Libera Università degli Studi Sociali
Guido Carli

Dipartimento di Impresa e Management
Corso di Laurea Magistrale in Corporate Finance

Volatility models forecasting power: a comparison under the framework of the VaR

Candidato:
Alfredo Cinelli

Matricola:
708471

Relatore:
Prof. Paolo Porchia

Correlatore:
Marco Pirra

Luglio 2020

Anno Accademico 2019-2020

Acknowledgments

First of all I would like to thank my family for their endless support during my university journey. I want to thank also some of my colleagues for their valuable advices and discussions.
Last but not least I am grateful to some professors met during these years. Alessio Sancetta, who first introduced me to the fascinating world of econometrics. Stan Hurn, who told me about the birth of financial econometrics back in the 80s, a very precious witness of a great econometrician and one of the protagonists of that era. Paolo Santucci de Magistris and Federico Carlini, for their suggestions, advices and for pushing me diving deeper into the complex econometrics realm.
Finally, I want to thank my supervisor Paolo Porchia for believing in me, for his time and patience and for his inspiring genuine interest in learning and sharing knowledge.

Abstract

This thesis is aimed at comparing Value-at-Risk estimates from GARCH, GJR-GARCH, HAR-RV and RealizedGARCH models. The work first gives a theoretical background regarding risk management, VaR and realized measures of volatility. Then, the applied time-series models are described under a theoretical and estimation perspective. Moreover, also the testing framework used to evaluate the VaR estimates is fully discussed.

The empirical part starts by selecting the most proper specification for the ARCH-type models on the time series of IBM stock log-returns from 2002 to 2009.

The selected models turn out to be: GARCH(1,1), GJR-GARCH(1,1), HAR, LHAR and RealizedGARCH(1,1). For each of them the Gaussian and the Student-t densities specifications have been considered. The models are evaluated at the 1% Value-at-Risk by using a two step procedure. The out-of-sample period used goes from 2006 to 2009, thus it includes both a calm and a turbulent period for financial markets.

The results of the models evaluation showed that when evaluating volatility models under the VaR framework what matter is mainly the density assumed for the log-returns process. Indeed, models using a density which fits better stylized facts of returns are able to give more accurate VaR estimates than the baseline Gaussian assumption. Moreover, models using also intra-daily data give back a better VaR projections compared to traditional models. Hence, whenever possible is suggested to use high-frequency models over daily models.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The recent financial crisis in 2007-2009 demonstrated that the financial institutions risk management systems were not as adept as thought in tracking and anticipating extreme price movements during that highly volatile period.
Almost all financial institutions experienced multiple consecutive violations. Therefore, several doubts was cast and many questions were raised regarding the accuracy and the reliability of the implemented VaR models, procedures and systems.

The foundation of financial econometrics and modern risk management were laid with the seminal works of Engle in 1982 and Bollerslev in 1986, introducing ARCH-type models. However, further research achievements allowed the formal development of theory of quadratic variation as backbone of realized measures of volatility, and thus a lot of things has changed since the 80s.
Moreover, with technological developments such as cheap computational power and high-frequency databases several new volatility models have been proposed and their implementation became easier.

Therefore, the main purpose of this thesis is to assess whether the newer models using also intra-daily data are able to give back better volatility forecasts compared to the commonly used traditional models based on daily data only. However, several authors already demonstrated the superiority of realized volatility models over ARHC-type models (e.g. Koopman et al., 2005). This is the reason why in this work the assessment of the models is performed under the framework of the Value-at-Risk.

The assessment of volatility models under the VaR is a new approach initiated with the work of Giot and Laurent (2004) and is not only a valid alternative to traditional approaches (e.g. those based on measures as the MAE, MSE, etc.), but is also of practical importance due to the widespread application of the VaR as a risk measure.

The work is clearly divided into a theoretical part, developed in chapters 2 and 3 and an empirical part developed in chapter 4. In chapter 2 preliminary notions are given, thus an overview of risk management story, purpose and regulation is discussed with a focus on the formal description of the Value-at-Risk and the realized measures of volatility. Chapter 3 instead gives the theoretical background necessary to fully understand how the empirical work is performed, thus all the volatility models used are theoretically described with their estimation method. Furthermore also the testing framework adopted is discussed at the end of the chapter.

The empirical results are covered in chapter 4, comparing the traditional models GARCH and GJR-GARCH with high-frequency models such as the HAR-RV and the RealizedGARCH. Additionally, also a comparison with the widely used methods of Historical Simulation and Weighed Historical Simulation is carried out.
In particular, the comparison is based on a two step procedure. The first step is based on the evaluation of the statistical accuracy of the models by using two tests for conditional efficiency, the Christoffersen test and the Engle and Manganelli test.
The second step is instead based on the assessment of the models in terms of a loss function, thus meaningful measures of efficiency are used and in particular the loss functions of Sarma and Lopez.

The results are shown with respect to the IBM stock for a period going from 2002 to 2009, thus considering also the turbulent period of the financial crisis.
The choice of the models has a precise rationale, the GARCH is chosen because is the most used discrete time volatility model. The GJR-GARCH is used in order to allow to capture the leverage effect, thus it is an extension of the standard GARCH. The RealizedGARCH shares the same structure of the ARCH-type models, but instead of using squared returns it utilizes realized volatility hence it exploits informations in high-frequency data. Moreover, the RealizedGARCH is able to handle the leverage effects but in a different way compared to the GJR and in general to ARCH-type models. Finally the HAR-RV model is chosen because it takes into account some important stylized facts of realized volatility such as the long memory property.

Chapter 2

Preliminary Notions

In this chapter several notions are given, these are crucial in order to guide thorough the steps of this thesis. Therefore concepts as Risk, the story and the reasons behind risk management, the idea, the formalization and the computation methods for the Value-at-risk, which is the most spread risk measure, are discussed. The last section of chapter 2 is devoted to introducing ideas and a rigorous description of relatively new measures of volatility, namely measures arising from high-frequency data and theory of quadratic variation.
Most of the notions discussed in this chapter are the cornerstone of the procedures and the idea addressed in this work.

2.1 A Journey into Risk, Risk Management and Regulation

2.1.1 A Very Short Story about Risk, Finance and Risk Management

The Oxford Dictionary defines risk as 'hazard, a chance of bad consequences, loss or exposure to mischance'.
In the terms of financial risk it can be defined in one sentence, although roughly, as "the quantifiable loss more than expected". Regardless of any context, risk is tightly related to uncertainty and thus to the concept of randomness.
In the 1933 the Russian mathematician A.N. Kolmogorov gave a coincided definition of both probability and randomness (Kolmogorov, 1933).
In Kolmogorov's language a probabilistic model is defined by a triplet $(\Omega, \mathscr{F}, \mathbb{P})$, where Ω is the sample space and ω is one of its realizations, often referred to be (in economics) a state of the world. While $\mathscr{F}$ is the set of events, called σ-algebra under some good properties, and P(A) is the probability (a measure) defined on events such that A $\in \mathscr{F}$.
Finally, X can be defined as a function on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and is also known as *random variable*, usually in applications more interest is posed to the *distribution function*, (CDF), $F_X(x) = \mathbb{P}(X \leq x)$.

In this thesis the focus is on financial risk and more specifically on *market risk*. There are many types of financial risks, the first, and maybe the most important, is the market risk mentioned before, namely the risk of an adverse change in the value of a financial position due to changes in the value of the underlying components the position depends upon (e.g. bond, stock and commodity prices, exchange rate, etc.).

Crucial is also the *credit risk* that is, the risk of not receiving payments on outstanding investments such as bond or loans, due to the default of the borrower.

Another category is the *operational risk*, which is the risk coming from failed or inadequate internal processes, people, systems or external events. Usually the boundaries amongst these risk categories are not neat, nor they form a full list of all possible types of risks. Moreover there are notions of risks associated to all categories, such as the *model risk*, namely the risk of using a wrong approach in measuring a particular type of risk.

Risk management has been defined by Steinherr (1998) as 'one fo the prominent innovations of the 20th century', indeed the vast majority of the theory behind is modern. However some concepts have been around for long time.
The concept of derivative, a financial instrument whose value depends upon an underlying asset, dates back to Babylon. Dunbar (2000) interpreted a passage of the *Code of Hammurabi* (1800 B.C.) as an early evidence of the concept of option aimed at providing financial cover against crop failure.

An explicit mention of option can be found by the end of the seventeenth century in Amsterdam and in particular in the (beautiful) book *Confusion de Confusiones* wrote by Joseph de la Vega (1688). In particular in a discussion between a trader, a philosopher and a lawyer observing the activity of the Beurs of Amsterdam, actually the discussion was about what we nowadays call European call and put options, and in particular their use in risk and investment.
So it is clear how financial derivatives are not so new, moreover they appear to be an useful instrument in managing risk, rather than an invention of the capitalist devil.

Regardless of the "age" of derivatives, a theory about valuation and risk management cannot be found before the 1950s. In the 1952 a ground-breaking paper came out, written by the Nobel Prize Harry Markowitz, this paper laid the foundation of the *Modern Portfolio Theory*, which developed the concept of mapping an investment onto a risk-return diagram and via the notion of *effiecient frontier* an investor can optimize the return given the level of risk.
The decades following the Markowitz publication experienced a fast growth in risk-management and finance methodologies, including relevant works such as the Capital Asset Pricing Model (CAPM), Sharpe ratio end Arbitrage Pricing Theory (APT).
Another milestone in the current financial (risk management) theory is the *Black-Scholes-Merton formula* appeared in the 1973, it is basically a partial differential equation that, given some boundary conditions, determines the price of an European option (under rather restrictive assumptions).
This pricing formula was so ground-breaking that started the field of derivative pricing (let's think about Cox and Ross in 1976 or Harrison and Kreps in 1979), indeed in the 1997 Myron Scholes and Robert Merton were awarded of the Nobel Prize (Fischer Black died two years before). The methodologies regarding the risk neutral pricing and hedging of financial derivatives changed upside-down finance.

Almost parallel to the pure financial theory a strictly related and fascinating new discipline called financial econometrics became so important to be considered aside the econometrics itself.
The birth of financial econometrics dates back to the 1982 when an astonishing publication on a new time series model named *Autoregressive Conditional Heteroskedasticity* (ARCH)

came out by the hands of (the Nobel Prize) Robert Engle, from 1982 onward a plethora of financial econometrics and in particular volatility modelling publications came out.

2.1.2 The Basel Accords and the Value-at-Risk

The concept of regulation goes back a long way, at least to the time of the Venetian banks and insurance companies in London's coffee shops in the eighteenth century. However, most of the current regulation originated from the Basel Committee of Banking Supervision (BCBS). This entity was founded by the Governors of the G-10 Central Banks, even though the Basel Committee has no legal authority its conclusions, the so called Basel Accords, play a pivotal role in the nowadays financial and banking sector.

