. Thus, the Simplex algorithm consists on movements along the edges of the feasible set until it finds the optimal solution, that is, the solution associated with the minimum value of the objective function.

   Although the Simplex algorithm is commonly used, it can not be very efficient for big optimisation problems.

2. Interior point method:

   This iterative method consists on finding the optimal solution from the interior of the feasible set. Introduced by Karmakar in 1984, it has been proved to be more efficient with respect to the Simplex method, specially on very large problems.

3. Cutting plane method:

   Let f be the general quantile regression function, $f = \sum_{i=1}^{n} \rho_\tau(Y_i - \beta^T X_i)$. The cutting plane method uses the subgradients of f so that the feasible set is cut by new planes (restrictions) and the size of the feasible set is reduced until the optimum is found.

   Let γ be the subgradient of f. The method constructs new points (solutions) $\beta^1, \beta^2, \ldots, \beta^k, \ldots$, where the solution of iteration k, β^{k+1}, is the solution of the following optimisation problem:

$$\begin{array}{rl} \min & z \\ \text{subject to} & \gamma(f,\beta^1)^T\beta - z \leq \gamma(f,\beta^1)^T\beta^1 - f(\beta^1) \\ & \gamma(f,\beta^2)^T\beta - z \leq \gamma(f,\beta^2)^T\beta^2 - f(\beta^2) \\ & \vdots \\ & \gamma(f,\beta^k)^T\beta - z \leq \gamma(f,\beta^k)^T\beta^k - f(\beta^k) \end{array}$$

It is usual to find the Simplex algorithm combined with the Cutting plane method. The search of the solution is easier in this case than in the original Simplex because the feasible set becomes smaller in each iteration of the cutting plane algorithm and therefore less iterations are needed in the Simplex.

See [35] for more details.

4. Finite smoothing algorithm:

   Numerical comparison shows that the finite smoothing algorithm significantly outperforms the simplex algorithm in computing speed for relatively large datasets (n > 3000 or d > 50). For small to moderate datasets, it is competitive. In addition, it is significantly faster than the interior point algorithm when the design matrix in quantile regression has a large number of covariates.

   The algorithm's aim is to approximate the function $\sum_{i=1}^{n} |Y_i - X_i\beta|$ by a smooth function $D_\gamma(\beta)$, which depends on a threshold γ, where

$$D_\gamma(\beta) = \sum_{i=1}^{n} H_\gamma(Y_i - X_i^T\beta)$$

   and

$$H_\gamma(t) = \begin{cases} t^2/(2\gamma) & if \quad |t| \leq \gamma \\ |t| - \gamma/2 & if \quad |t| > \gamma \end{cases}$$

   The difference between the function ρ_τ and its smooth function can be seen in Figure 3.2.

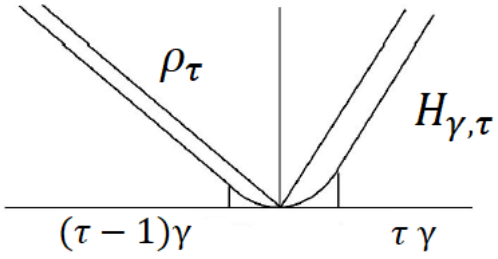


Figure 3.2: Objective functions $H_{\gamma,\tau}$ and ρ_τ

   In this case, the general quantile regression, $\sum_{i=1}^{n} \rho_\tau(Y_i - X_i^T\beta)$, is approximated by the following smooth function:

$$D_{\gamma,\tau}(\beta) = \sum_{i=1}^{n} H_{\gamma,\tau}(Y_i - X_i^T\beta)$$

   where

$$H_{\gamma,\tau}(t) = \begin{cases} t(\tau-1) + \frac{1}{2}(\tau-1)^2\gamma & if \quad t \leq (\tau-1)\gamma \\ \frac{t^2}{2\gamma} & if \quad (\tau-1)\gamma \leq t \leq \tau\gamma \\ t\tau - \frac{1}{2}\tau^2\gamma & if \quad t \geq \tau\gamma \end{cases}$$

   See [13] for more details.

3.3. Robustness of QR

Saying that a method is "robust" (see [26]) means that it is insensible to outliers or to the omission of model assumptions concerned to the data (sample).

It shall be noticed that under the Linear Regression Model (LRM), estimates can be sensitive to outliers. In contrast, the Quantile Regression Model (QRM) estimates are not sensitive to outliers. This robustness is due to the nature of the function that is minimized. If the value of the response variable for a data point lying above (or below) the fitted quantile-regression line is modified, as long as that data point remains above (or below) the line, the fitted quantile-regression line does not change. In other words, if the values of the response variable are modified without changing the sign of the residual, the fitted line remains the same. In this way, the influence of outliers is quite limited.

This behaviour seems logical as soon as the definition of sample quantile is remained. If the sign does not change, the proportion of the sample above (or below) the fitted quantile does not change and therefore, the quantile and the model do not change.

In addition, since the covariance matrix of the estimates is calculated under the normality assumption, a violation of the normality assumption could be thought to cause inaccuracy in standard errors. However, the QRM is robust to distributional assumptions because the estimator weighs the local behaviour of the distribution near the specific quantile more than the remote behaviour of the distribution.

3.4. Estimation and inference

There are numerous approaches to inference for quantile estimation in the literature. The ideal result would be having a finite-sample estimation of parameters. The reality is that, in general, this result cannot be found due to the amount of assumptions that shall be made and the exhausting computational effort (see [30, Chapter 3]).

Therefore, asymptotic estimations are provided. An important issue which cannot be forgotten is: even though some asymptotic results are found and explicitly calculated, do the results converge in some appropriate sense?

Suppose that the τ-th conditional quantile function of Y given $X = x$ takes the parametric form $Q_Y(\tau|X = x) = g(x, \beta)$. As it has been seen in previous sections, the estimator of β, $\widehat{\beta}_n$ in the linear quantile regression problem, is given by Equation (3.4). The question is: under what circumstances does the estimator $\widehat{\beta}_n$ converge to β as $n \longrightarrow \infty$?

Estimations and rates of convergence are given in this section.

3.4.1. Linear Quantile Regression

Let $Y_1, Y_2, \ldots$ be independent random variables with distribution functions $F_1, F_2, \ldots$ and let the τth conditional quantile function be

$$Q_{Y_i}(\tau|X) = x^T \beta(\tau)$$

Thus, the conditional quantile function is linear. Let F_i be the conditional distribution function of Y_i and let

$$Q_{Y_i}(\tau|X) = F_i^{-1}(\tau|X) = \xi_i(\tau)$$

In order to obtain results of asymptotic behaviour of convergence of $\widehat{\beta}$, two Conditions shall be given:

- Condition A1: The distribution functions $\{F_i\}$ are absolutely continuous, with continuous densities $f_i(\xi)$ uniformly bounded away from 0 and ∞ at the points $\xi_i(\tau)$.
- Condition A2: There exist positive definite matrices D_0 and $D_1(\tau)$ such that

  1. $\lim_{n\to\infty} n^{-1} \sum x_i \, x_i^T = D_0$.
  2. $\lim_{n\to\infty} n^{-1} f_i(\xi_i(\tau)) \sum x_i \, x_i^T = D_1(\tau)$.

3. $max_{i=1,\dots,n}||x_i||/\sqrt{n} \to 0$.

Theorem 3.3 ([30]). *Under Conditions A1 and A2,*

$$\sqrt{n}\left(\widehat{\beta}_\tau - \beta\right) \longrightarrow N\left(0, \tau(1-\tau)D_1^{-1}D_0\, D_1^{-1}\right)$$

In the case of iid error model,

$$\sqrt{n}\left(\widehat{\beta}_\tau - \beta\right) \longrightarrow N\left(0, \omega^2 D_0^{-1}\right)$$

where $\omega^2 = \tau(1-\tau)/f_i^2(\xi_i(\tau))$.

3.4.2. Non-linear Quantile Regression

Linear quantile models are the usual choice for many applications and practical cases because of the simplicity of the model (comparing to other QR models) and the knowledge in the field of linear QR estimation, as it has been seen in Section 3.4.1. However, there is also a certain interest to offer the analogous asymptotic approach in the case of non-linear quantile regression models.

The conditional quantile model with nonlinear parameters is expressed as:

$$Q_{Y_i}(\tau|X) = g(X, \beta)$$

where g is a nonlinear function. The nonlinear quantile regression estimator, $\widehat{\beta}_\tau$ is given by

$$\widehat{\beta}_\tau = \min_{b\in B} \sum_{i=1}^{n} \rho_\tau\left(Y_i - g\left(X_i^T, b\right)\right)$$

where $B \in \mathbb{R}^p$ is compact. Condition A1 presented in the previous Section is maintained in order to obtain asymptotic results, with $\xi_i(\tau) = g(X_i, \beta)$. Other assumptions related to the function g shall be made (see [30]):

- Condition G1: There exist constants k_0, k_1 and n_0 such that, for $\beta_1, \beta_2 \in B$ and $n > n_0$,

$$k_0||\beta_1 - \beta_2|| \leq \left(n^{-1}\sum_{i=1}^{n}\left(g(X_i,\beta_1) - g(X_i,\beta_2)\right)^2\right)^{1/2} \leq k_1||\beta_1 - \beta_2||.$$

- Condition G2: There exist positive definite matrices D_0 and $D_1(\tau)$ such that, with $\dot{g}_i = \partial g(X_i, \dot{\beta})/\partial\dot{\beta}|_{\dot{\beta}=\beta}$,

  1. $\lim_{n\to\infty} n^{-1}\sum \dot{g}_i\, \dot{g}_i^T = D_0$.
  2. $\lim_{n\to\infty} n^{-1} f_i(\xi_i)\, \sum \dot{g}_i\, \dot{g}_i^T = D_1(\tau)$
  3. $max_{i=1,\dots,n}||\dot{g}_i||/\sqrt{n} \longrightarrow 0$

Under these conditions, the estimator of β satisfies

$$\sqrt{n}\left(\widehat{\beta}_\tau - \beta\right) \longrightarrow N\left(0, \tau(1-\tau)D_1^{-1}D_0\, D_1^{-1}\right)$$

As in the linear context, a uniform linear representation can be made, but, in contrast to the linear quantile regression estimation, the computation of the process is considerably more tedious.

3.4.3. Scalar Sparsity Estimation

The sparsity function can be defined as the reciprocal function of the density quantile function. That is, given the density quantile function $\phi(\tau) = f\left(F^{-1}(\tau)\right)$, the sparsity function is given by

$$s(\tau) = \frac{1}{f\left(F^{-1}(\tau)\right)}$$

In the case of iid error quantile regression models, the sparsity is analogous to the standard deviation of the least square errors of the iid errors of a regression model. The precision of quantile estimates should depend on the sparsity since it reflects the density of observations near the quantile of interest. If the data is very sparse at the quantile of interest it will be difficult to estimate. On the other hand, when the sparsity is low, so observations are very dense, the quantile will be more precisely estimated.

Since $s(\tau) = dF^{-1}(\tau)/d\tau$, it is natural to consider that a good an estimator of $s(\tau)$ can be a simple difference quotients of an empirical quantile function:

$$\widehat{s_n}(\tau) = \frac{\left[F_n^{-1}(\tau + h_n) - F_n^{-1}(\tau + h_n)\right]}{2h_n}$$

The optimal h_n which minimize the mean squared error of the expression above was obtained by E. Bofinger (see [8]) and is of the form:

$$h_n = n^{1/5}\left[\frac{4.5s^2(\tau)}{s''(\tau)}\right]^{1/5}$$

Once h_n is chosen, F_n^{-1} can be estimated by the empirical quantile function of the residuals from the quantile regression fit. Therefore,

$$F_n^{-1}(t) = \widehat{u}_{(i)}\ ,\ \ t \in [(i-1)/n, i/n)$$

where $\widehat{u}_i = y_i - x_i^T\widehat{\beta}$. Thus, for $t \in [0,1]$, the function F is defined from the ordered regression quantile estimations.

3.5. QR approach to VaR: CAViaR model

Engle and Manganelli (see [21]) proposed an approach to Value at Risk based on quantile estimation. That is, instead of trying to model the entire distribution, they focused on modelling the quantiles of the distribution.

Empirical results show that the distribution of the volatility of stock market returns is autocorrelated over time. As VaR is closely related to the standard deviation of this distribution, it should experiment a similar behaviour. Therefore, an autoregressive behaviour is expected. The Conditional Autoregressive Value at Risk model (CAViaR) was developed under this premise.

3.5.1. The CAViaR method

Definition 3.4 (CAViaR process)**.** *Let $\{r_t\}_{t=1}^T$ be a vector of observed portfolio returns, let α be the probability associated to VaR and $\mathbf{x}_t$ a vector of observable variables at time t, which represents the information set at time t. Let β be a vector of unknown parameters and let $q_t(\beta,\alpha) \equiv q_t(\mathbf{x}_{t-1},\beta)$ denote the time t α-quantile of the distribution of portfolio returns formed at time $t-1$. The generic CAViaR specification is of the form:*

$$q_t(\beta,\alpha) = \beta_0 + \sum_{i=1}^{s}\beta_i q_{t-i}(\beta,\alpha) + \sum_{j=s+1}^{r+s+1}\beta_j l(\mathbf{x}_{t-j}) \tag{3.5}$$

where $p = s+r+1$ is the dimension of β (vector of βs to be estimated) and l is the function of a finite number of rates of return (number of lagged values of observables, that is, the function connecting Value at Risk with the data set).

The second term of the expression, $\sum_{i=1}^{s} \beta_i q_{t-i}(\beta)$, is autoregressive and enables smooth transition of the quantile over time. Engle y Mananelli (2004) (see [21]) say that the first order is sufficient for practical use.

It is true that the behaviour of the dependence of VaR on lagged returns could be thought as similar for negative and positive returns. GARCH models suggest that if the return r_{t-1} reaches a very negative value, the absolute value of VaR is expected to decrease in case another crash happens.

On the other hand, very positive returns could decrease this value as well. The reason is that if an extreme positive return is given at time t, that is, the value of the asset suffers a significant increase, it is highly probable that the value of the asset at time $t+1$ is lower than the value at time t. The reason is that if the value of the asset at time t is higher than expected, the value at time $t+1$ does not usually keep these high values. Therefore, the return at time $t+1$ is negative as $P_{t+1} - T_t < 0$. In these cases, the VaR decreases.

Engle and Manganelli propose different objective functions, l, giving various alternatives to the CAViaR model:

1. Adaptative:
$$q_t(\beta) = q_{t-1}(\beta) + \beta \left\{ [1 + \exp(G(r_{t-1} - q_{t-1}(\beta)))]^{-1} - \alpha \right\}$$
where G is a positive finite number. If $G \longrightarrow \infty$, then $q_t(\beta_1)$ tends to the indicator function almost surely.

2. Symmetric Absolute Value (SAV):
$$q_t(\beta) = \beta_0 + \beta_1 q_{t-1}(\beta) + \beta_2 |r_{t-1}|$$

3. Asymmetric Slope (AS):
$$q_t(\beta) = \beta_0 + \beta_1 q_{t-1}(\beta) + \beta_2 (r_{t-1})^+ + \beta_3 (r_{t-1})^-$$

4. Indirect GARCH(1,1):
$$q_t(\beta) = \left(\beta_0 + \beta_1 q_{t-1}^2(\beta) + \beta_2 r_{t-1}^2\right)^{1/2}$$

Notice that the SAV and the Indirect GARCH models do not distinguish the magnitude by which the VaR is exceeded. Thus, when it is exceeded by a large value, VaR is simply increased by the same amount in the same way as it would have been a small margin (see [12]) and they respond equally to positive and negative returns. Plus, the Indirect GARCH model returns positive values only. Therefore, in practice, for high confidence levels $1-\alpha$, the estimation of the minus series of returns is used to calculate the $1-\alpha$ Value at Risk, which is equivalent to $-VaR_\alpha$ of the original series of returns.

In the case of the AS model, the positive and negative returns are treated separately and therefore, the model allows the responds to positive and negative returns to be different.

The CAViaR method can be used when the error distributions are not independent and identically distributed, or volatilities or error distributions change.

Note 3.5. *It should be notice that once the definition of the CAViaR model is defined, the $VaR_{t,\alpha}(h)$ is given by the value $-q_t(\mathbf{x}_{t-1}, \beta)$. This fact is trivial from the definition of VaR at certain time t and time horizon h with probability α and the definition of the function q.*

Therefore, Equation (3.5) *can be also written as follows:*

$$-VaR_{t,\alpha}(h) = \beta_0 + \sum_{i=1}^{s} \beta_i \cdot (-VaR_{t-i,\alpha}(h)) + \sum_{j=s+1}^{r+s+1} \beta_j l(\mathbf{x}_{t-j})$$

3.5.2. Parameter estimation

The parameters of the CAViaR models are estimated by regression quantiles. Let $\mathbf{x}_t$ be a vector of regression quantiles of dimension p and consider the following system:

$$r_t = \mathbf{x}'_t\beta + \varepsilon_{\alpha,t}\ ,\ \ q(\varepsilon_{\alpha,t}|\mathbf{x}_t) = 0 \tag{3.6}$$

where $q(\varepsilon_{\alpha,t}|\mathbf{x}_t)$ is the α-quantile conditional function of $\varepsilon_{\alpha,t}$ given $\mathbf{x}_t$.

Now, let $q_t(\beta) \equiv \mathbf{x}_t\beta$. The α-regression quantile is given by any solution $\widehat{\beta}$ of the following minimization problem:

$$\min_{\beta} \frac{1}{n}\sum_{t=1}^{n} [\alpha - I(r_t < q_t(\beta))]\,[r_t - q_t(\beta)] \tag{3.7}$$

Regression quantiles are estimated and solved using least absolute deviation models. These types of models are consistent and more robust than least square methods, specially when a fat-tailed distribution of errors is given.

Consider the model given in Equation (3.6). This model can be rewritten using q_t as follows:

$$r_t = f(r_{t-1}, \mathbf{x}_{t-1}, \ldots, r_1, \mathbf{x}_1; \beta) + \varepsilon_{\alpha,t} \equiv q_t(\beta) + \varepsilon_{\alpha,t} \quad t = 1, \ldots, n \tag{3.8}$$

where $q(\varepsilon_{\alpha,t}|\mathbf{x}_t) = 0$, as q_t is a linear combination of $r_{t-1}, \mathbf{x}_{t-1}, \ldots, r_1, \mathbf{x}_1$ depending on the vector of parameters β. Let $\widehat{\beta}$ be the estimator of β, that is, the parameter vector that minimizes the objective function given in Equation (3.7). It can be shown that this estimator is consistent and asymptotically normal. These results are explained in the following theorems.

Theorem 3.6. *Let the model in Equation 3.8 be under the following assumptions ([21, Appendix A]):*

1. *(Ω, F, P) is a probability space in which the random vectors $\{\varepsilon_{t,\alpha}, \mathbf{x}_t\}$, $t = 1, \ldots, n$, are allocated.*
2. *The function q_t is such that for each compact $\beta \in B$, $q_t(\beta)$ is measurable and q_t is continuous in B, for the given information at time t, Ω_t and $t = 1, \ldots, n$.*
3. *The conditional error $\varepsilon_{t,\alpha}$ given past information, Ω_t, forms a stationary process with continuous conditional density function $h_t(\varepsilon|\Omega_t)$.*
4. *There exists $h > 0$ such that $\forall t, h_t(0|\Omega) \geq h$.*
5. *There exists a stochastic function of variables that belong to the information set with upper-bounded expectation, $K(\Omega_t)$, such that $|q_t(\beta)| < K(\Omega_t)$*
6. *The expectation of the absolute values of the errors at time t is finite for all t.*
7. *The function $\{[\alpha - I(r_t < q_t(\beta))][r_t - q_t(\beta)]\}$ obeys the law of large numbers.*
8. *For every $\xi > 0$, there exists a $\tau > 0$ such that if $||\beta' - \beta|| \geq \xi$ then $\liminf_{n\to\infty} n^{-1}\sum P[|q_t(\beta') - q_t(\beta)| > \tau] > 0$*

Then, $\widehat{\beta} \xrightarrow{p} \beta$, where $\widehat{\beta}$ is the solution to the minimization problem given in Equation (3.7).

Theorem 3.7. *Under asymptotical normality assumptions given in [21, Appendix A] and the conditions in Theorem 3.6,*

$$\sqrt{n}A_n^{-1/2}D_n\left(\widehat{\beta} - \beta\right) \xrightarrow{d} N(0, I)$$

where A_n and D_n are given as follows:

$$A_n \equiv E\left[n^{-1}\alpha(1-\alpha)\sum_{t=1}^{n} \nabla' q_t(\beta)\nabla q_t(\beta)\right],$$

$$D_n \equiv E\left[n^{-1}\sum_{t=1}^{n} h_t(0|\Omega_t)\nabla' q_t(\beta)\nabla q_t(\beta)\right]$$

where $\nabla q_t(\beta)$ *is the gradient of* $q_t(\beta)$.

The last estimator that shall be mentioned is the Variance-Covariance matrix estimator. This type of estimations is very popular in literature (see [30]). Thus, the Variance-Covariance matrix can be estimated using regression quantiles as follows.