The *first Basel Accord* of 1988 was the first building-block towards an international minimum capital framework. It was mainly focused on credit risk, considered at that time the most important source of risk in the banking sector.
However, the first Basel Accord had a coarse approach in measuring risk an in treating derivatives.

In 1996 a crucial Amendment to Basel I Accord prescribed a *standardized model* for market risk, at the same time allowing more sophisticated techniques for larger banks, the so called *internal model* (VaR-based). Moreover from 1996 onward the important difference between banking and trading book arose.
Although the amendment gave a wider operational window to banks, the problem of coarseness for credit risk was still there, leading to too few incentives for banks to diversify credit portfolios. Because of the "expensive" regulatory capital related to particular credit positions, banks started moving business away form some market segments they perceived less attractive in terms of risk-return profile.

The *Second Basel Accord* of 2004 provided a more risk-sensitive approach in assessing credit risk in banks portfolios, the so called *internal-rating-based* methods (IRB). Other novelties of the Basel II were the introduction of the three pillars (i.e. control typologies) and the birth of regulation regarding another important source of risk, namely the operational risk.
Despite the several improvements, Basel II proved not to be adequate during the 2007-2009 Financial Crisis for many reasons. The capital of most banks was not properly quantified and the institutions abused of both on and off balance sheet leverage.
Moreover many banks relied too much on wholesale short-term funding to finance long-term illiquid assets and structured products, hence the banking system was not able to absorb the resulting systemic trading and credit losses. Finally, also the VaR-based capital regime was not shaped to measure a peculiar kind of risk, namely the one coming form the massive illiquid credit exposure in the banks portfolios. These elements created the basis for the *Third Basel Accord*.

With the Basel III the aim was to 'strengthen global capital and liquidity regulations with the goal of promoting a more resilient banking sector' and to 'improve the banking sector's ability to absorb shocks arising from financial and economic stress' (BCBS, 2009).
Therefore a capital reform took place, providing a more consistent capital base able to capture all risks and, at the same time, to control leverage and buffers. Moreover explicit liquidity standards born, a short term measure called *liquidity coverage ratio* (LCR) and a long term one as well, called *net stable funding ratio* (NSFR).

In 1993 the body G-30 published a seminal document addressing the off-balance sheet products, such as derivatives, in a systematic fashion. More or less at the same time, at JPMorgan, the Weatherstone 4.15 report asked for a brief one-day-one-page summary of the bank exposure to market risk to be delivered at the CEO office by the late afternoon (thus ”4.15”).
That was the moment in which the *Value-at-Risk* (VaR) as a key market risk measure was born and RiskMetrics set an industry-wide standard.

In a financial world in which timing and speed are crucial, the need for a quick market valuation of trading position, the so called marking-to-market, became a necessity. Moreover, markets in which many position on the same underlying arose, made the activity of managing risk based on the simple aggregation of nominal positions completely unsuitable.
These are the reasons why in 1996 the amendment to Basel I provided with a standardized model for dealing with market risk and at the same time allowing bigger banks to use VaR-based internal models.

There are many ways to compute the VaR, it is in fact possible to distinguish among a fully parametric approach called Variance-Covariance method, a Monte Carlo approach (which is parametric as well) and Historical Simulation (which is not parametric). Of course all the approaches have their pros and cons, in this thesis the focus in solely on the first category, because it allows a flexible implementation of time series volatility models.
The aforementioned approach also allows to evaluate volatility models not only in terms of forecasting power per se, but also in terms of forecasts under a VaR testing framework.

In general when talking about VaR, the focus is on the so called *Profit and Losses* (P&L) of a financial position over time thus make sense to use a time series notation, let $\{L_t\}_{t\in\mathbb{N}}$ be the sequence of losses over time defined as

$$L_t := L_{[(t-1)\Delta, t\Delta]} = (V_t - V_{t-1}) \tag{2.1}$$

where V_t is the value of the financial position at generic time t, and Δ is the time horizon of the loss, usually in practical applications Δ is set to be daily, which means $\Delta \approx \frac{1}{250}$, where 250 trading days are by convention assumed in one year.

It worth stating that V_t is modelled as function of time and a k-dimensional random vector $\mathbf{Z}_t = [Z_{t,1} \dots Z_{t,k}]'$ of *state variables*, hence

$$V_t = f(t, \mathbf{Z}_t) \tag{2.2}$$

for a measurable function $f : \mathbb{R}_+ \times \mathbb{R}^k \to \mathbb{R}$. Of course the state variables have to be measurable so that $\mathbf{Z}_t$ is available in t. Moreover the choice of the variables depends upon the assets of which the portfolio at hand is made of.
Usually, as in this thesis, the state variables are logarithmic prices of stocks/stock indices, but they can also be logarithmic exchange rates, yields, etc. The formulation in equation 2.2 is called *mapping of risk*.

It is of more interest to deal with the time series of changes in the value of the state variable, namely the *risk factors* $\{\mathbf{X}_t\}_{t\in\mathbb{N}}$, where $\mathbf{X}_t := \mathbf{Z}_t - \mathbf{Z}_{t-1}$. Another important concept is the loss operator, $l_{[t]} : \mathbb{R}^k \to \mathbb{R}$ which is a function able to map risk factors into losses, thus

$$l_{[t]}(\mathbf{x}) = f(t+1, \mathbf{Z}_t + \mathbf{x}) - f(t, \mathbf{Z}_t), \quad \mathbf{x} \in \mathbb{R}^k \tag{2.3}$$

where by construction $L_{t+1} = l_{[t]}(\mathbf{X}_{t+1})$.
When f is differentiable the focus is usually on a linear (i.e. first-order) approximation labeled L^{Δ}_{t+1} of the loss, which is expressed based on the Taylor expansion as

$$L^{\Delta}_{t+1} = f_t(t, \mathbf{Z}_t) + \sum_{i=1}^{k} f_{Z_i}(t, \mathbf{Z}_t)\mathbf{X}_{t+1} \quad where \tag{2.4}$$

$$f_t(t, \mathbf{Z}_t) := \frac{\partial f(t, \mathbf{Z}_t)}{\partial t}$$

$$f_{Z_i}(t, \mathbf{Z}_t) := \frac{\partial f(t, \mathbf{Z}_t)}{Z_i}$$

thus as stated in equation (2.4) the subscripts to f represent the partial derivatives. In the same way is possible to compute a linear approximation for the linear operator.
The first-order approximation is useful because allows for the representation of the loss as a *linear function* of the risk factors. Of course, the approximation is more precise the smaller are the risk factors and/or the more the value of the portfolio is linear in the risk factors (i.e. $f(\cdot)$ has small second-order derivatives).

The applications of the following chapter are based on the time series of losses, in which each observation is computed as the logarithmic return of the asset at hand (stock or stock index) because the positions are in one instrument at a time, then $L_t = r_t$ and the risk factors are the log-returns themselves. For risk-management and finance purposes is crucial the difference between the conditional and unconditional distribution of losses (they are by nature random variables).
The above difference is strongly related to the time series of risk factors, for the purposes of this thesis $\{\mathbf{X}_t\}_{t\in\mathbb{N}}$ and $\{\mathbf{L}_t\}_{t\in\mathbb{N}}$ become $\{r_t\}_{t\in\mathbb{N}}$, of course the $\{r_t\}$ series is deemed to be stationary and is proved to be via several statistical tests (e.g. Dickey-Fuller, Phillips-Perron and KPSS). Going back to the difference between the conditional and unconditional distribution, is crucial to introduce the concept of *information set* labeled as $\mathscr{F}_t$, which is a σ-algebra containing all the informations available up to time t, more formally $\mathscr{F}_t = \sigma([r_s : s \leq t])$.
Therefore, as in the majority of applications the emphasis is posed to the *conditional distribution* of $\{r_t\}_{t\in\mathbb{N}}$, given the information set $\mathscr{F}_{t-1}$, also labeled as $F_{r_t|\mathscr{F}_{t-1}}$.
It worth pointing out that $F_{r_t|\mathscr{F}_{t-1}}$ is not equal to the (stationary) unconditional distribution F_r. A representative example are the GARCH models, in which the unconditional variance (σ) is constant (under stationarity condition) whilst the conditional variance (σ_t) is time varying and depends upon a bunch of variables depending on the functional form of the variance equation. Of course whenever $r_t \sim i.i.d.$ then $F_{r_t|\mathscr{F}_{t-1}} = F_r$.

However for a generic portfolio, the conditional loss distribution $F_{L_{t+1}|\mathscr{F}_t}$ is specified as the loss operator distribution under $F_{\mathbf{X}_{t+1}|\mathscr{F}_t}$, where as stated above $\mathbf{X}_{t+1}$ is the vector of risk factors. Thus for $l \in \mathbb{R}$,

$$F_{L_{t+1}|\mathscr{F}_t}(l) = \mathbb{P}(l_{[t]}(\mathbf{X}_{t+1}) \leq l \mid \mathscr{F}_t) = \mathbb{P}(L_{t+1} \leq l \mid \mathscr{F}_t) \tag{2.5}$$

Whereas the unconditional loss distribution of the generic portfolio $F_{L_{t+1}}$ is the distribution of $l_{[t]}(\cdot)$ under the stationary distribution of the risk factors, $F_{\mathbf{X}}$.

The Value-at-Risk is a modern risk measure based on the portfolio loss distribution, this is the reason why losses are of paramount importance in risk management.

Moreover the concept of loss distribution allows all levels of aggregation, from a position in a single instrument (as in the forthcoming applications) to a portfolio composed of many instruments. Finally, the loss distribution reflects the beneficial effect of diversification, and allows for comparison across portfolios. For instance, is possible to compare a portfolio of equity derivatives with a book of fixed-income, as long as the time horizon Δ taken into consideration is the same.
The two main practical drawbacks of the loss distribution approach are first that the distribution estimation is based on past data so it is backward-looking, and second that even under stationarity the estimation of the loss distribution for a large portfolio is not so trivial. These are the reasons why, when using a loss distribution-based approach is convenient to use in conjunction forward-looking measures (e.g. implied volatility) and multiple hypothetical scenarios.

The VaR can be defined as the maximum potential loss that can arise over an arbitrary time period and for an arbitrary confidence level $\alpha \in]0,1[$, therefore given the loss (L) which is the random variable of interest and its distribution function F_L, then

$$VaR^{\alpha}(L) := \inf[l \in \mathbb{R} : \mathbb{P}(L > l) \leq (1-\alpha)] = \inf[l \in \mathbb{R} : F_L(l) \geq \alpha]. \tag{2.6}$$

Hence in other terms the VaR (through the work VaR^{α} and $VaR(\alpha)$ are used indifferently) is the *quantile* of the loss distribution, usually in practice α is set $\alpha = 0.95$ or $\alpha = 0.99$ while the time horizon Δ is 1 or 10 days. Clearly the VaR depends on the choice of both α and Δ.

The biggest drawback of the VaR is that given the confidence level α, there are no informations about the severity of losses exceeding the level itself, this is the reason why another crucial measure called *Expected Shortfall* has been introduced (also the non-subadditivity of VaR is another relevant drawback).
The figure 2.1 shows the difference amongst the measure described above in terms of probability density. Before diving deeper into the VaR and ES, is crucial to introduce the concept of

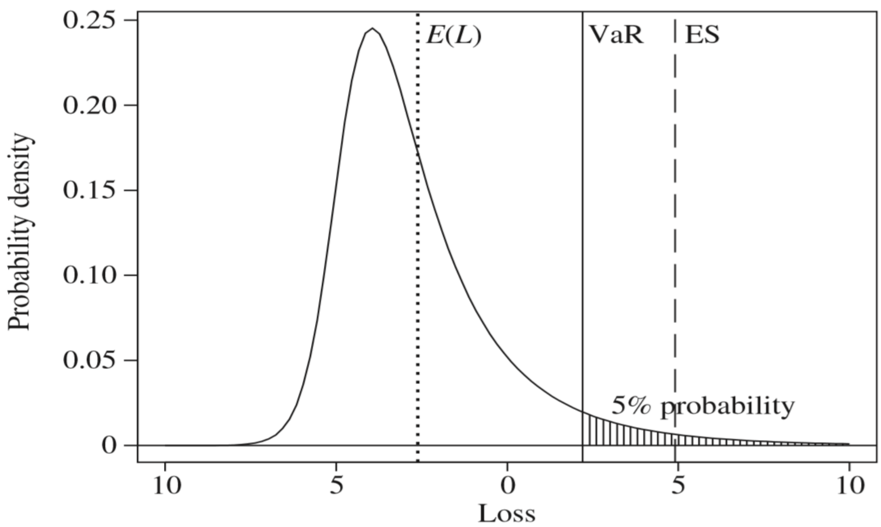


Figure 2.1: Example of loss distribution with VaR and ES.

generalized inverse function. Given an increasing function $T : \mathbb{R} \to \mathbb{R}$, its *generalized inverse* denoted as $T^{\leftarrow}(y)$ is formally defined as

$$T^{\leftarrow}(y) := \inf[x \in \mathbb{R} : T(x) \geq y]. \tag{2.7}$$

This concept can be automatically extended to a generic distribution function F, its generalized inverse is also known as *quantile function*.
Given an $\alpha \in]0,1[$, the α-quantile of F is

$$q_\alpha(F) := F^{\leftarrow}(\alpha) = \inf[x \in \mathbb{R} : F(x) \geq \alpha]. \tag{2.8}$$

For a random variable X the following alternative notation is also used $q_\alpha(X) := q_\alpha(F)$. Whenever F is both strictly increasing and continuous (as usually is the CDF), the quantile function is simply the ordinary inverse $q_\alpha(x) := q_\alpha(F) = F^{-1}(\alpha)$.
The above definitions are crucial for the introduction of both the *delta-normal* and *delta-student* formulation of the VaR, namely the ones based respectively on the assumption of Gaussian and Student-t distribution for the conditional density of log-returns (as the models in the next chapter). Whenever $L \sim N(\mu, \sigma^2)$ and given the confidence level $\alpha \in]0,1[$ the VaR^α equation is

$$VaR^\alpha = \mu - \sigma\Phi^{-1} \quad where \tag{2.9}$$

$$\mu := \mathbb{E}[L] \quad and \quad \sigma^2 := \mathrm{Var}(L)$$

where μ and σ^2 are respectively the *unconditional mean* and the *unconditional variance*.
The proof to equation 2.9 is first based on the strictly increasing behaviour of $\Phi(\cdot)$, then it has to be shown only that $\Phi(VaR^\alpha) = \alpha$. Now

$$\mathbb{P}(L \leq VaR^\alpha) = \mathbb{P}\left(\frac{L-\mu}{\sigma} \leq \Phi^{-1}(\alpha)\right) = \Phi(\Phi^{-1}(\alpha)) = \alpha. \tag{2.10}$$

The same result can be derived for any location-scale family [1].
An useful result is derived for the Generalized Student-t density (the Student-t distribution expressed with a location parameter and a non-negative scale parameter). Assuming that $L \sim t_v(\mu, \sigma^2)$, then $\mathbb{E}[L] = \mu$ if $v > 1$ where μ is actually the location parameter and $\mathrm{Var}(L) = \left(\frac{v}{v-2}\right)\sigma^2$, whenever $v > 2$ otherwise no second moment exists and noting that σ^2 is not the variance of the distribution but the scale parameter. Then

$$VaR^\alpha = \mu - \lambda t_v^{-1}(\alpha) \quad where \;\; \lambda := \sigma\sqrt{\frac{v}{v-2}} \tag{2.11}$$

where t_v is the Generalized Student-t distribution and v denotes its number of degrees of freedom, clearly whenever L follows a Generalized Standard Student-t distribution $\lambda = \sigma$.
The same results derived above hold also in a conditional setting (and for a loss simply identified as the log-return when dealing with only one asset at a time as in the next chapter)

[1] A location-scale family allows given a standard r.v. $X \sim D_X(0,1)$ to express another r.v. as function of X, namely $Y := \mu + \sigma X$ where μ is the mean of Y and the location parameter, whilst σ is the standard deviation of Y and the non-negative scale parameter. Then given the density of X is possible to derive in a standard fashion the density of Y (based on the Jacobian of the transform) as $f_Y(y) = f_X(x)\frac{1}{\sigma}$ where $x = \frac{y-\mu}{\sigma}$

thus whenever $L_t \mid \mathscr{F}_{t-1} \sim N(\mu_t, \sigma_t)$ then the delta-normal VaR formula becomes

$$VaR_t^{\alpha} = \mu_t - \sigma_t \Phi^{-1}(\alpha) \quad where \tag{2.12}$$

$$\mu_t := \mathbb{E}[L_t \mid \mathscr{F}_{t-1}] \quad and \quad \sigma_t^2 := \mathrm{Var}(L_t \mid \mathscr{F}_{t-1})$$

where μ_t and σ_t^2 are the conditional mean and conditional variance respectively.
The same holds true also when the log-return conditional density is assumed to be a Standard Generalized Student-t distribution, namely $L_t \mid \mathscr{F}_{t-1} \sim t_v(\mu_t, \sigma_t^2)$, then

$$VaR_t^{\alpha} = \mu_t - \sigma_t t_v^{-1}(\alpha) \quad given \ \lambda_t = \sigma_t \tag{2.13}$$

As stated above a drawback of the VaR measure is the absence of informations whenever the loss exceeds the threshold. Moreover the VaR in non-subbaditive, which means that the VaR of a merged portfolio is not necessarily bounded above by the sum of the VaRs of the individual portfolios. Thus the last drawback goes against the crucial evidence of diversification. These are the two main reasons why another important measure has born, the *Expected Shortfall* (ES).

The measure of Expected Shortfall is thightly related to VaR, the shift form unconditional to conditional setting is feasible in a similar fashion to the latter measure. Theoretically the ES is superior to the Value-at-Risk in the sense that the former overcomes the problem of non-subbadditivity.
The ES often denoted as *Conditional Value-at-Risk* (CVaR), whenever the loss L itself is integrable with a continuous distribution F_L and given a confidence level $\alpha \in]0,1[$, can be expressed as

$$ES^{\alpha}(L) = \frac{\mathbb{E}[L; L \geq q_{\alpha}(L)]}{1-\alpha} = \mathbb{E}[L \mid L \geq VaR^{\alpha}]. \tag{2.14}$$

Thus the ES as defined in equation 2.14 can be interpreted as the expected loss arising whenever the VaR is exceed (for sake of space no proof is given to equation 2.14).
Another useful way to express the Expected Shortfall is to use the Acerbi's integral formula, (its proof can be found in Acerbi and Tasche, 2002),

$$ES^{\alpha}(L) = \frac{1}{1-\alpha}\int_{\alpha}^{1} q_{\beta}(F_L)d\beta = \frac{1}{1-\alpha}\int_{\alpha}^{1} VaR^{\beta}(L)d\beta, \tag{2.15}$$

where as stated before $q_{\beta}(F_L) := F_L^{\leftarrow}(\beta)$ is the quantile function of F_L.
From the ES definition given in equation 2.15 the risk measure can be interpreted as the average VaR^{β} for all levels $\beta \in [\alpha, 1]$.

In the same way as for the VaR is possible to derive a compact formula for the ES when the loss L is distributed according to a Normal distribution, namely when $L \sim N(\mu, \sigma^2)$ and given $\alpha \in]0,1[$, then

$$ES^{\alpha} = \mu + \sigma \frac{\Big(\phi(\Phi^{-1}(\alpha))\Big)}{1-\alpha}, \tag{2.16}$$

where $\Phi^{1-}(\cdot)$ is the inverse CDF and $\phi(\cdot)$ is the density of the standard normal distribution.

The proof can be given noting that

$$ES^{\alpha} = \mu + \sigma \mathbb{E}\left(\frac{L-\mu}{\sigma} \mid \frac{L-\mu}{\sigma} \leq q_{\alpha}\left(\frac{L-\mu}{\sigma}\right)\right).$$

Thus is only needed to compute the Expected Shortfall for the standardized loss $L^* := \frac{(L-\mu)}{\sigma}$, therefore

$$ES^{\alpha}(L^*) = \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^{\infty} l\phi(l)dl = \frac{1}{1-\alpha}[-\phi(l)]_{\Phi^{-1}(\alpha)}^{\infty} = \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

In the same fashion is possible to get a compact formula to express the ES for any location-scale family, by simply computing the ES of the standardized loss.
When $L \sim t_{\nu}(\mu, \sigma^2)$, then the compact formula for the ES in case of Student-t distribution is

$$ES^{\alpha}(L) = \mu + \sigma ES^{\alpha}(L^*) \quad where$$

$$ES^{\alpha}(L^*) := \left[\left(\frac{g_{\nu}(t_{\nu}^{-1}(\alpha)}{1-\alpha}\right)\left(\frac{\nu + (t_{\nu}^{-1}(\alpha))^2}{\nu - 1}\right)\right].$$

It worth noting that g_{ν} is the density of the standard Student-t distribution, whilst t_{ν} is its distribution function.