Theorem 3.8. *Let* A_n *and* D_n *be matrices as in Theorem 3.7 and let the conditions of Theorems 3.6 and 3.7 be satisfied. Under the following assumptions:*

there exists two matrices $\widehat{A_n}$ *and* $\widehat{D_n}$ *such that* $\widehat{A_n} - A_n \xrightarrow{p} 0$ *and* $\widehat{D_n} - D_n \xrightarrow{p} 0$, *where*

$$\widehat{A_n} = n^{-1}\alpha(1-\alpha)\ \nabla' q(\widehat{\beta})\ \nabla q(\widehat{\beta})$$

$$\widehat{D_n} = (2n\widehat{c_n})^{-1}\sum_{t=1}^{n} I\left(|r_t - q_t(\widehat{\beta})| < \widehat{c_n}\right)\ \nabla' q_t(\widehat{\beta})\ \nabla q_t(\widehat{\beta})$$

and $\widehat{c_n}$ *satisfies:*

$$\frac{\widehat{c_n}}{c_n} \xrightarrow{p} 1\ ,\ \ c_n = o(1)\ \ and\ \ c_n^{-1} = o(n^{1/2})$$

The matrix $\nabla q(\widehat{\beta})$ *is a* $n \times p$ *matrix with typical row* $\nabla q_t(\widehat{\beta})$

Other assumptions of the estimation problem can be found in [21].

3.5.3. A new test and validation method

The dynamic quantile (DQ) proposed by Engle and Manganelli in [21], is a test for the evaluation of the alternative specifications and has better power properties than other existing tests.

Let

$$Hit_t(\beta) \equiv I(r_t < q_t(\beta)) - \alpha$$

be the so called Hit function. It shall be noticed that this function takes a value $(1-\alpha)$ if the return of the sample at time t is less than the quantile and $-\alpha$ otherwise. It is clear that both the expected value and the conditional expectation of $Hit_t(\beta)$ is 0 and the function is uncorrelated with its lagged values and with $q_t(\beta)$.

Therefore, a natural test in order to measure the goodness of fit of the model is to check whether the test statistic $n^{1/2}X(\widehat{\beta})\ Hit(\widehat{\beta})$ is significally different from 0, where $X_t(\widehat{\beta}), t = 1,\ldots,n$, is the typical row of $X(\widehat{\beta})$ and $Hit(\widehat{\beta}) = (Hit_1(\widehat{\beta}),\ldots,Hit_n(\widehat{\beta}))'$.

Let

$$M_n \equiv X'(\beta) - E\left[n^{-1}X'(\beta)\ H\ \nabla q(\beta)\right] D_n^{-1}\ \times\ \nabla' q(\beta)$$

where H is the diagonal matrix with typical entry $h_t(0|\Omega_t)$ and h_t is the function obtained as in Theorem 3.6. The two tests proposed by Engle and Manganelli are the in-sample Dynamic Quantile test and the out-of-sample Dynamic Quantile test.

In this practical case, the in-sample DQ test is the useful case of study. The in-sample method is a specification for CAViaR process under study and it can be found very useful for model selection. In the case of the out-of-sample test, it can be used by external regulators of the model to check whether the VaR estimates submitted are good enough and satisfy certain conditions.

Therefore, the in-sample method is the useful one in the later empirical study. Using assumptions proposed in [21, Appendix A], the test holds.

Theorem 3.9 (In-sample Dynamic Quantile test)**.** *Under assumptions of Theorem 3.6, Theorem 3.7 and [21, Appendix A], the DQ satisfies:*

$$DQ \equiv \frac{Hit'(\widehat{\beta})X(\widehat{\beta})\ (\widehat{M_n}\widehat{M_n}')^{-1}\ X'(\widehat{\beta})Hit'(\widehat{\beta})}{\alpha(1-\alpha)} \sim \chi^2_q \quad as\ n \longrightarrow \infty$$

where

$$\widehat{M_n} \equiv X'(\beta) - \left\{ (2n\widehat{c_n})^{-1} \sum_{t=1}^{n} I\left(|r_t - q_t(\widehat{\beta})| < \widehat{c_n}\right) \times X_t'(\widehat{\beta})\nabla q_t(\widehat{\beta}) \right\} \widehat{D_n}\nabla' q(\widehat{\beta})$$

Thus, the aim of this test is to check the occurrence of autocorrelation among the exceedances of the VaR forecasts and also verify whether the number of exceedances agrees with the expectation (α-level).

The null hypothesis of the DQ test is that, all the coefficients of the following quantile regression

$$Hit_t = \beta_0 + \sum_{i=1}^{p} \beta_i Hit_{t-i} + \sum_{j=1}^{q} \mu_j r_{t-j} + \varepsilon_t,$$

are zero, and the alternative hypothesis is that at least one of the coefficients is significantly different from zero.

Theorem 3.10 (Out-of-sample Dynamic Quantile test)**.** *Let T_R be the number of in-sample observations and N_R the number of out-of-sample observations. Under assumptions of Theorem 3.6, Theorem 3.7 and [21, Appendix A], the DQ satisfies:*

$$DQ \equiv \frac{N_R^{-1} Hit'(\widehat{\beta}_{T_R})\ X(\widehat{\beta}_{T_R}) \left[X'(\widehat{\beta}_{T_R}) \cdot X(\widehat{\beta}_{T_R})\right]^{-1} \times X'(\widehat{\beta}_{T_R})\ Hit'(\widehat{\beta}_{T_R})}{\alpha(1-\alpha)} \sim \chi^2_q \quad as\ R \longrightarrow \infty$$

The principal advantage of the out-of-sample DQ test is that it does not depend on the estimation procedure, that is, in order to implement it, the evaluation consists only on comparing the sequence of the obtained VaR forecasts at echa time t and the corresponding values of the return at time t for each VaR.

Chapter 4

Practical example: VaR estimation for historical series of returns

This chapter will show a practical example of how to fit different methods, including quantile regression models, to a data set in order to estimate the daily VaR for a range of dates. The data sets that has been used are daily rates of returns calculated based on the formula:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Due to the desire of analysing VaR forecasting in the Spanish Market, three companies from the IBEX35 has been selected. The data series can be found in [27]. Each of them belongs to a different economic sector:

- Banco Santander: it belongs to the Financial sector and it has been selected in this analysis because the behaviour of its returns is very similar to the IBEX35 itself due its high weight of 15.48 in the IBEX35 calculation.
- Endesa: it belongs to the sector of Energy and it has a weight of 1.57.
- Indra: this company belongs to the consultancy sector and has a weight of 0.41.

The analysis of the Euro-Dollar exchange rate has also been made. As not all the models are expected to work well on the considered series, different sorts of data are studied in order to see the accuracy of the predictions of each model depending on the nature of the observations. The difference of behaviour in terms of length, scale, mean and deviation of data are described in Table 4.1 and Table 4.2.

In line with the regulatory suggestions made by the Basel Committee, Value at Risk forecasts have been calculated at a confidence level of 99% and one-day ahead forecast.

The data will be analysed using the system for statistical computation and graphics, R. The R language (see [40]) is worldwide used by mathematicians for the development of statistical software and data analysis. The complete R code used in this chapter can be found in Appendix A.

4.1. Rolling Window and Backtesting criterion

In order to obtain a Value at Risk forecast for a given series of returns in this practical example, the Rolling Window method and the Backtesting Criterion are applied for every chosen model.

Basing on [37], the application of the back-testing criterion can be given by the following steps:

1. Get adjusted n closing price and returns of stock data for any company.
2. Proceed as follows (see Figure 4.1):

Series	Beginning	End	Observations
Banco Santander	January 2, 1996	July 31, 2018	5672
Endesa	January 2, 1996	July 31, 2018	5667
Indra	March 24, 1999	July 31, 2018	4853
Euro-Dollar	May 15, 1997	July 31, 2018	6186

Table 4.1: General data of the four analysed series

a) Calculate VaR_α on the basis of w days data (width of the window of observations) for the next day of the last window observation.

b) Compare VaR_α with actual loss of the next day.

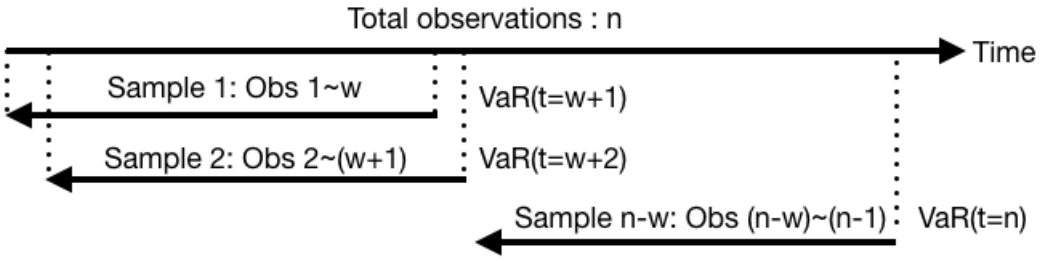


Figure 4.1: Value at Risk forecast method of Rolling Window

c) Repeat these two steps $n - w$ times, each time with a new window shifted by 1 day forward.

d) Measure the accuracy of the model forecasts and predictions with the Performance Tests and Cost functions and compare it to the rest of the models considered in the analysis.

4.2. Data and exploratory analysis

For each of the considered series, the sample length and detailed dates can be found in 4.1.

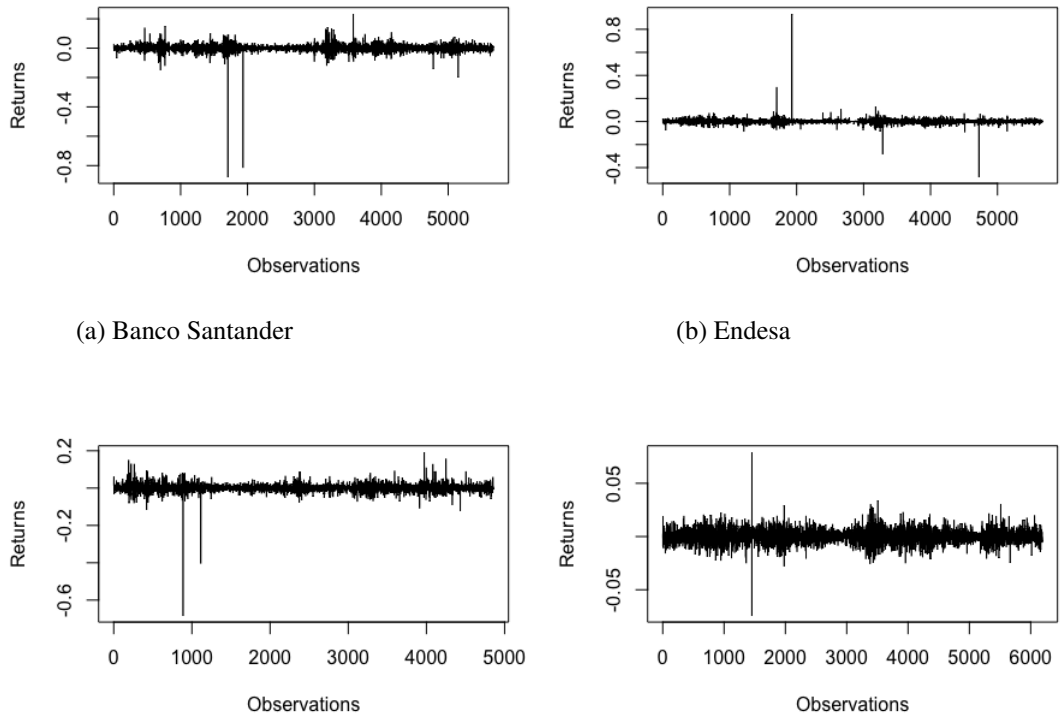
The difference between the number of estimations can be explained by business decisions of companies and by the number of closing and opening days markets have, specially in the case of the Euro-Dollar series. The Returns of the considered series can be seen in Figure 4.2.