There are three main approaches to compute the Value-at-Risk.
First the variance-covariance method (the one used in this work), which is a parametric method because needs the estimation of some parameters. It can be turned into a conditional or unconditional method. The vector of risk-factors $\mathbf{X}_{t+1}$ is assumed to follow, either conditionally or unconditionally, a multivariate Gaussian distribution $\mathbf{X}_{t+1} \sim MN(\mu, \Sigma)$, where μ and Σ are respectively the mean vector and the variance-covariance matrix. Then by exploiting the assumption that the linearized loss is a decent proxy of the actual loss, the problem shifts to $L_{t+1}^{\Delta} := l_{[t]}^{\Delta}(\mathbf{X}_{t+1})$ as specified in equation 2.4. The linearized loss operator is a function with a standard structure of the type

$$l_{[t]}^{\Delta}(\mathbf{X}_t) = -(c_t + \mathbf{b}_t'\mathbf{X}_t) \tag{2.17}$$

where c_t is a constant and $\mathbf{b}_t$ is a constant vector, both known at time t. For instance in a stock portfolio $l_{[t]}^{\Delta}(\mathbf{X}_t) = -V_t\mathbf{w}_t'\mathbf{X}_t$ where V_t is the value of the portfolio, $\mathbf{w}_t$ the vector of portfolios weights and $\mathbf{X}_t$ the vector of returns.
A crucial property of the multivariate normal is that a linear function of $\mathbf{X}_{t+1}$ has an univariate normal distribution. From rules on the mean and variance of a linear combination of a random vector,

$$L_{t+1}^{\Delta} = l_{[t]}^{\Delta}(\mathbf{X}_{t+1}) \sim N(-c_t - \mathbf{b}_t'\mu, \mathbf{b}_t'\Sigma\mathbf{b}_t). \tag{2.18}$$

Hence for this loss distribution the VaR can be computed via equation 2.9 or 2.12 depending whether operating in an unconditional or a conditional setting. The parameters Σ and μ in the unconditional setting can be easily computed via sample moments, by using the implicit (and more than reasonable) assumption that risk-factors data come from a stationary stochastic process. The approach used in this thesis is the conditional one, in which (in the applications of chapter 4 it is an univariate series because only the log-return is considered as risk-factor)

data are treated as realizations of a time series $\mathbf{X}_{t+1} \mid \mathscr{F}_t \sim MN(\mu_{t+1}, \Sigma_{t+1})$, where μ_{t+1} and Σ_{t+1} are respectively the conditional mean vector and the conditional variance-covariance matrix given the information set at time t. The estimates of μ_{t+1} and Σ_{t+1} can be obtained via time series models such as GARCH, HAR-RV, etc. (see chapter 3).
The variance-covariance method provides an analytical solution to the risk-measurement problem, but paying at the same time two simplifying assumptions. First, the linearization may not be a good approximation of the relationship between the true loss and the risk-factors. The second one regards the distribution assumption, indeed in order to get the result in equation 2.18 the multivariate distributions assumed have to be closed under linear operators, the multivariate Gaussian is of course one of these distributions, but it does not take into account some facts such as the fat tail phenomenon. However other multivariate distribution are closed under linear operators and better fits some facts, for instance the multivariate Student-t or the multivariate generalized hyperbolic distributions, actually this problem does not occur when dealing with one risk-factor only because univariate distributions are used.

Another method is the Historical Simulation (HS), which is still today a benchmark to compete against. The HS is based on estimating the distribution of the loss operator under the empirical distribution of data of the risk-factors $\mathbf{X}_t, \mathbf{X}_{t+1}, ..., \mathbf{X}_T$. Therefore an univariate dataset is constructed by applying the linear operator to each historical observation of the risk-factors, thus is possible to get simulated losses

$$\{\widetilde{L}_s = l_{[t]}(\mathbf{X}_s) : s = t, t+1, ..., T\}$$

The VaR(α) is given by the percentile of the sequence of simulated losses sorted in ascending order, namely

$$VaR(\alpha)_{t+1} = Percentile(\{\widetilde{L}_s\}_{s=t}^{T},\ \alpha).$$

Whenever the VaR falls in between two observation an easy linear interpolation can be used. This approach is an unconditional method, assuming that the process of risk factors is stationary (e.g. log-returns) with distribution $F_{\mathbf{X}}$, then the empirical distribution of the data is a consistent estimator of $F_{\mathbf{X}}$. The main advantages of the Historical Simulation are for sure its simplicity especially in terms of implementation because it reduces risk-measure estimation problems to one-dimensional problem, but also its non-parametric (i.e. model free) nature.
The first advantage means that no relatively complex estimation technique has to be employed (e.g. Maximum likelihood, GMM, etc.), the second advantage is crucial because it implies that no parametric model (e.g. GARCH) needs to be used, thus no assumptions beyond the stationarity of the risk factor need to be made (fat tails and extreme event are considered as long as included in the dataset).
However, its non-parametric nature is at the same time a drawback, because there is non clear cut rule on choosing the length ($M = T - t$) of the data sample, but at the same time the length matters a lot in computing the Value-at-Risk, as it can be see in figure 2.2. In choosing the dataset length it has to be noted that a too small sample might lead to not considering large losses, and if M is too large then small weights are attached to more recent observations (i.e. the most relevant). Thus the choice of the length of the sample is an ad hoc choice depending on the case at hand.
Furthermore, the lack of a dynamics in the HS causes to overlook some stylized facts on return dependence, such as variance clustering which leads the HS-VaR to react too slowly

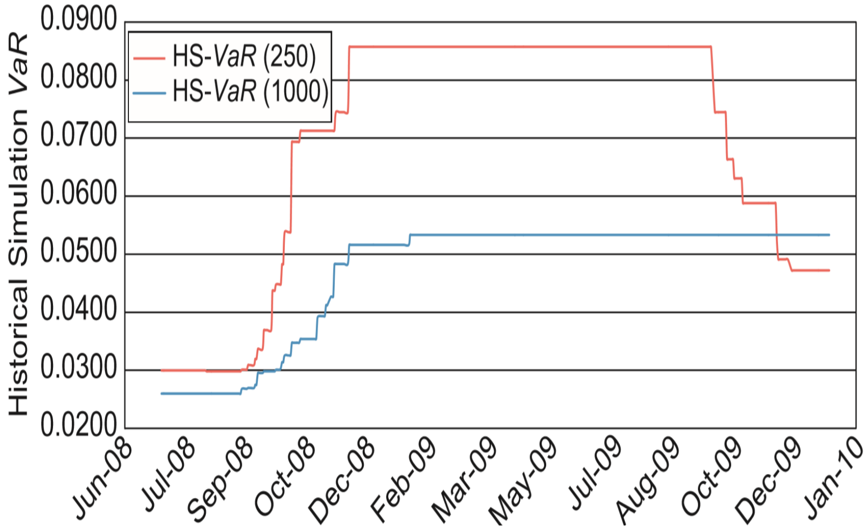


Figure 2.2: S&P500 VaRs from Historical Simulation using 250 and 1,000 return days in a rolling window: July 1, 2008–December 31, 2009.

to changes in the market environment. Finally its unconditional nature could be seen as a further weakness, because conditional approaches are more relevant for day-to-day market risk management. However there are some extension that can be used in order to overcome some problems of the plain Historical Simulation, such as the Weighted Historical Simulation (WHS) or the Filtered Historical Simulation (FHS).

The WHS is aimed at attaching a weight based on a decay factor to each observation in the series of risk factors, in order to give more weight to recent observation and vice versa. The WHS reacts faster to changes in market volatility than the plain HS, the decay factor η is usually a value between 0.95 and 0.99 and for each observation the weight is $\omega_s = \{\eta^{s-1}\frac{1-\eta}{(1-\eta)^m}\}_{s=t}^{T}$.

The last but not least method for VaR computation is the Monte Carlo method [2] which is still parametric and is based on simulating many paths for the risk-factors assuming proper stochastic processes (that is the reason why it falls in the parametric methods) in order to get simulated losses. Of course the Monte Carlo approach is computational burden, especially for large portfolios in which there are many derivatives that cannot be evaluated via closed form formulas, this requires variance reduction techniques (e.g. antithetic variates, control variates, stratification, etc.). Moreover is important to remark that the delta-gamma approximations rely on the assumption that the portfolio value changes in a linear or quadratic fashion with risk factors, thus this assumption is sometimes a limit to their accuracy. However, it worth noting that the Monte Carlo approach can be applied on any model of risk factors, therefore in some cases the MC is preferred over the delta-gamma method, for instance when the latter is not accurate (i.e. portfolio mappings highly non-linear), or when the delta-gamma is not feasible, for instance when dealing with a time horizon greater than just one day.

[2] The first version of Monte Carlo traced back to the Buffon's needle experiment to compute π by randomly dropping needles on a floor made of parallel and equidistant strips. Another relevant first application was made by Enrico Fermi in the 1930s to study the diffusion of neutrons. The modern version of MCMC is mainly attributed to Stanislaw Ulam and in part also to John von Neumann, it was developed during the Manhattan Project for the nuclear bomb in the Los Alamos laboratories. The name Monte Carlo is instead based on the casino in the homonymous town in Monaco, this name was suggested by Nicholas Metropolis (Metropolis is indeed not by chance a name associated, with Hastings, to an MCMC method for generating random samples).

2.2 Notions about high frequency data

The birth of high-frequency trading in the beginning of the 21th century has started horizons to study in a different way financial markets. The access to *high-frequency data* allows analysts, traders and researchers to study the dynamics of market micro-structure and the impact of agents on it.
Moreover with the exponential growth of both storage capacity and computing power, is now possible to look at the financial markets form a slightly different angle.

It worth distinguish amongst four types of datasets, according to the level of detail. First the *trade data*, which contains the time stamp of trades, the volume of shares traded and the price at which the trades took place.
The second level is the *trade and quote*, which includes also the best bid/ask quote and the trade direction up to identification rules. Then there is the *fixed level order book data*, which gives informations about limit order activities.
The last category of dataset also includes messages on all limit order activities, allowing to recreate the trading flow and the limit order book at any point in time.

It is important to remark that on one hand high-frequency data allows to explore new frontiers in empirical works, but on the other hand entry errors are inevitable and often encountered in these types of dataset.
Therefore is crucial to ensure data to be as free as possible form errors, for this purpose a cleaning procedure was recommended by Barnodrff-Nielsen, Hansen, Lunde and Shepard (2008).

Another crucial remark is that tick-by-tick data are not sampled at evenly spaced intervals, therefore is necessary to use interpolations techniques, for instance Hansen and Lunde (2006) suggested to replace missing observations with the last recorded prices. Of course the more liquid is the asset the lower is the proportion of non-observable prices replaced with the last observation itself.
Figure 2.3 shows in Panel (a) 6326 irregularly spaced logarithmic prices for the IBM stock on 3 January 2006, whereas Panel (b) shows 23400 log-prices interpolating with the Hansen and Lunde suggested approach. The availability of high-frequency data allows to compute several

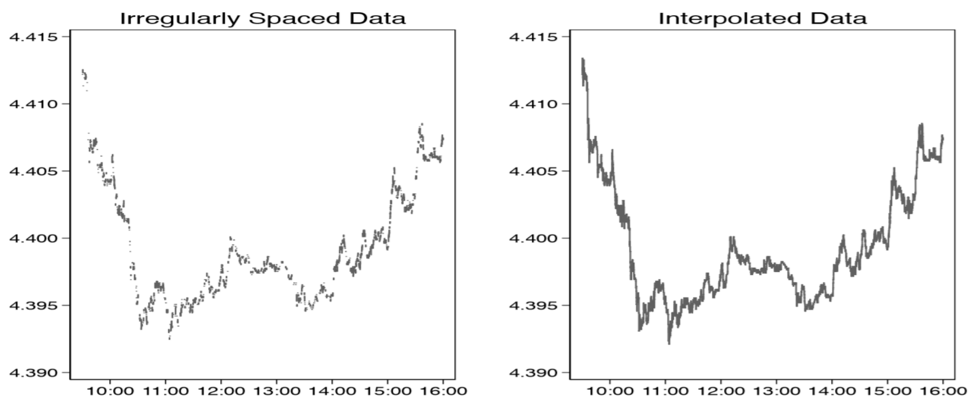


Figure 2.3: Irregularly spaced and interpolated log-prices for IBM on 3 January 2006.

efficient non-parametric estimator for the variance of an asset. However it is first useful to specify some concepts.

For an univariate risky log-price process p_t defined on a filtered probability space $(\Omega, (\mathscr{F})_{t\in[0,T]}, \mathscr{F}, \mathbb{P})$ and assuming also that $(p_t)_{t\in[0,T]}$ is a real valued process, Protter in 1992 showed that the continuously compounded return associated can be uniquely represented as

$$r_t := p_t - p_{t-1} = \mu_t + M_t = \mu_t + M_t^C + M_t^J \tag{2.19}$$

therefore the instantaneous log-return can be decomposed into an expected return component μ_t (a finite variation process), and a martingale innovation [3] M_t, where M_t^C and M_t^J are respectively an infinite variation local martingale and a compensated jump martingale. Moreover in the most general setting and following the notation of Corsi, Pirino and Renò (2008) the log-price $(X_t)_{t\in[0,T]}$, (or p_t), is assumed to evolve over time according to a precise stochastic differential equation (SDE) of the following form (this is an Ito's Process),

$$dX_t = \mu_t dt + \sigma_t dW_t + k_t dq_t \quad t \in [0,T]. \tag{2.20}$$

Where μ_t is the instantaneous conditional expected return, σ_t^2 is the instantaneous variance (and σ_t is càdlàg), W_t is a standard Brownian motion [4] able to capture the shocks to log-returns over time with its stochastic differential $dW_t \sim N(0, dt)$, k_t is the so called size of the jump and q_t is a Poisson process which serves the role of counting, thus its value is 1 if there is a jump in t and 0 otherwise. It worth noting that when in the above SDE there are no jumps and the drift and the continuous volatility are time invariant, then the price process is the one assume under the Black-Scholes-Merton framework. Figure 2.4 shows 100 simulated paths of a Brownian motion, this figure is recurrent in many fields, for instance when simulating stock prices in Black-Scholes context (under the assumption that prices follow a Geometric Brownian motion).

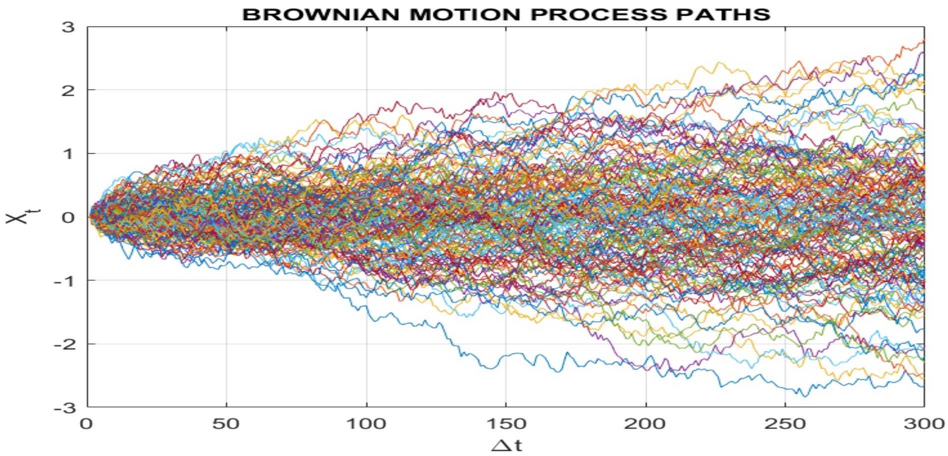


Figure 2.4: The one hundred simulated paths with three hundred observations each one from a standard Brownian motion

[3]A martingale is a particular stochastic process X_t such that $\mathbb{E}[|X_t|] < \infty$, and $\mathbb{E}[X_t \mid (X_s)_{s\leq u}] = X_u, \forall u \leq t$, hence the best prediction is its current value. The concept of martingale was introduced in 1934 by the great French mathematician Paul Lévy.

[4]The name Brownian motion comes form the botanist Robert Brown who first described the phenomenon in 1827, after him in 1905 the great(est) physicist Albert Einstein published a paper using the concept of Brownian motion to describe the motion of pollen in the water. However with the mathematician Norbert Wiener finally a formal mathematical description of the Brownian motion, or not by chance, Wiener process was given. It worth pointing out that the Brownian motion is an example of continuous time martingale.

Another relevant topic for the description of realized measures is the concept of *Quadratic Variation* (QV), given a fixed time window T, the total quadratic variation of $(X_t)_{t\in[0,T]}$ is

$$[X]_t^{t+T} := X_{t+T}^2 - X_t^2 - 2\int_t^{t+T} X_{s^-} dX_s \tag{2.21}$$

where T represents the time interval, usually a day, and the integral is actually a stochastic integral of the adapted càdlàg process X_{s^-}.
Under the assumption of jump diffusion model in equation 2.20, the QV can be decomposed into a continuous component and a discontinuous component as

$$[X]_t^{t+T} = [X^C]_t^{t+T} + [X^J]_t^{t+T} \tag{2.22}$$

where the continuous component $[X^C]_t^{t+T} := \int_t^{t+T} \sigma_s^2 ds$ is the *Integrated Variance* (IV), and the discontinuous component $[X^J]_t^{t+T} := \sum_{t\in(0,1]} k_t^2$ is the so called jump contribution.

The realized variance [5] (RV) is a non-parametric estimator of the above quadratic variation (often defined in the financial context as *notional volatility*), for this purpose is necessary to track prices at intra-daily frequencies and divide the trading day into n subintervals, assuming any reasonable time interval $[t, t+T]$, then

$$RV^{(n)}(X)_t = \sum_{i=1}^{n} (\Delta_i X)^2$$

where for this thesis applications and in general for financial applications $X_t := p_t$ is the log-price, therefore by construction $\Delta X_t := r_t$ is the log-return of the asset. For sake of simplicity in the notation usually the daily realized variance is also defined as

$$RV_t^{(n)} = \sum_{i=1}^{n} r_{t,i}^2. \tag{2.23}$$

A useful property of the realized variance is being a consistent estimator of the quadratic variation, namely $RV_t(X)_t \xrightarrow{p} [X]_t^{t+T}$, actually the RV is a consistent and an unbiased estimator of the daily variance whenever returns have zero mean and are uncorrelated over time (see Andersen et al., 2001).

In the applications of the next chapter the equation 2.20 is assumed to have no jumps and moreover is assumed a market with non frictions, this because the interest is posed on the continuous component only, therefore the necessity of a consistent estimator of the IV only. Therefore the asset logarithmic price process is a semi-martingale defined by the following SDE (now using a more familiar notation),

$$dp_t = \mu_t dt + \sigma_t dW_t \quad where \;\; t \in [0,T] \;\; and \;\; p_t := X_t \tag{2.24}$$

where the components are the same as defined in equation 2.20 and noting that σ_t is strictly positive and square integrable, namely $\mathbb{E}[\int_0^t \sigma^2(s) ds] < \infty$.
Then based on the Protter decomposition given in equation 2.19 is possible to further decom-

[5] The concept of realized variance is not so new, Robert Merton mentioned it in the 1980. However, its introduction into modern financial econometrics date to the late 1990s (see Andersen and Bollerslev, 1998).

pose the daily log-return as

$$r_t := p_t - p_{t-1} = \int_{t-1}^{t} \mu_s ds + \int_{t-1}^{t} \sigma_s dW_s \tag{2.25}$$

where, given that the quadratic variation of a finite variation process is zero, the QV is equal to the IV only, namely

$$QV_t = IV_t := \int_{t-1}^{t} \sigma_s^2 ds.$$

Therefore under this setting, which means under the assumptions of no jumps and no microstructure noise, the RV is a consistent non-parametric estimator of the integrated variance, indeed as shown by Barndorff-Nielsen and Shepard (2002) the following is true $RV_t \xrightarrow{p} IV_t$ as $n \to \infty$. Moreover $\sqrt{n}(RV_t^{(n)} - IV_t) \xrightarrow{d} N(0, 2IQ_t)$, where $IQ_t := \int_{t-1}^{t} \sigma_s^4 ds$ is the so called *Integrated Quarticity*.

However in the market there are several microstructure noises, such as the bid/ask bounce, decimalization effects, etc. Therefore the observed price is equal to the sum of the efficient unobserved price p_t^* and an independent mean zero microstructure noise v_t,

$$p_t = p_t^* + v_t$$

as a consequence in the same fashion the log-return becomes $r_t = r_t^* + u_t$.
The noise clearly affects the RV in the sense that the realized measure becomes an upward biased estimator of the RV computed with the latent efficient prices, namely $RV_t^{(n)} \approx RV_t^{(n,*)} + n\sigma_u^2$ where σ_u^2 is the variance of u_t. Moreover is crucial to note that the bias is proportional to the sampling frequency n, furthermore asymptotically RV converges to a constant proportional to the variance of the microstructure noise itself. In order to try to overcome this problem Andersen, Bollerslev, Diebold and Labys (2001) proposed a conservative cut-off point by choosing 5-minutes intervals to perform the computation.

It is important to remark that the cut-off of 5-minutes is an empirical choice and is actually a compromise, a trade-off between the bias which increases as the frequency (n) increases and the convergence which of course increases with the frequency.
Figure 2.5 shows as the $RV_t^{(n)}$ changes with the chosen time frequency. There are other approaches in order to overcome the problem of microstructure bias, for instance the two-scale estimator proposed by Zhang, Mykland and Ait-Sahalia (2005) which is based on using several realized variance estimators computed at different sparse frequencies, and then computing an average of them adjusting for the bias by using the estimator at 1-second frequency.