Some properties of the analysed series are given as follows:

In light of this descriptive analysis of the series, it can be seen that the behaviour of data is different in terms of skewness and kurtosis, which may have an effect in model estimations. For this reason, series from several sectors of the IBEX35 have been selected and analysed.

It shall be noticed that Kurtosis and Skewness coefficients of the first three series are considerably higher than expected. This could be happening because of the presence of outliers.

Different methods in order to detect outliers are commonly used. In this case, all the series have been analysed using the *Outliers* R package [31]. The first step of the method consists of detecting the value or values which are the farthest from the mean of the series. If a lot of outliers are detected, the median is a more robust measure. After removing the outliers, normality tests are applied and QQ-plots are analysed in order to check the normality of the values of the sample. This method works for large series, which is the case in this practical examples. All the notorious outliers detected in the series are



(a) Banco Santander

(b) Endesa

(c) Indra

(d) Euro-Dollar

Figure 4.2: Returns of the analysed series

found in the years corresponding to the financial crisis (2008). Thus, these values do not interfere in the estimation of the VaR given in Section 4.3.

It shall be remarked that in the case of GARCH models, as the mean of all the series of returns remain approximately constant and close to 0, for each considered window of the rolling method, the mean of the series is assumed to be constant.

4.3. Estimation of VaR and Evaluation of predictions

The evaluation of models in order to estimate the VaR for each considered series is presented in this section.

Both the 5% and the 1% VaR are calculated in the case of the Euro-Dollar Exchange Rate series. Once all the models are fitted, results of performance tests and cost function values are given. In the light of the Basel Accords, the 1% VaR is fitted in the case of the IBEX35 series.

The models presented in this analysis are Historical Simulation, Variance-Covariance with normal distribution assumptions, GARCH(1,1) with both normal and t-Student error distributions, GJR-GARCH and APARCH with t-Student error distribution, Extreme Value functions (Pareto and GEV) and the four CAViaR models presented in Section 3.5.1.

In order to compare the goodness of fit of the models, performance tests has been applied. Thus, the p-value of the Unconditional Coverage (UC), Conditional Coverage (CC) and DQ tests have been obtained.

A model is considered to be accurate to the series if one can accept the three respective null hypothesis, that is, their p-values are greater than 0.05. However, there exist some exceptions to this fact. If the Excess Ratio (ER) differs from the α level, the UC and CC fail the null hypothesis. Although the

	Banco Santander	Endesa	Indra	Euro-Dollar
Mean	$6.501 \cdot 10^{-5}$	$5.255 \cdot 10^{-4}$	$1.156 \cdot 10^{-4}$	$2.108 \cdot 10^{-5}$
Standard Deviation	0.027	0.022	0.023	0.006
Skewness	-10.448	11.427	-5.822	0.131
Kurtosis	330.805	601.974	172.510	9.809
Minimum	-0.877	-0.480	-0.683	-0.074
Maximum	0.232	0.932	0.192	0.079
Jarque-Bera p-value	$< 2.2 \cdot 10^{-16}$	$< 2.2 \cdot 10^{-16}$	$< 2.2 \cdot 10^{-16}$	$< 2.2 \cdot 10^{-16}$

Table 4.2: Summary of Descriptive Measures of the analysed series

α	Basel Light Test	Num. Observations
0.05	Green	$0 \leq Obs \leq 61$
	Yellow	$62 \leq Obs \leq 76$
	Red	$Obs \geq 77$
0.01	Green	$0 \leq Obs \leq 14$
	Yellow	$15 \leq Obs \leq 23$
	Red	$Obs \geq 24$

Table 4.3: Frontiers of the Basel Light Test for 1000 observations and $\alpha = 0.01$ and $\alpha = 0.05$

ER is not equal to α this can happen because the number of excesses is lower than expected. In this cases, the model is considered a good fit.

The Basel Light Test (BSL) is also applied in this case. As 1000 out-of-sample observations are taken for each series, the BLT values, *Green*, *Yellow* and *Red*, depend on the α level, the number of out-of-sample observations and the number of excesses using a cumulative Binomial distribution. In the case of the 1% and 5% levels and 1000 observations, the number of observations delimiting each Zone is detailed in Table 4.3.

Final conclusions about the performance of the CAViaR model in comparison with the rest of the considered models for each series of returns is given at the end of this Section.

4.3.1. Euro-Dollar Exchange Rate VaR

In order to obtain the estimation of the Value at Risk for the returns of the Euro-Dollar exchange rate series, different models have been considered.

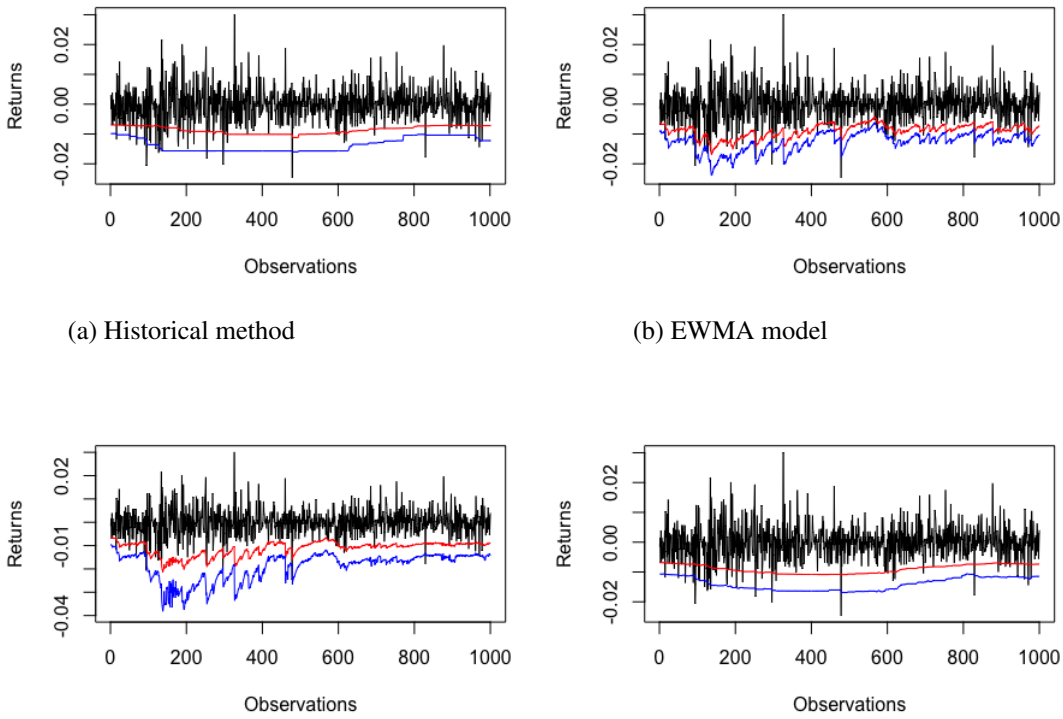
In Chapter 2, classical methods have been explained. In order to obtain estimations from them, the window length, w, shall be given by a fixed number. This width of the window is actually the number of returns considered in the VaR estimation.

The Rolling window method is used and therefore, for each estimation, a length of 500 observations has been considered. It corresponds approximately to two years of observations.

The analysis of results should begin with Historical Simulation, Variance-Covariance method with Normal distribution, GARCH models and Extreme Value distributions. In the case of GARCH models, different parameters have been selected in order to fit the series of returns.

In practice, it is not necessary to consider p and q greater than 1, where p and q are the parameters of a GARCH model. This happens because, in financial return series, correlations of order greater than two do not contribute to the better explanation of the series and the complexity of models grows very

fast (as more parameters need to be estimated). Therefore, p and q are usually considered to be equal to 1, and estimation methods give estimated values for the parameters. The values of the parameters can be analysed during time in order to study the real influence of a lag in the series of returns (in case any of the parameters' estimation is close to zero during time).



(a) Historical method

(b) EWMA model

(c) GARCH-T distribution

(d) Pareto distribution

Figure 4.3: Euro-Dollar VaR estimations with several methods for $\alpha = 0.01$ (blue) and $\alpha = 0.05$ (red)

Thus, estimations of one day ahead VaR forecasts for the 1000 last days (four years) of the series of returns are given and orders of the models (e.g. p and q from GARCH) are calculated with the rest of the sample.

In Figure 4.3a, estimations of VaR using the Historical method are given for $\alpha = 0.01$ (blue) and $\alpha = 0.05$ (red) for the last 1000 days. For each observation, the previous 500 days have been taken in order to calculate the parameters of the model and the VaR has been obtained from the fitted model.

Once the ACF and the PACF several types of models are proposed: a heterocedastic (GARCH(1,1)) model, a model which takes asymmetry into account as well as heterocedasticity (APARCH ($\delta = 2$)), a GJR-GARCH(1,1) and an EWMA($\lambda = 0.94$) model. Different types of distributions have also been considered for the series of errors: normal and t-student distributions.

The estimation of VaR for the last 1000 observations for the EWMA model with normal distribution of errors is given in Figure 4.3b and the VaR estimations for the GARCH(1,1) with t-Student distribution of errors is given in Figure 4.3c.

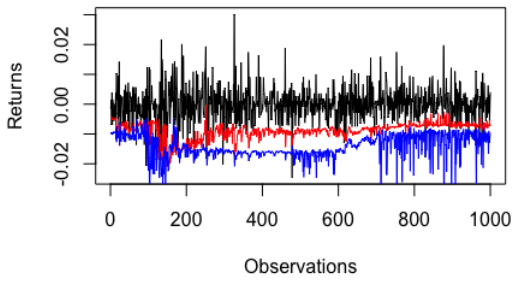
In the same line, GEV and Pareto distributions can also be fitted to the series of returns (see Figure 4.3d). The parameters of the GEV distribution are calculated using MLE for each observation using the previous 500 days of the series and block maxima has been applied. The GEV distribution has been fitted with a length block equal to 5. Thus, parameters are fitted using ta series where each observation is the minimum return within 5 days (one week).

In the case of the Pareto distribution, a threshold shall be chosen. Therefore, data under the 15% - quantile is considered as the extreme events of the series of returns for each sample of the rolling

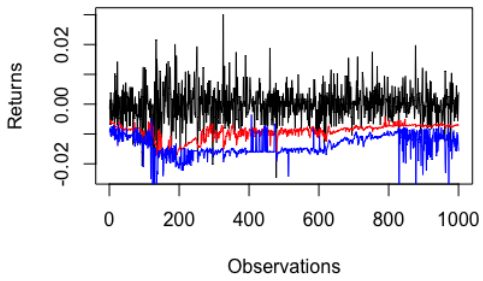
window.

It should be noticed that this series of exchange rate presents high volatility. Therefore, evaluation of VaR in these circumstances usually results as a counter-productive process.

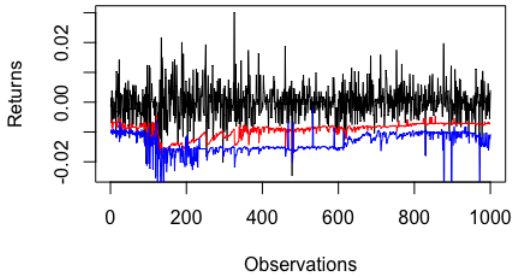
Once all the classical methods are applied and VaR estimations are obtained, a similar process can be done with CAViaR methods. Therefore, the Value at Risk at 1% and 5% of the last 1000 observations with a window of length 500 has been taken for the Adaptative, GARCH, Asymmetric Slope and Symmetric Absolute Value methods have been used.



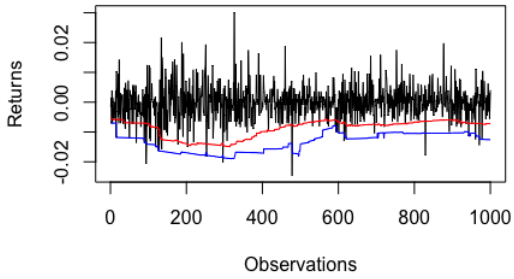
(a) Asymmetric Slope



(b) Symmetric Absolute Value



(c) GARCH



(d) Adaptative

Figure 4.4: Euro-Dollar VaR estimations with CAViaR methods for $\alpha = 0.01$ (blue) and $\alpha = 0.05$ (red)

In Figure 4.4, it can easily be seen that CAViaR methods do not remain stable in high volatility scenarios.

Once all the models have been analysed, results of all the tests are presented in Table 4.4. For each model and confidence level (95% and 99%), the number of excesses, Excess ratio, Basel light test and Conditional and Unconditional coverage and DQ tests' p-values have been calculated. A model is considered to fit the series if the p-values of the three tests are greater than the significance level of 0.05. Cost function results (QL, FC and Basic Loss function) are also illustrated in this Table.

It can easily be seen that in the case of the 95% confidence level models, all of them are in the Green Zone of the Basel Light Test but it does not mean that all of them are good fits. In general, CAViaR models accept the null hypothesis for the three tests presented in the Table (p-value greater than 0.05). Comparing the four models, the CAViaR GARCH can be seen as the best of them, as its cost functions values are lower than in the rest of the CAViaR models.

However, GARCH models also have a good behaviour in general and the number of excesses of the GARCH-T and the GARCH-GJR-T make the UC and the CC tests fail the null hypothesis. That is, the ER is not similar to the α level, that is, the number of excesses obtained in the forecasts is lower than expected (the expected number of excesses for the α level and a number of forecasts, n, is $\alpha \cdot n$). In fact, they are the lowest in each considered series. For this reason, the UC and the CC tests fails the

Euro-Dollar Exchange Rate										
Model	$\alpha\%$	E	ER	LT	UC	CC	DQ	QL	FC	BLF
Hist. sim.	5	60	0.060	Green	**0.159**	**0.349**	0.000	$6.5 \cdot 10^{-4}$	**0.0090**	$1.597 \cdot 10^{-6}$
	1	17	0.017	Yellow	0.043	**0.074**	0.007	$1.90 \cdot 10^{-4}$	0.0136	$2.937 \cdot 10^{-7}$
Delta	5	59	0.059	Green	**0.204**	**0.428**	0.000	$6.64 \cdot 10^{-4}$	0.0091	$1.653 \cdot 10^{-6}$
	1	27	0.027	Red	0.000	0.000	0.000	$2.27 \cdot 10^{-4}$	**0.0125**	$5.969 \cdot 10^{-7}$
EWMA	5	53	0.053	Green	**0.666**	**0.709**	**0.205**	$6.178 \cdot 10^{-4}$	0.0091	$1.208 \cdot 10^{-6}$
	1	19	0.019	Yellow	0.011	0.027	0.009	$1.880 \cdot 10^{-4}$	0.0126	$4.664 \cdot 10^{-7}$
APARCH-T	5	30	0.030	Green	0.000	0.000	0.023	$6.416 \cdot 10^{-4}$	0.0111	$6.945 \cdot 10^{-7}$
	1	8	0.008	Green	**0.510**	**0.755**	**0.378**	$2.064 \cdot 10^{-4}$	0.0185	$1.504 \cdot 10^{-7}$
GARCH-N	5	52	0.052	Green	**0.773**	**0.395**	**0.435**	$6.170 \cdot 10^{-4}$	0.0092	$1.184 \cdot 10^{-6}$
	1	21	0.021	Yellow	0.002	0.006	0.000	$1.915 \cdot 10^{-4}$	0.0128	$4.222 \cdot 10^{-7}$
GARCH-T	5	29	0.029	Green	0.000	0.004	0.013	$6.551 \cdot 10^{-4}$	0.0113	$\mathbf{6.874 \cdot 10^{-7}}$
	1	7	0.007	Green	**0.313**	**0.573**	**0.792**	$2.099 \cdot 10^{-4}$	0.019	$\mathbf{1.120 \cdot 10^{-7}}$
GARCH-GJR-T	5	25	0.025	Green	0.000	0.000	0.021	$6.479 \cdot 10^{-4}$	0.011	$7.081 \cdot 10^{-7}$
	1	6	0.006	Green	**0.170**	**0.376**	**0.813**	$2.078 \cdot 10^{-4}$	0.019	$1.461 \cdot 10^{-7}$
GEV	5	57	0.057	Green	**0.320**	**0.384**	0.000	$6.563 \cdot 10^{-4}$	0.0093	$1.485 \cdot 10^{-6}$
	1	14	0.014	Green	**0.231**	**0.204**	0.005	$1.891 \cdot 10^{-4}$	0.0153	$1.770 \cdot 10^{-7}$
Pareto	5	56	0.056	Green	**0.393**	**0.691**	0.001	$6.478 \cdot 10^{-4}$	0.0093	$1.417 \cdot 10^{-6}$
	1	17	0.017	Yellow	0.043	**0.074**	0.001	$1.931 \cdot 10^{-4}$	0.014	$2.841 \cdot 10^{-7}$
CAViaR SAV	5	51	0.051	Green	**0.885**	**0.491**	**0.302**	$6.222 \cdot 10^{-4}$	0.0096	$1.114 \cdot 10^{-6}$
	1	16	0.016	Yellow	**0.079**	**0.166**	0.006	$1.924 \cdot 10^{-4}$	0.014	$3.200 \cdot 10^{-7}$
CAViaR AS	5	54	0.054	Green	**0.566**	**0.847**	0.010	$6.394 \cdot 10^{-4}$	0.0093	$1.431 \cdot 10^{-6}$
	1	16	0.016	Yellow	**0.079**	**0.112**	0.029	$1.975 \cdot 10^{-4}$	0.014	$3.055 \cdot 10^{-7}$
CAViaR GARCH	5	46	0.046	Green	**0.557**	**0.566**	**0.807**	$\mathbf{6.144 \cdot 10^{-4}}$	0.0096	$9.996 \cdot 10^{-7}$
	1	14	0.014	Green	**0.139**	**0.266**	**0.276**	$1.877 \cdot 10^{-4}$	0.014	$2.940 \cdot 10^{-7}$
CAViaR Adap.	5	60	0.060	Green	**0.159**	**0.362**	0.000	$6.476 \cdot 10^{-4}$	0.0092	$1.404 \cdot 10^{-6}$
	1	21	0.021	Yellow	0.002	0.007	0.000	$\mathbf{1.846 \cdot 10^{-4}}$	0.013	$2.877 \cdot 10^{-7}$

Table 4.4: Test Results of Euro-Dollar Exchange Rate series: Number of Excesses (E), Excess Ratio (ER), Basel Light Test (LT), Kupiec (UC), Christoffersen (CC) and DQ p-values, Quantile Loss (QL), Caporin (FC) and Basic Loss function (BLF) cost functions' values

Banco Santander series of Returns									
Model	E	ER	LT	UC	CC	DQ	QL	FC	BLF
Historical sim.	14	0.014	Green	**0.231**	**0.399**	0.000	$8.659 \cdot 10^{-4}$	0.048	$3.473 \cdot 10^{-5}$
Delta method	16	0.016	Yellow	**0.079**	**0.166**	0.000	$8.819 \cdot 10^{-4}$	0.047	$3.598 \cdot 10^{-5}$
EWMA	17	0.017	Yellow	0.043	**0.096**	0.005	$8.081 \cdot 10^{-4}$	**0.045**	$2.981 \cdot 10^{-5}$
APARCH-T	5	0.005	Green	**0.079**	**0.208**	**0.861**	$8.464 \cdot 10^{-4}$	0.062	$2.178 \cdot 10^{-5}$
GARCH-N	13	0.013	Green	**0.362**	**0.556**	**0.186**	$\mathbf{7.987 \cdot 10^{-4}}$	0.046	$2.978 \cdot 10^{-5}$
GARCH-T	6	0.006	Green	**0.170**	**0.376**	**0.929**	$8.484 \cdot 10^{-4}$	0.063	$1.910 \cdot 10^{-5}$
GARCH-GJR-T	7	0.007	Green	**0.314**	**0.573**	0.022	$8.637 \cdot 10^{-4}$	0.064	$2.120 \cdot 10^{-5}$
GEV	10	0.01	Green	**1.000**	**0.904**	0.002	$1.708 \cdot 10^{-3}$	0.054	$3.014 \cdot 10^{-5}$
Pareto	10	0.01	Green	**1.000**	**0.904**	0.002	$1.605 \cdot 10^{-3}$	0.054	$3.040 \cdot 10^{-5}$
CAViaR SAV	20	0.02	Yellow	0.005	0.013	0.002	$8.354 \cdot 10^{-4}$	0.051	$2.414 \cdot 10^{-5}$
CAViaR AS	20	0.02	Yellow	0.005	0.004	0.000	$8.969 \cdot 10^{-4}$	0.059	$\mathbf{1.753 \cdot 10^{-5}}$
CAViaR GARCH	19	0.019	Yellow	0.012	0.027	0.006	$8.501 \cdot 10^{-4}$	0.053	$2.327 \cdot 10^{-5}$
CAViaR Adap.	16	0.016	Yellow	**0.079**	**0.166**	0.000	$8.354 \cdot 10^{-4}$	0.047	$3.020 \cdot 10^{-5}$

Table 4.5: Test Results of Banco Santander series: Number of Excesses (E), Excess Ratio (ER), Basel Light Test (LT), Kupiec (UC), Christoffersen (CC) and DQ p-values, Quantile Loss (QL), Caporin (FC) and Basic Loss function (BLF) cost functions' values

null hypothesis of ER equal to α because the null hypothesis of both tests include to check whether the number of excesses is approximately equal to the expected number of excesses. Thus, all the cases of rejecting the null hypothesis are due to an excess of conservatism, that is, as the number of excesses is lower than expected, the VaR forecasts are lower than it should be if their fits to the series of returns were adequate. Although this could lead to high values for the cost functions, the truth is that, for the considered GARCH models with t-student distribution of innovations, the values of the considered cost functions are average or low comparing to other analysed models. Therefore, GARCH models might be seen as a better choice in general.