Going back to the most general description of the log-price process given in equation 2.20 it is crucial to remark that the RV is a consistent estimator of the quadratic variation, thus of both the continuous and discontinuous component. Therefore at a daily time horizon the relationship is clearly

$$RV_t^n \xrightarrow{p} QV_t = \int_{t-1}^{t} \sigma_s^2 ds + \sum_{t \in (0,t]} k_t^2 \tag{2.26}$$

where the component $\sum_{t \in (0,1]} k_t^2$ is the jump contribution.
Therefore when the log-price follows a jump diffusion process, then the RV is no longer a consistent estimator of the integrated variance which is the continuous part, namely the part

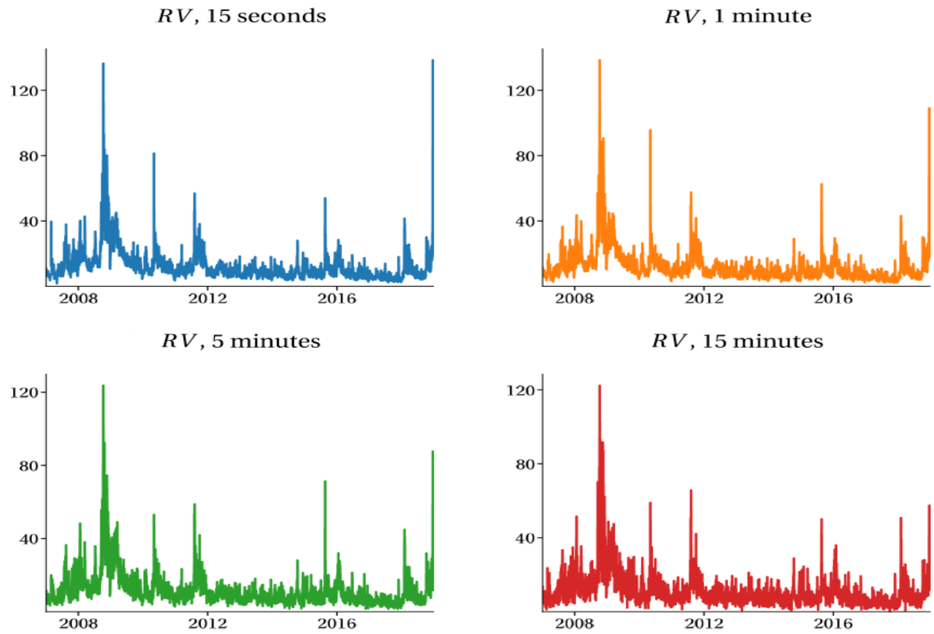


Figure 2.5: RV_t for the SPY computed at different frequencies

of interest. Thus when the jumps are added to the model setting, the RV is clearly an upward biased estimator of the IV.

When the jumps are infrequent is possible to use the *bipower variation* (BPV) (see Barndorff-Nielsen and Shephard, 2004), using again for consistency the notation of Corsi, Pirino and Renò

$$BPV^{(n)}(X)_t = \mu_1^{-2} \sum_{j=2}^{n} |\Delta_{j-1}X| \cdot |\Delta_j X|$$

where μ_1 is given and is approximately equal to 0.7979. Using the same notation as in equation 2.23, then likewise the work of Caporin, Rossi and Santucci de Magistris (2016) is possible to express the BPV as

$$BPV_t^{(n)} = \frac{\pi}{2} \sum_{j=2}^{n} |r_{t,j-1}| \cdot |r_{t,j}| \tag{2.27}$$

where $j = 2, ..., n$ are the fixed intra-day intervals, t is as usual the daily time interval (it can be any horizon).

The crucial characteristic of the BPV is to be a consistent estimator of the integrated variance even in case of jumps, namely

$$BPV_t^{(n)} \xrightarrow{p} \int_{t-1}^{t} \sigma_S^2 ds. \tag{2.28}$$

Moreover, given the results in equations 2.26 and 2.28 is possible to get a rough estimate of jump contribution, $[RV_t^{(n)} - BPV_t^{(n)}] \approx \sum_{t \in (0,1]} k_t^2$, this measure could be used in the context of testing for jumps.

The necessity of disentangling the continuous component (IV) from the jumps lies in the fact that even if volatility jumps are relevant in economic and financial applications, a plethora of past literature had not managed to show the effect of jumps in determining the future volatil-

ity. An explicative example is given in the work of Andersen, Bollerslev and Diebold (2007) where the jumps are shown to bring a negative or null impact in determining future volatility. However, it is shown that volatility is related with dispersion of beliefs and heterogeneous information.
If the occurrence of a jump increases the uncertainty of fundamental values, it is likely to have a positive impact on future volatility. Hence, the statement that jumps are not relevant in determining volatility becomes a "puzzle". It worth noting that the bipower variation behaves well in estimating the integrated variance asymptotically, whereas in finite sample it turns out to be an upward biased estimator of the IV in presence of jumps. Therefore the BPV leads to an underestimation of the discontinuous component.

Given the above reasons a way better estimator for the integrated variance in presence of jumps has been developed mainly by Corsi, Pirino and Renò (2010). Therefore the TBPV in the general notation is

$$TBPV^{(n)}(X)_t = \mu_1^{-2}\sum_{j=2}^{n}\left|\Delta_{j-1}X\right|\cdot\left|\Delta_j X\right| I_{(|\Delta_{j-1}X|^2\leq\delta_{j-1})} I_{(|\Delta_j X|^2\leq\delta_j)}$$

where $I_{(.)}$ is an indicator function taking the value 1 whenever its condition is satisfied and 0 otherwise, and δ_j is a positive threshold function designed to remove the jumps from the time series of returns.
Using as usual the more familiar notation the TBPV becomes

$$TBPV_t^{(n)} = \frac{\pi}{2}\sum_{j=2}^{n}\left|r_{t,j-1}\right|\cdot\left|r_{t,j}\right| I_{(|r_{t,j-1}|^2\leq\delta_{j-1})} I_{(|r_{t,j}|^2\leq\delta_j)} \tag{2.29}$$

where $j = 2,...,n$ defines the fixed length intra-daily intervals, $t = 1,...,T$ is the time period and as usual $r_{t,j}$ is the intradaily log-returns.
It worth noting that the equation 2.29 is very similar to the BPV in equation 2.27, the difference is in the presence of the two indicator functions which removes the log-returns larger than the specified thresholds δ_{j-1} and δ_j. While this difference is completely irrelevant asymptotically, it has been shown by Corsi et al. (2010) to be crucial in small samples, indeed very often $TBPV_t < BPV_t$ whenever there are jumps, this because the latter measure is biased.

Likewise the log-returns also for the realized variance some stylized facts have been ascertained (see Anderse et al.,2001):

- The unconditional distribution of the realized variance is both skewed and kurtosed, while the unconditional distribution of the logarithmic RV is nearly Gaussian.
- The logarithmic realized variance appears to be fractionally integrated, therefore volatility shocks die out very slowly.
- The logarithmic RV of stock indexes is not linear in returns, therefore the so called leverage effects occur, which means that past shocks on returns have a larger impact then past positive shocks of the same magnitude (see Black, 1976). This holds true also for stocks, however for this asset class the leverage effects are of minor economic importance.

Moreover is crucial to point out that even if daily returns standardized by their ex-post realized variance are nearly Gaussian, the same is not true when standardizing them by the one-day-ahead forecast realized volatility. Therefore, even if the aforementioned change seems to be of minor change, it has instead far-reaching consequences on the density of z_t (i.e. standardized log-returns).

Chapter 3

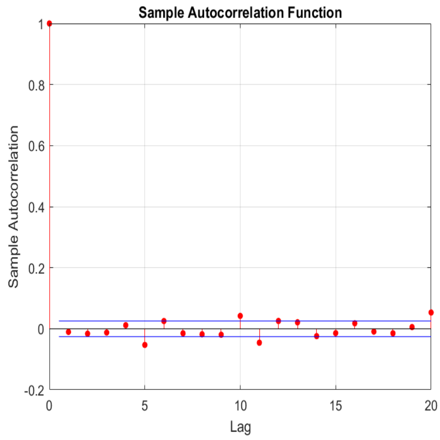
Methodology

In this chapter a bunch of different models of volatility are discussed, of course the list below is far to be exhaustive. The models described in this chapter are therefore the ones employed in the empirical analysis discussed in the last chapter.
Moreover in the last section various tests about the accuracy and the efficiency of the VaR measures are also described, the purpose of this section is to introduce the theoretical framework of the testing tools applied in chapter 4. Therefore the topics of this chapter are needed in order to perform the empirical investigation that is actually the core of the thesis.

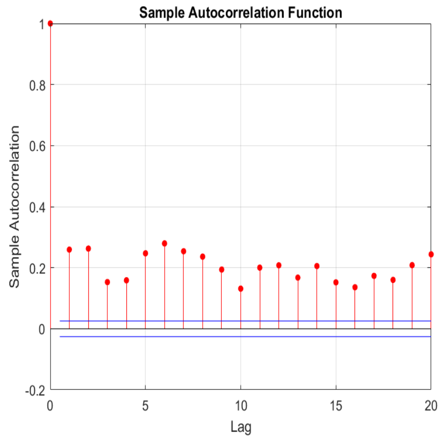
3.1 Overview of logarithmic returns and ARCH model

Starting with the Polish mathematician Benoit Mandelbrot in the 1956 and afterwards, many stylized facts regarding prices and especially logarithmic returns of financial assets have been ascertained. First of all, assets prices are not stationary and is usually assumed that they follow a Random Walk (RW), this is the main reason why the focus is posed on the log-returns which instead are stationary (they are a first order difference).
Moreover, usually log-returns exhibit no autocorrelation and therefore are often treated as White Noises (WN) [1].
However unlike returns the squared log-returns or their absolute value tend to exhibit correlation over time as showed in figures 3.1. Returns have several stylized facts, first their distribution is not Gaussian, because it has fat tails, high peak and is not perfectly symmetric. Another crucial stylized fact is that the returns exhibit volatility clusters, which means that large returns, in absolute value, tend to be followed by large returns (in absolute value), and vice versa. The figure 3.2 helps in displaying some of the above stylized facts regarding both log-returns and prices, indeed the former are stationary and swings around zero and clearly shows clusters of volatility especially in the dot-com bubble and financial crisis period, whereas the latter is not stationary and seems to show a trend and maybe even some changes in the structure (i.e. structural breaks, these features are further investigated when dealing with time series). Moreover from the QQ plot it can be seen that the log-returns show fat-tails and finally the plot of squared log-returns show that they are a first rough proxy of

[1] A white noise is a weakly stationary stochastic process which has some properties: let $u_t \sim WN(0,\sigma^2)$ then $\mathbb{E}[u_t] = 0$ and $\mathbb{E}[u_t^2] = \sigma^2 < \infty$ and finally $\mathbb{E}[u_t u_{t-s}] = 0 \; \forall s \neq 0$. Therefore a white noise is a stochastic process with mean zero, constant finite variance and zero autocovariance over time.