In the case of the 99% confidence level, not all the tests are in the green zone of the Basel Light Test. GARCH models with t-Student error distribution behave well in general. They accept UC, CC and DQ null hypothesis and are in the Green Zone of the BL Test and their BLF values are lower than for the rest of the models. However, the values of the FC and QL are, in general, higher than in the rest of the models. The lowest FC values are found in the CAViaR models.

Notice that the CAViaR GARCH model accepts all the null hypothesis and also, the QL function, and the other two functions in general, take low values. Thus, as it has been seen in the graphics above, series with high volatility are well explained with GARCH models, and CAViaR models might not be stable (except for the CAViaR GARCH model, which explains volatility using Quantile Regression).

4.3.2. IBEX35 Series VaR

According to the Basel Committee regulations, the Value at Risk at the significance level of 1% has been calculated for 1000 one-day-ahead out of sample last observations of the returns of series of the IBEX35 closing values of Banco Santander, Endesa and Indra. The obtained results are shown in Tables 4.5 to 4.7.

In VaR forecasts analysis, the GARCH(1,1) model with normal distribution function is considered the reference point and basic benchmark. In the case of Banco Santander and Indra, the GARCH-N model is qualified to the green zone of the Basel Light Test but in the case of Endesa, the model is in

Endesa series of Returns									
Model	E	ER	LT	UC	CC	DQ	QL	FC	BLF
Historical sim.	10	0.010	Green	**1.000**	**0.904**	**0.802**	$9.090 \cdot 10^{-4}$	0.036	$1.996 \cdot 10^{-4}$
Delta method	6	0.006	Green	**0.170**	**0.376**	**0.859**	$9.365 \cdot 10^{-4}$	0.045	$2.014 \cdot 10^{-4}$
EWMA	20	0.020	Yellow	0.005	0.014	0.010	$9.852 \cdot 10^{-4}$	0.036	$2.034 \cdot 10^{-4}$
APARCH-T	3	0.003	Green	0.009	0.033	**0.662**	$9.773 \cdot 10^{-4}$	0.054	$\mathbf{1.738 \cdot 10^{-4}}$
GARCH-N	20	0.020	Yellow	0.005	0.014	0.018	$9.428 \cdot 10^{-4}$	**0.033**	$2.006 \cdot 10^{-4}$
GARCH-T	2	0.002	Green	0.002	0.008	**0.489**	$9.811 \cdot 10^{-4}$	0.054	$1.837 \cdot 10^{-4}$
GARCH-GJR-T	3	0.003	Green	0.009	0.033	**0.657**	$1.002 \cdot 10^{-3}$	0.055	$1.853 \cdot 10^{-4}$
GEV	7	0.007	Green	**0.314**	**0.573**	**0.989**	$\mathbf{9.055 \cdot 10^{-4}}$	0.041	$1.990 \cdot 10^{-4}$
Pareto	6	0.006	Green	**0.170**	**0.376**	**0.965**	$9.125 \cdot 10^{-4}$	0.042	$1.984 \cdot 10^{-4}$
CAViaR SAV	10	0.01	Green	**1.000**	**0.904**	**1.000**	$9.066 \cdot 10^{-4}$	0.038	$1.928 \cdot 10^{-4}$
CAViaR AS	11	0.011	Green	**0.754**	**0.842**	**0.221**	$9.167 \cdot 10^{-4}$	0.037	$1.977 \cdot 10^{-4}$
CAViaR GARCH	10	0.01	Green	**1.000**	**0.904**	**0.241**	$9.426 \cdot 10^{-4}$	0.037	$1.974 \cdot 10^{-4}$
CAViaR Adap.	12	0.012	Green	**0.5378**	**0.715**	**0.294**	$9.206 \cdot 10^{-4}$	0.037	$2.066 \cdot 10^{-4}$

Table 4.6: Test Results of Endesa series: Number of Excesses (E), Excess Ratio (ER), Basel Light Test (LT), Kupiec (UC), Christoffersen (CC) and DQ p-values, Quantile Loss (QL), Caporin (FC) and Basic Loss function (BLF) cost functions' values

Indra series of Returns									
Model	E	ER	LT	UC	CC	DQ	QL	FC	BLF
Historical sim.	11	0.011	Green	**0.754**	**0.842**	**0.111**	$\mathbf{7.704 \cdot 10^{-4}}$	**0.049**	$1.33 \cdot 10^{-5}$
Delta method	12	0.012	Green	**0.538**	**0.715**	0.046	$7.838 \cdot 10^{-4}$	0.052	$1.187 \cdot 10^{-5}$
EWMA	16	0.016	Yellow	**0.079**	**0.166**	0.009	$7.952 \cdot 10^{-4}$	**0.049**	$1.413 \cdot 10^{-5}$
APARCH-T	4	0.004	Green	0.030	**0.094**	0.000	$9.300 \cdot 10^{-4}$	0.082	$3.989 \cdot 10^{-6}$
GARCH-N	13	0.013	Green	**0.362**	**0.556**	0.022	$8.118 \cdot 10^{-4}$	0.053	$1.396 \cdot 10^{-5}$
GARCH-T	5	0.005	Green	**0.079**	**0.208**	0.001	$9.288 \cdot 10^{-4}$	0.084	$3.159 \cdot 10^{-6}$
GARCH-GJR-T	5	0.005	Green	**0.079**	**0.208**	0.001	$9.902 \cdot 10^{-4}$	0.086	$5.352 \cdot 10^{-6}$
GEV	10	0.010	Green	**1.000**	**0.904**	**0.081**	$7.788 \cdot 10^{-4}$	0.053	$1.178 \cdot 10^{-5}$
Pareto	9	0.009	Green	**0.746**	**0.875**	**0.0785**	$7.792 \cdot 10^{-4}$	0.054	$1.152 \cdot 10^{-5}$
CAViaR SAV	13	0.013	Green	**0.362**	**0.556**	0.009	$8.895 \cdot 10^{-4}$	0.060	$1.231 \cdot 10^{-5}$
CAViaR AS	17	0.017	Yellow	0.043	**0.096**	0.000	$9.151 \cdot 10^{-4}$	0.054	$1.758 \cdot 10^{-5}$
CAViaR GARCH	13	0.013	Green	**0.362**	**0.243**	0.000	$8.587 \cdot 10^{-4}$	0.058	$\mathbf{1.159 \cdot 10^{-5}}$
CAViaR Adap.	11	0.011	Green	**0.754**	**0.842**	**0.227**	$7.812 \cdot 10^{-4}$	0.050	$1.488 \cdot 10^{-5}$

Table 4.7: Test Results of Indra series: Number of Excesses (E), Excess Ratio (ER), Basel Light Test (LT), Kupiec (UC), Christoffersen (CC) and DQ p-values, Quantile Loss (QL), Caporin (FC) and Basic Loss function (BLF) cost functions' values

the yellow zone. The number of excesses in the three cases is higher than expected, although for Indra and Santander, the UC, CC and DQ tests accept the null hypothesis of uncorrelation and independence of excesses. However, the loss function values are, in general, higher than in the rest of the considered models but institutions need to keep savings for contingency plans and this model overestimate non-excesses. Thus, the GARCH-N model cannot be considered as a potential model for VaR estimation.

Historical simulation and Variance-Covariance (Delta) models are classic approaches to VaR estimation. The use of historical simulation in the three series leads to a classification in the green BLT zone for the three considered series. It shall be noticed that in the case of Santander, the number of excesses is higher than expected. In fact, the DQ test fails the null hypothesis. Also, the values of the QL and BLF functions are, in general, higher than in order considered models.

The Delta method leaves the estimations for the Banco Santander series in the BLT yellow zone and the DQ test fails the null hypothesis. It should be remarked that in Endesa, the number of excesses is lower than expected and, in this case, the behaviour of estimations could lead to a good estimator. However, the values of the cost functions are higher than in the rest of the models, that is, this model underestimates the VaR.Similar conclusions can be obtained from the Indra series of returns.

Apart from the GARCH-N model, other GARCH models have been considered. Comparing the performance of the GARCH-N model and the rest of them, it can be observed that the choice of t-Student error functions lead to better forecasts. In the case of the APARCH-T, GARCH-T and GARCH-GJR-T models, the number of excesses obtained in the forecasts is lower than expected (the expected number of excesses for the α level and a number of forecasts, n, is $\alpha \cdot n$). In fact, they are the lowest in each considered series. For this reason, the UC and the CC tests fails the null hypothesis of ER equal to 0.01 because the null hypothesis of both tests include to check whether the number of excesses is approximately equal to the expected number of excesses. For this reason, the considered cost functions have large values (in the scale of each series).

Both GEV and Pareto distributions are fitted to the three series in order to analyse the performance of Extreme Value functions in order to predict Value at Risk. In the case of the GEV distribution, Block Minima method has been applied. Thus, the observations used to fit the series where the minimum returns within 5 days. In the three series, the number of excesses obtained adequate to the confidence level and the UC, CC adn DQ p-values are greater than 0.05. Plus, cost functions' values are, in general, average values (there are models with higher values and also with lower values).

The Pareto distribution belongs to the green zone of the BLT and p-values of the UC, CC and DQ are greater than 0.05 except the DQ test of the Santader return series. As well as in the case of the GEV distribution, cost functions' values are average, even lower than the average values. Therefore, the proposed Extreme Value models are assumed to fit well to the proposed series.

Finally, the four CAViaR model approaches have been calculated. In the 4 models for the Santander series and the the Asymmetric Slope CAViaR of the Indra series, the forecasts belong to the yellow zone of the BLT. In general, the UC and CC tests accept the null hypothesis, which means that excesses are uncorrelated and independent. In the case of the Santander series, due to the high volatility, these models are not as stable as in the other two series (as in the Euro-Dollar series of returns).

It shall be noticed that the value of cost functions for the CAViaR model are low, in general. This means that, although these models have a higher number of excesses than the rest of the considered models, they adjust better to the series of returns. Thus, the difference between the value of the returns and the value of the estimated VaR is lower in the CAViaR models. This can also be seen in the case of Extreme Value models, in general.

4.3.3. Conclusions

Even though each model has its strengths and weaknesses, non-basic GARCH models with t-Student distribution have shown their predictive power. However, the values of their cost functions are higher than in other considered models. This is caused by the excessive persistence in the models which make high volatility periods more conservative in their Value at Risk estimations. That is, the VaR is lower than expected and the number of excesses is low. However, the values of the considered cost functions

might not be very good comparing to the rest of the models, as the difference between the return and the VaR estimation is large.

In the Historical and Variance-Covariance models, the results obtained in the performance tests are good in the case of the UC and CC models. However, the DQ test fails in the majority of the series and cases, that is, the observed excesses can be fitted to a quantile regression function.

Although the CAViaR model can be though as a well predictive model, this analysis has shown that the number of excesses of the VaR forecasts exceeds the Excess Ratio and therefore, the number of excesses is larger than expected. However, if the series does not present periods of changing volatility (increasing or decreasing volatility), the CAViaR adjustment to the series leads to low values of the cost functions, which is essential in terms of defining a contingency plan for an institution.

Models with high quality forecasts are Extreme Value models. Therefore, Quantile Regression estimations of VaR via Extreme Value Theory (EVT) could lead to accurate fits with a small number of excesses and with low loss function values, that is, a model combining the strength of both the EV models and the CAViaR. This type of models is gaining reputation day by day (see [14]). Quantile regression is an important tool for estimation of conditional quantiles of a response Y given a vector of covariates X. It can be used to measure the effect of covariates not only in the center of a distribution, but also in the upper and lower tails. This modelling set-up combines restrictions of extreme value theory with leading homoscedastic and heteroscedastic linear specifications of regression analysis.

Appendix A

R Code

A.1. Data and exploratory analysis Code

```
#libraries
library(readxl)
library(lubridate)
library(MASS)
library(sn)
library(e1071)
library(tseries)
library(extRemes)
library(evir)
library(rugarch)
library(GAS)
library(data.table)
source("caviar.R") #From Buzczynsky and Chlebus

dataset <- read_excel("series_returns.xlsx")
#create returns
dataset$Return<-NA
for(i in 2:length(dataset$Ultimo)){
  dataset$Return[i]<-(dataset$Ultimo[i]-dataset$Ultimo[i-1])/dataset$Ultimo[i-1]
}
#basic statistics
mean(dataset$Return)
sd(dataset$Return)
skewness(dataset$Return)
kurtosis(dataset$Return)
min(dataset$Return)
max(dataset$Return)
jarque.bera.test(dataset$Return)
#define window and confidence level
w=500
conf_level<-c(0.95,0.99)
```

A.2. Functions for Value at Risk Estimation Code

```
# HISTORICAL METHOD #
hist_method<-function(dataset,w,conf_level){
  hist_var<-data.frame()
  for(i in w:(length(dataset$Return)-1)){
```

```
    serie<-dataset$Return[(i-w+1):i]
    quant<-quantile(serie,1-conf_level)
    hist_var<-rbind(hist_var,quant)
    }
  vec<-vector()
  for (i in 1:length(conf_level)){
    vec<-c(vec,paste(as.character(conf_level[i]*100),"% VaR",sep=""))
    }
  colnames(hist_var)<-vec
  hist_var$Fecha<-dataset$Fecha[(w+1):length(dataset$Fecha)]
  return(hist_var)
  }
  #hist_var(dataset,500,c(0.9,0.95,0.99))
```

```
# VARIANCE-COVARIANCE METHOD #
delta_method<-function(dataset,w,conf_level,distribution){
  if(distribution=="normal"){
    delta_var<-data.frame()
    for(i in w:(length(dataset$Return)-1)){
      serie<-dataset$Return[(i-w+1):i]
      quant<-(-(mean(serie)+(qnorm(conf_level,0,1)*sd(serie))))
      delta_var<-rbind(delta_var,quant)}
    vec<-vector()
    for (i in 1:length(conf_level)){
      vec<-c(vec,paste(as.character(conf_level[i]*100),"% VaR",sep=""))
      }
    colnames(delta_var)<-vec
    delta_var$Fecha<-dataset$Fecha[(w+1):length(dataset$Fecha)]
    }
  else if(distribution=="t-student"){
    delta_var<-data.frame()
    for(i in w:(length(dataset$Return)-1)){
      serie<-dataset$Return[(i-w+1):i]
      fit<-fitdistr(serie,"t")
      quant<-
          fit$estimate[["m"]]-fit$estimate[["s"]]*qt(conf_level,df=fit$estimate[["df"]])
      delta_var<-rbind(delta_var,quant)
    }
    vec<-vector()
    for (i in 1:length(conf_level)){
      vec<-c(vec,paste(as.character(conf_level[i]*100),"% VaR",sep=""))}
    colnames(delta_var)<-vec
    delta_var$Fecha<-dataset$Fecha[(w+1):length(dataset$Fecha)]
  }
  else{delta_var<-"Error: revise los datos introducidos"}
  return(delta_var)
}
#delta_method(dataset,500,c(0.9,0.95,0.99),"normal")
#delta_method(dataset,500,c(0.9,0.95,0.99),"t-student")
```

```
# EWMA MODEL (normal) #
ewma_model<-function(dataset,w,conf_level,lambda){
  sigma2<-vector()
  for(i in (w+1):(length(dataset$Return))){
    if(i==w+1){
      sigma2<-c(sigma2,var(dataset$Return[(i-w):(i-1)])) #inicializar con una ventana
          temporal
```

```
    } else{
      sigma2<-c(sigma2,((1-lambda)*dataset$Return[i-1]^2) + (lambda*sigma2[i-w-1]))
    }
  }
  ewma_var<-data.frame()
  for(i in w:(length(dataset$Return)-1)){
    serie<-dataset$Return[(i-w+1):i]
    quant<-(-mean(serie)-qnorm(conf_level,0,1)*sqrt(sigma2[i-w+1]))
    ewma_var<-rbind(ewma_var,quant)
  }
  vec<-vector()
  for (i in 1:length(conf_level)){
    vec<-c(vec,paste(as.character(conf_level[i]*100),"% VaR",sep=""))}
  colnames(ewma_var)<-vec
  ewma_var$Fecha<-dataset$Fecha[(w+1):length(dataset$Fecha)]
  return(ewma_var)
}
#ewma_model(dataset,250,c(0.9,0.95,0.99),0.94)
```

```
# GEV DISTRIBUTION #
gev_model<-function(dataset,w,conf_level,len_block){
  gev_var<-data.frame()
  for(i in w:(length(dataset$Return)-1)){
    serie<-dataset$Return[(i-w+1):i]
    gev_dist<-gev(-serie,block=len_block)
    quant<-(-(gev_dist$par.ests[["mu"]]-(gev_dist$par.ests[["sigma"]]
    /gev_dist$par.ests[["xi"]])*(1-(-len_block*log(conf_level))^
    (-gev_dist$par.ests[["xi"]]))))
    gev_var<-rbind(gev_var,quant)
  }
  vec<-vector()
  for (i in 1:length(conf_level)){
  vec<-c(vec,paste(as.character(conf_level[i]*100),"% VaR",sep=""))}
  colnames(gev_var)<-vec
  gev_var$Fecha<-dataset$Fecha[(w+1):length(dataset$Fecha)]
  return(gev_var)
}
#gev_model(dataset,250,c(0.9,0.95,0.99),1)
```

```
# PARETO DISTRIBUTION #
gpd_model<-function(dataset,w,conf_level,porcentaje){
  gpd_var<-data.frame()
  for(i in w:(length(dataset$Return)-1)){
    serie<-dataset$Return[(i-w+1):i]
    fit<-gpd(-serie,threshold = quantile(-serie,(porcentaje/100)))
    quant<-(-riskmeasures(fit,conf_level)[,"quantile"])
    gpd_var<-rbind(gpd_var,quant)
  }
  vec<-vector()
  for (i in 1:length(conf_level)){
    vec<-c(vec,paste(as.character(conf_level[i]*100),"% VaR",sep=""))}
  colnames(gpd_var)<-vec
  gpd_var$Fecha<-dataset$Fecha[(w+1):length(dataset$Fecha)]
  return(gpd_var)
}
#gpd_model(dataset,250,c(0.9,0.95,0.99),85)
```

A.3. Application of models Code

```
hist_series<-hist_method(dataset,500,c(0.99))
delta_series<-delta_method(dataset,500,c(0.99),"normal")
ewma_series<-ewma_model(dataset,500,c(0.99),0.94)

#GARCH MODEL para un valor fijo de garch p,q
library(rugarch)
garch_var<-data.frame()
p=1
q=1
#normal
# 1. APARCH-N
spec_aparch<-ugarchspec(variance.model =
    list(model="fGARCH",submodel="APARCH",garchOrder=c(p,q)),mean.model=
list(armaOrder=c(0,0),include.mean=TRUE),distribution.model = "norm")
# 2. GARCH-N
spec_aparch<-ugarchspec(variance.model =
    list(model="fGARCH",submodel="GARCH",garchOrder=c(p,q)),mean.model=
list(armaOrder=c(0,0),include.mean=TRUE),distribution.model = "norm")

for(i in (length(dataset$Return)-1600):(length(dataset$Return)-1)){
  print(i)
  serie<-dataset$Return[(i-w+1):i]
  fit<-ugarchfit(data=serie,spec = spec_aparch,out.sample = 0,solver="hybrid")
  sigma<-as.numeric(sigma(ugarchforecast(fit,n.ahead = 1)))
  quant<-(-mean(serie)-qnorm(conf_level,0,1)*sigma)
  garch_var<-rbind(garch_var,quant)
}

#t-student
# 3. GARCH-T
spec_aparch<-ugarchspec(variance.model =
    list(model="fGARCH",submodel="GARCH",garchOrder=c(p,q)),mean.model=
list(armaOrder=c(0,0),include.mean=TRUE),distribution.model = "std")
# 4. GARCH GJR T
spec_aparch<-ugarchspec(variance.model =
    list(model="fGARCH",submodel="GJRGARCH",garchOrder=c(p,q)),mean.model=
list(armaOrder=c(0,0),include.mean=TRUE),distribution.model = "std")
# 5. APARCH-T
spec_aparch<-ugarchspec(variance.model =
    list(model="fGARCH",submodel="APARCH",garchOrder=c(p,q)),mean.model=
list(armaOrder=c(0,0),include.mean=TRUE),distribution.model = "std")

for(i in (length(dataset$Return)-1600):(length(dataset$Return)-1)){
  print(i)
  serie<-dataset$Return[(i-w+1):i]
  fit<-ugarchfit(data=serie,spec = spec_aparch,out.sample = 0,solver="hybrid")
  sigma<-as.numeric(sigma(ugarchforecast(fit,n.ahead = 1)))
  fit<-fitdistr(serie,"t")
  quant<-(-mean(serie)+qt((1-conf_level),df=fit$estimate[["df"]])*sigma)
  garch_var<-rbind(garch_var,quant)
}

gev_series<-gev_model(dataset,500,conf_level,5)
par_series<-gpd_model(dataset,500,conf_level,85)
```

```
# CAVIAR #
w=500
cav_sav_series<-vector()
for(i in (length(dataset$Return)-1600):(length(dataset$Return)-1)){
  print(i)
  serie<-dataset$Return[(i-w+1):i]
  modelo<-caviarOptim(serie,model=1,pval=0.01)
  cav_sav_series<-rbind(cav_sav_series,(-1)*modelo$VarPredict)
}
cav_sav_99<-as.vector(cav_sav_series[601:1600,1])

cav_as_series<-vector()
for(i in (length(dataset$Return)-1600):(length(dataset$Return)-1)){
  print(i)
  serie<-dataset$Return[(i-w+1):i]
  modelo<-caviarOptim(serie,model=2,pval=0.01)
  cav_as_series<-rbind(cav_as_series,(-1)*modelo$VarPredict)
}
cav_as_99<-as.vector(cav_as_series[601:1600,1])

cav_garch<-vector()
for(i in (length(dataset$Return)-1600):(length(dataset$Return)-1)){
  print(i)
  serie<-dataset$Return[(i-w+1):i]
  modelo<-caviarOptim(serie,model=3,pval=0.01)
  cav_garch<-rbind(cav_garch,(-1)*modelo$VarPredict)
}
cav_garch_99<-as.vector(cav_garch[601:1600,1])

cav_adap<-vector()
for(i in (length(dataset$Return)-1600):(length(dataset$Return)-1)){
  print(i)
  serie<-dataset$Return[(i-w+1):i]
  modelo<-caviarOptim(serie,model=4,pval=0.01)
  cav_adap<-rbind(cav_adap,(-1)*modelo$VarPredict)
}
cav_adap_99<-as.vector(cav_adap[601:1600,1])
```