(a) *Autocorrelation Function of log-returns*



(b) *Autocorrelation Function of squared log-returns*

Figure 3.1: Autocorrelation Function of log and squared log-returns of the General Electric stock

the variance and it seems to be a kind of autocorrelation. Even though the unconditional variance of the log-returns is constant, their conditional one is not, this crucial point lead to the development of volatility models starting with the ARCH model with the ground-breaking paper of Robert Engle in the 1982. Therefore the focus is posed on modelling $Var(r_t \mid \mathscr{F}_{t-1})$, where the information set $\mathscr{F}_t$ is function of several variables depending on the functional form assumed for the *variance equation*.

Usually when modelling log-returns by using the ARCH (and in general discrete time volatility models) the return is assumed to be a weak white noise [2], namely a white noise with time varying (conditional) variance. The general setting of the ARCH(q) is:

$$r_t = \sqrt{\sigma_t^2} z_t \quad where \quad z_t \mid \mathscr{F}_{t-1} \sim D(0,1)$$

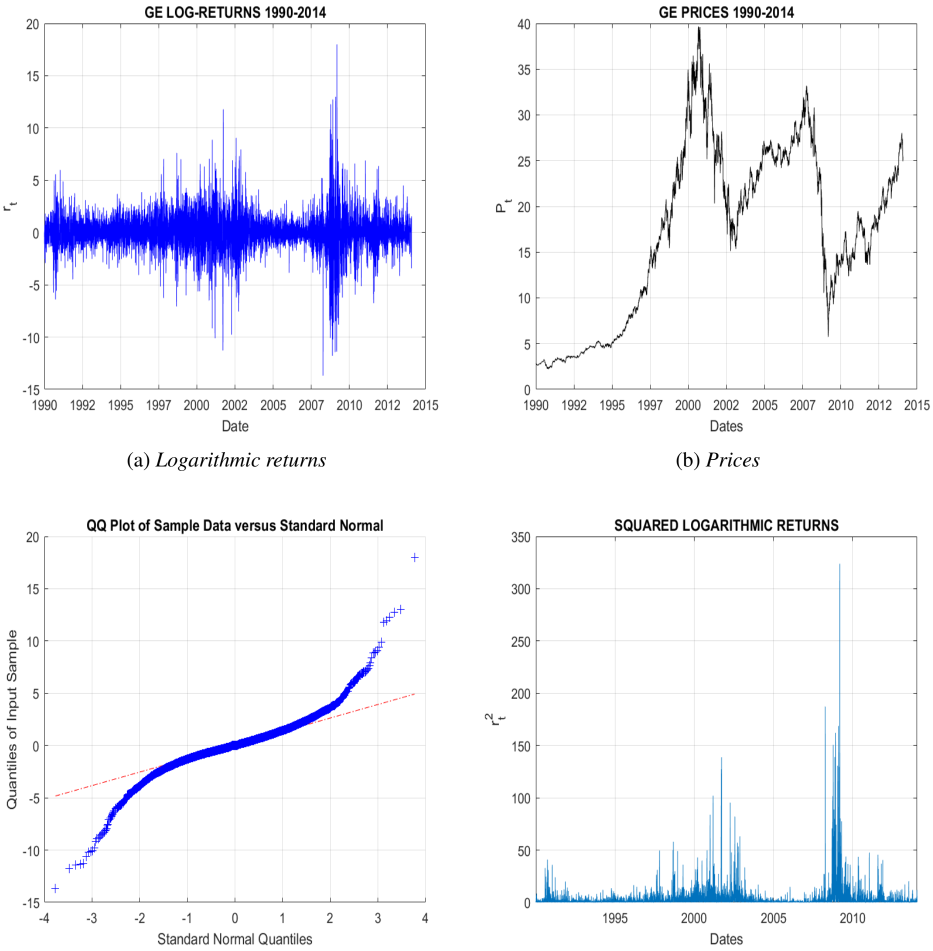
$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \quad where \quad \varepsilon_t := \sqrt{\sigma_t^2} z_t \tag{3.1}$$

where D is the distribution assumed for the standardized log-returns (or more precisely, standardized innovations), z_t, therefore $r_t \mid \mathscr{F}_{t-1} \sim D(0, \sigma_t^2)$ where the information set $\mathscr{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ is function of some elements measurable in t, of course is possible to give a conditional mean structure to the log-returns but is usually ignored because of its small magnitude.

The distributional assumption plays a crucial role because of the aforementioned stylized facts, moreover this assumption plays a pivotal role in the VaR computation as well.

In order for the ARCH model to be ”well behaved” (and in general for the models belonging to the so called ARCH family) two conditions are imposed, they becomes effective when estimating the model via Maximum Likelihood. Firs is the *positivity condition*, namely $\omega > 0 \vee \alpha_i \geq 0 \, \forall i = 1, 2, \ldots, q$, the second one is the *stationarity condition* which requires

[2] It worth noting that the assumption of the log-return to be a weak white noise makes it a zero correlation stochastic process, which is a coherent assumption with the theory of Efficient Market Hypothesis because it cannot be predicted (see Fama, 1970). Indeed the assumption of the log-return to be a white noise implies by construction that the price behaves as a random walk.



(a) *Logarithmic returns* (b) *Prices*

(c) *QQ plot against the Gaussian* (d) *Squared log-returns*

Figure 3.2: Log-returns , prices, QQ plot and squared log-returns of the General Electric stock

$\sum_{i=1}^{q} \alpha_i < 1$. Clearly the first condition is necessary because dealing with a centered second moment which is positive in nature, the second one is crucial in order to avoid the model to take an explosive dynamic. Under the stationarity assumption the unconditional mean of the log-returns is $\mathbb{E}[r_t] = 0$ and its unconditional variance is instead $Var(r_t) = \frac{\omega}{1-\sum_{i=1}^{q} \alpha_i}$, therefore it is important to stress that only the conditional mean and variance of r_t are time varying and not its unconditional ones. Another strong advantage of the ARCH model is the possibility to easily get forecasts of the conditional variance (without shocking) as shown, for an ARCH(1), in the equation 3.2 , under the assumption of stationarity the forecast converges to the unconditional variance, which is indeed often referred to as the long-run variance of the process.

$$\sigma^2_{t+1|t} := \mathbb{E}[\sigma^2_{t+1} \mid \mathscr{F}_t] = \omega + \alpha \varepsilon_t^2$$

$$\sigma^2_{t+k|t} := \mathbb{E}[\sigma_{t+k} \mid \mathscr{F}_t] = \omega \left(\sum_{i=0}^{k-1} \alpha^i \right) + \alpha^k \varepsilon_t^2. \quad (3.2)$$

The main drawback of the ARCH models is that in practice they turn out to be too rich, namely they requires to many parameters in order to fit adequately financial log-returns.

However, largely parametrized models can be hard to estimate and more important they are sometimes unstable when it comes to forecast, which is the pivotal usage of these models.

3.2 GARCH and GJR-GARCH

3.2.1 Description

In order to overcome the drawback of the ARCH models in the 1986 Tim Bollerslev, a PhD student of Robert Engle at UCSA, pubblished a papaer of paramount importance in financial econometrics called Generalized Autoregressive Conditional Heteroskedasticity (it was actually also the name of his PhD thesis). The GARCH models allow to fit financial returns adequately while keeping the number of parameters small, indeed in practice usually a GARCH(1,1) is enough.

Let $\varepsilon_t(\theta)$ be a discrete-time stochastic process with conditional mean and variance parametrized by a finite-dimensional vector $\theta \in \Theta \subseteq \mathbb{R}^m$, where θ_0 is the true unobservable parameters vector and Θ is the (compact) parameters space. Then as shown by Bollerslev et al.(1996) the process $\varepsilon_t(\theta_0)$ follows an ARCH model if $\mathbb{E}[\varepsilon_t(\theta_0 \mid \mathscr{F}_{t-1})] = 0$ for any $t = 1,2,...,T$, and the conditional variance $Var(\varepsilon_t(\theta_0) \mid \mathscr{F}_{t-1}) := \sigma_t^2(\theta_0)$ depends non-trivially on the $\sigma - algebra$ generated by the past observations.

Let $r_t(\theta_0)$ be the stochastic process of interest with conditional mean $\mathbb{E}[r_t \mid \mathscr{F}_{t-1}] := \mu_t(\theta_0)$, of course both $\mu_t(\theta_0)$ and $\sigma_t^2(\theta_0)$ are measurable with respect to the information set $\mathscr{F}_{t-1}$.

Then the process $\varepsilon_t(\theta_0) := r_t - \mu_t(\theta_0)$ and is possible to define the standardized process as $z_t(\theta_0) := \frac{\varepsilon_t(\theta_0}{\sqrt{\sigma_t^2(\theta_0)}}$ $\forall t = 1,2,...,T$ and $z_t \mid \mathscr{F}_{t-1} \sim D(0,1)$. Therefore the standardized process has a zero conditional mean, $\mathbb{E}[z_t(\theta_0) \mid \mathscr{F}_{t-1}] = 0$ and a unit conditional variance $Var(z_t(\theta_0) \mid \mathscr{F}_{t-1}) = 1$, this plays a crucial role in estimation because it is the variable with respect to the distribution assumption is made.

As stated above, in order to fit financial returns in a parsimonious fashion, Bollerslev developed in 1986 the GARCH(p,q) model, its general setting is

$$r_t = \mu_t + \varepsilon_t \quad where \;\; \varepsilon_t := \sqrt{\sigma_t^2} z_t \tag{3.3}$$

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2 \tag{3.4}$$

using the lag polynomial notation, hence $\alpha(L) := \alpha_1 L^1 + ... + \alpha_q L^q$ is the ARCH lag polynomial and $\beta(L) := \beta_1 L^1 + ... + \beta_p L^p$ is the GARCH polynomial. In practice, due to its parsimony, the GARCH(1,1) is enough to fit log-returns, thus the following specification is pretty popoular $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$.

It is possible to rewrite a GARCH(p,q) as an ARCH(∞) in the following way

$$\sigma_t^2 = \left(1 - \sum_{i=1}^{p} \beta_i L^i\right)^{-1} \left(\omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2\right)$$

$$= \omega^* + \sum_{k=0}^{\infty} \phi_k \varepsilon_{t-k-1}^2$$

thus the parsimony of the GARCH(p,q) model compared to the ARCH(q) one is pretty evident.

In order to guarantee the non-negativity, namely $\sigma_t^2 > 0 \,\forall\, t$, it requires that $\omega^* \geq 0 \vee \phi_k \geq 0$, therefore the non-negativity of the parameters ω^* and $\{\phi_k\}_{k=0}^{\infty}$ implies the one on σ_t^2. In order to make ω^* and $\{\phi_k\}_{k=0}^{\infty}$ well defined it is important to assume: the roots of the polynomial $\beta(x) = 1$ lie outside the unit circle (i.e. are greater than 1 in modulus, because they could be complex), and $\omega \geq 0$. Furthermore the polynomials $\alpha(x)$ and $[1-\beta(x)]$ must have no common roots.