A.4. Backtest Code

```
VaR<-serie
alpha=0.01
BackTest = BacktestVaR(data, VaR, alpha)
vec<-vector()
for (i in 1:length(data)){
  vec[i]<-ifelse(data[i]-VaR[i]<0,1,0)
}
#number of excesses
sum(vec)
#Excess Ratio
sum(vec)/length(data)

#Unconditional Coverage
BackTest$LRuc

#Conditional Coverage
```

```
BackTest$LRcc

#DQ Test
BackTest$DQ

#QL Loss funcion
mean(BackTest$Loss$LossSeries)

#Caporin
vec<-vector()
for (i in 1:length(data)){
  vec[i]<-ifelse(data[i]-VaR[i]<0,VaR[i]-data[i],data[i]-VaR[i])
}
sum(vec)/length(data)

#Basic Loss function
vec<-vector()
for (i in 1:length(data)){
  vec[i]<-ifelse(data[i]-VaR[i]<0,(VaR[i]-data[i])^2,0)
}
sum(vec)/length(data)
```

Bibliography

[1] ABAD P., BENITO S. & LOPEZ C., *A comprehensive review of Value at Risk methodologies*, Span Rev Financ Econ., 2013, http://dx.doi.org/10.1016/j.srfe.2013.06.001.

[2] ABAD P. & BENITO S., *A detailed comparison of value at risk estimates*, Mathematics and Computers in Simulation, Vol. 94, pp. 258-276, Elsevier, 2013.

[3] ADAMKO P., SPUCHL'ÁKOVÁ E., VALÁŠKOVÁ K., *The history and ideas behind VaR*, International Conference on Applied Economics, ICOAE 2015, Kazan (Russia), 2-4 July 2015.

[4] ALEXANDER C., *Market Risk Analysis Volume IV*, John Wiley & Sons Ltd, ISBN 978-0-470-99788-8 (HB), 2008.

[5] ANGELIDIS T., BENOS A. AND DEGIANNAKIS S., *The Use of GARCH Models in VaR Estimation*, Dept. of Banking and Financial Management, University of Piraeus, 12.2003.

[6] BALKEMA A. & DE HANN L., *Residual life time at great age*, Vol. 2, pp. 792-804, Annals of Probability, 1974.

[7] BO D., *Value at Risk*, Department of Mathematics, National University of Singapore, 2001.

[8] BOFINGER E., *Non-parametric estimation of density for regularly varying distributions*, Australian Journal of Statistics, Volume 17, Issue 3, 1975, `https://doi.org/10.1111/j.1467-842X.1975.tb00957.x`.

[9] BOLLERSLEV T., *Generalized Autoregressive Conditional Heterocedasticity*, Journal of Econometrics, 31, pp. 307-327, 1986.

[10] BOLLERSLEV T., CHOU R. & KRONER K., *ARCH Modelling in Finance: A review of the Theory and Empirical Evidence*, Vol. 52, pp 5-59, Journal of Econometrics, 1992.

[11] BUCZYŃSKI M. & CHLEBUS M., *Is CAViaR model really so good in Value at Risk forecasting? Evidence from evaluation of a quality of Value-at-Risk forecasts obtained based on the: GARCH(1,1), GARCH-t(1,1), GARCH-st(1,1), QML-GARCH(1,1), CAViaR and the Historical Simulation Models depending on the stability of Financial Markets*, Working Papers 2017-29, Faculty of Economic Sciences, University of Warsaw, 2017, `https://ideas.repec.org/p/war/wpaper/2017-29.html`.

[12] BURDOVÁ D., *Modeling Conditional Quantiles of Central European Stock Market Returns*, Institute of Economic Studies, Charles University in Prague, 2014.

[13] CHEN C., *A Finite Smoothing Algorithm for Quantile Regression*,Journal of Computational and Graphical Statistics, Vol. 16, no. 1, pp. 136-164., JSTOR, JSTOR, 2007, `www.jstor.org/stable/27594233`.

[14] CHERNOZHUKOV V., *Extremal Quantile Regression*,Vol. 33, No 2, pp. 806-839, The Annals of Statistics, 2005.

[15] CHRISTOFFERSEN P. & PELLETIER D., *Backtesting Value-at-Risk: A Duration-Based Approach*, 10.2003.

[16] COLES S.G. , BAWA J., TRENNER L. & DORAZIO P., *An introduction to statistical modeling of extreme values*, Springer, Vol. 208, 2001.

[17] DAVINO C. , FURNO M. , VISTOCCO D., *Quantile Regression: Theory and Applications*, Vol. 988, Wiley Series in Probability and Statistics, John Wiley and Sons, ISBN 111997528X 9781119975281, 2013.

[18] DING Z., GRANGER C. & ENGLE R., *A long memory property of stock market returns and a new model*, Vol. 1, pp. 83-106, Journal of Empirical Finance, 1993.

[19] ENGLE R., *Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom Inflation*, Vol. 50, No 4, Econometrica, 1982.

[20] ENGLE R., *The Use of ARCH/GARCH Models in Applied Econometrics*, Journal of Economic Perspectives, Volume 15, Number 4, Pages 157-168, Fall 2001.

[21] ENGLE R. & MANGANELLI S., *CAViaR: conditional autoregressive value at risk by regression quantiles*, Journal of Business & Economic Statistics, 22, 367-381, 2004.

[22] FERNANDEZ V., *Extreme Value Theory and Value at Risk*, Revista de Análisis Económico, Vol. 18, Number 1, pp. 57-85, 06.2003.

[23] GLOSTEN L., JAGANNATHAN R. & RUNKLE D., *On the relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks*, Vol. 48, No 5, pp. 1779-1801, The Journal of Finance, 1993.

[24] GONZÁLEZ-RIVERA G., LEE T.H. & MISHRA S., *Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood*, International Institute of Forecasters, Elsevier, 2004, doi:10.1016/j.ijforecast.2003.10.003.

[25] GUMBEL E.J. , *Statistics of extremes*, New York: Columbia University Press, 1958.

[26] HAO L., & NAIMAN D. Q. , *Quantitative Applications in the Social Sciences: Quantile regression* Thousand Oaks, CA: SAGE Publications Ltd, 2007, doi: 10.4135/9781412985550.

[27] INVERTIA, *Datos de cierre diario de empresas seleccionadas: Banco Santander, Indra, Endesa y Cambio de Divida Euro-Dolar*, Data available at `https://www.invertia.com/es/portada`.

[28] JORION P., *Value at Risk. The New Benchmark for Managing Financial Risk*, Third Edition, The McGraw-Hill Companies, ISBN-13 9780071464956, 2007.

[29] KLUGMAN S. A., *Loss Models: From Data to Decisions*, John Wiley & Son, 2008.

[30] KOENKER R., *Quantile Regression*, Econometric Society Monographs, Cambridge: Cambridge University Press, 2005, doi:10.1017/CBO9780511754098.

[31] KOMSTA L., *R Package outliers: Tests for outliers*, Version 0.14, Lisence GPL-2 & GPL-3, 2011, `https://CRAN.R-project.org/package=outliers`

[32] KRAUSE A., *Exploring the Limitations of Value at Risk: How Good Is It in Practice?*, The Journal of Risk Finance, Vol. 4 Issue: 2, pp.19-28, 2003, `https://doi.org/10.1108/eb022958`.

[33] KUPIEC P. H., *Techniques for verifying the accuracy of risk measurement models*, Journal of Derivatives, 3(2): 73-84, 1995, `https://doi.org/10.3905/jod.1995.407942`.

[34] MARKOWITZ H. M. , *Portfolio Selection*, Vol. 7 (1), pp. 77-91, Journal of Finance, 1952.

[35] MORA ESCOBAR H.M., *Métodos numéricos para la estimación de parámetros en regresión cuantílica*, Revista Colombiana de Estadística, vol. 28, no. 2, pp. 221-231, 12.2005.

[36] NELSON D. B., *Conditional Heteroskedasticity in Asset Returns: a new approach*, Vol. 59, No 2, pp. 347-370, Econometrica, 1991.

[37] OLANREWAJU OLA-ADUA I., *Back-testing the VaR risk measure: an empirical study*, Faculty of Science and Technology, Institute of Mathematics and Statistics, University of Tartu, 2017.

[38] ORTEGA I., *Introduction to statistical modeling of extreme values*, Universidad de Zaragoza, 06.2017.

[39] PICKANDS J., *Statistical Inference using extreme order statistics*, Vol. 3, pp. 119-131, Annals of Statistics, 1975.

[40] R CORE TEAM, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017, `https://www.R-project.org`.

[41] ROY A.D., *Safety first and the holding of assets*, Vol. 20 (3), pp. 431-449, Econometrica, 1952.

[42] RUPPERT D., *Statistics and Finance. An introduction*, Springer Science+Business, 2004.

[43] SINGH A.K., ALLEN D.E., ROBERT P.J, *Extreme market risk and extreme value theory*, Mathematics and Computers in Simulation, 94, 310-328, Australia, 2013.

[44] TSAY R.S., *Analysis of Financial Time Series*, ISBN 9781118017098, CourseSmart, Wiley, 2010, `https://books.google.es/books?id=OKUGARAXKMwC`.

[45] TSAY R.S., *Value at Risk, Expected Shortfall & Risk Management*, Booth School of Business, University of Chicago, 2015.

[46] TUCKEY J., *The teaching of concrete mathematics*, American Mathematical Monthly, 1958, DOI 10.2307/2310294.

[47] XI Y., *Comparison of option pricing between ARMA-GARCH and GARCH-M models*, Electronic Thesis and Dissertation Repository, Western University, 2013.

[48] YANG Y., *Parameter Estimation of GARCH Model*, Optimitation, No. 1, Vol. 1, 12.2012.

Index