Therefore in general as proposed by Bollerslev (1986) the positivity conditions are $\omega \geq 0 \vee \alpha_i \geq 0 \; i = 1,...,q \vee \beta_j \geq 0 \; j = 1,...,p$. For the simple GARCH(1,1) almost sure positivity of σ_t^2 requires, (see Nelson and Cao, 1992), that $\omega \geq 0 \vee \alpha \geq 0 \vee \beta \geq 0$.

The positivity conditions are important, first because the variance is positive in nature, but also when using the (conditional) maximum likelihood for estimation, because some assumptions on the value of the pre-sample $(\sigma_{-1}^2,...,\sigma_{-p}^2,\varepsilon_{-1}^2,...,\varepsilon_{-q}^2)$ are necessary and therefore the positivity conditions defined above ensure that $\{\sigma_t^2\}_{t=0}^{\infty}$ is not negative for any arbitrary non-negative value of the pre-sample.

Another property of the GARCH models required in order to get a well behaved representation is the stationarity condition, in particular it plays a crucial role when performing forecasts. The process (ε_t) which follows a GARCH(p,q) is a martingale difference sequence (MDS), therefore in order to study the weak stationarity is sufficient to consider that $Var(\varepsilon_t) = Var(\mathbb{E}[\varepsilon_t \mid \mathscr{F}_{t-1}]) + \mathbb{E}[Var(\varepsilon_t \mid \mathscr{F}_{t-1})] = \mathbb{E}[\sigma_t^2]$, hence is only needed to show that it is asymptotically constant in time. Moreover a process $\{\varepsilon_t\}$ that behaves as a GARCH(p,q) with non-negative coefficients $\omega > 0 \vee \alpha_i \geq 0 \; i = 1,...,q \vee \beta_j \geq 0 \; j = 1,...,p$ is covariance stationary if and only if $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$, therefore for a GARCH(1,1) the stationarity condition becomes $\alpha + \beta < 1$.

The above is a sufficient but not necessary condition for strict stationarity, indeed any weakly stationary GARCH process is also strictly stationary.

For a well behaved GARCH(1,1) process its unconditional variance is $\mathbb{E}[\sigma_t^2] = Var(r_t^2) = \frac{\omega}{1-\alpha-\beta}$. Moreover its fourth moment is $\mathbb{E}[r_t^4] = 3\omega^2(1+\alpha+\beta)[(1-\alpha-\beta)(1-\beta^2-2\alpha\beta-3\alpha^2)]^{-1}$, hence its kurtosis is greater than 3, namely $K(r_t) = \frac{\mathbb{E}[r_t^4}{\mathbb{E}[r_t^2]^2} = \frac{3[(1+\alpha+\beta)(1-\alpha-\beta)}{1-\beta^2-2\alpha\beta-3\alpha^2} > 3$.

One of the most relevant applications of the GARCH models is forecasting future volatility, therefore an useful representation is the GARCH(p,q) as an ARMA for the precess ε_t^2, given that $\varepsilon_t := \sigma_t^2 + \nu_t$ where $\mathbb{E}[\nu_t \mid \mathscr{F}_{t-1}] = 0$ and $\nu_t \in [-\sigma_t^2, \infty]$, then $\varepsilon_t^2 \sim ARMA(m,p)$

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{max(p,q)} (\alpha_j + \beta_j)\varepsilon_{t-j}^2 + (\nu_t - \sum_{i=1}^{p} \beta_j \nu_{t-1})$$

where $m = max(p,q)$, therefore by writing the process in two parts, after and before time t

$$\sigma_{t+k}^2 = \omega + \sum_{i=1}^{n} (\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2) + \sum_{i=k}^{m} (\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2)$$

where $n = min(m, k-1)$.

Thus forecasting via GARCH(p,q), as shown by Engle and Bollerslev in 1986 boils down to the following general formula

$$\mathbb{E}[\sigma_{t+k}^2 \mid \mathscr{F}_t] := \sigma_{t+k|t}^2 = \omega + \sum_{i=1}^{n} \Big((\alpha_i + \beta_i)\mathbb{E}_t[\sigma_{t+k-i}^2] \Big) + \sum_{j=k}^{m} (\alpha_j \varepsilon_{t+k-j} + \beta_j \sigma_{t+k-j}^2). \quad (3.5)$$

In particular, for a GARCH(1,1) there are two different cases

$$\sigma^2_{t+k|t} = \sigma^2 + (\alpha+\beta)^{k-1}(\sigma^2_{t+1} - \sigma^2) \quad when \;\; k \geq 2$$

$$\sigma^2_{t+1|t} = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2 \quad when \;\; k = 1. \tag{3.6}$$

Therefore whenever the process is covariance stationary, then $\mathbb{E}[\sigma^2_{t+k} \mid \mathscr{F}_t]$ converges to its unconditional variance σ^2 as $k \to \infty$ (i.e. the process is mean reverting). Moreover, from the above formulas it is evident that the forecasts are based on the measurability of $\sigma^2_{t+1|t}$ with respect to $\mathscr{F}_t$.

The GARCH models are symmetric models in the sense that they treat shocks on returns regardless of the sign, thus positive and negative shocks have the same magnitude on volatility. The News Impact Curve (NIC) introduced by Pagan and Schwert (1990) and christened by Engle and Ng (1993) characterizes the impact of past returns shocks on the return volatility which is implicit in a volatility model.
Keeping constant the informations from $t-2$ and earlier, is possible to examine the implied relation between ε_{t-1} and σ_t^2. Therefore the NIC relates past returns shocks (news) to current volatility, namely it measures how news are incorporated into volatility estimates.
Therefore all lagged conditional variances are evaluated at the level of the unconditional one, that is $\sigma^2_{t-1} := \sigma^2$. Thus for the GARCH(1,1) the NIC has the following expression

$$\sigma^2 = A + \alpha\varepsilon^2_{t-1}$$

$$A := \omega + \beta\sigma^2.$$

The figure 3.3 shows graphically the symmetric relationship implied by the GARCH(1,1) between lagged shocks and current conditional variances. As shown above the GARCH

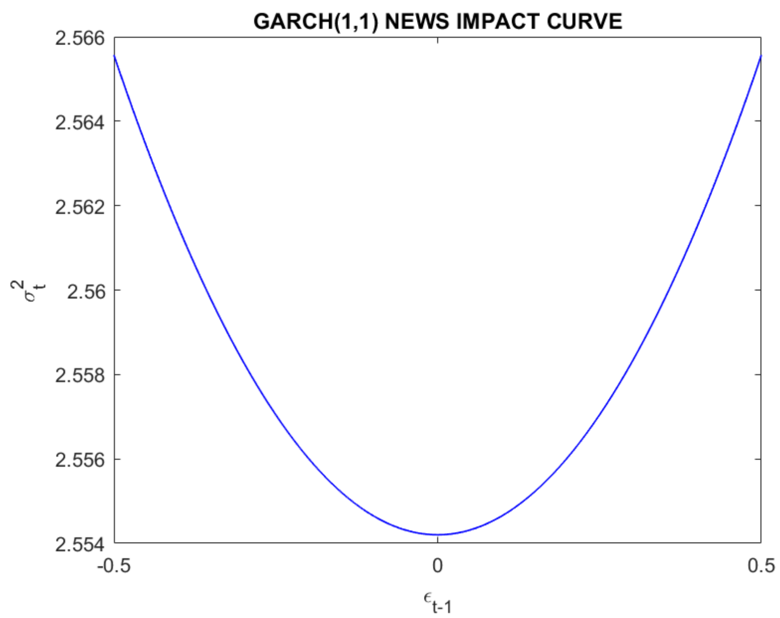


Figure 3.3: News Impact Curve for a GARCH(1,1)

specification implies the response to a shock to be independent of the sign of the shock, and

therefore it is just a function of the magnitude of the shock itself. However, a stylized fact of financial volatility is that bad news (i.e. negative shocks) tend to have a greater impact on volatility than good news (i.e. positive shocks), this is true especially for stocks and stock indexes.
That is, volatility tends to be greater in falling markets than in rising markets.
Fisher Black in 1976 attributed the aforementioned effect, also referred to as leverage effect, to the fact that bad news tend to drive down the stock price thus increasing the financial leverage (i.e. the debt-to-equity ratio) of the stock, causing the stock to be more volatile. Therefore a plethora of asymmetric GARCH models arose over the past years, in order to take into account the leverage effect. For instance the Exponential GARCH (E-GARCH) by Nelson (1991), the Asymmetric GARCH (A-GARCH) by Engle (1990), the Quadratic GARCH (Q-GARCH) by Sentana (1995), the Threshold GARCH (T-GARCH) by Zakoian (1994), etc. Nowadays there is a GARCH zoo as stated by Hansen, the asymmetric GARCH used in this thesis is the one developed in the 1993 by Glosten, Jagannathan and Runkle also known as GJR-GARCH (it can be view as a variant of the T-GARCH).

The GJR-GARCH is pretty similar to the standard GARCH, but at the same time is able to capture the leverage effects via and indicator function, namely a function that takes value 1 when a condition is satisfied and 0 otherwise. In the most general setting the GJR-GARCH(p,q) is expressed as

$$r_t = \mu_t + \varepsilon_t \quad \varepsilon := \sqrt{\sigma_t^2} z_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} (\alpha_i \varepsilon_{t-1}^2 + \gamma_i I_{t-1}^- \varepsilon_{t-i}^2) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{3.8}$$

where of course there is a distribution assumption of the type $z_t \mid \mathscr{F}_{t-1} \sim D(0,1)$.
It can be noted that the return equation is the same as the one of the standard GARCH model (see eq. 3.3), therefore is only the functional form assumed for the conditional variance that changes. The leverage effect, as stated above, is captured in the variance equation by the indicator function

$$I_t^- = \begin{cases} 1 & \text{if } \varepsilon_t < 0 \\ 0 & \text{if } \varepsilon_t \geq 0 \end{cases} \tag{3.9}$$

Therefore the conditional variance increases when $\varepsilon_t < 0$ due to the behaviour of the indicator function, of course this holds true whenever $\gamma_i > 0 \; \forall i = 1, 2, ..., q$.
As usual a GJR-GARCH(1,1) is taken into consideration because it is enough to adequately fit financial log-returns. The changes in the functional form of the variance equation affects the model in many ways. Indeed the unconditional variance of log-returns becomes $Var(r_t) = \mathbb{E}[\sigma_t^2] = \frac{\omega}{1-\alpha-\beta-\frac{1}{2}\gamma}$, this result comes not only from the stationarity assumption, but also from the following convention $\mathbb{E}[I_t^-] = \frac{1}{2}$.
Moreover the positivity condition becomes $\omega > 0 \vee \alpha \geq 0 \vee \beta \geq 0 \vee \gamma \geq 0$, whereas the stationarity condition, using the convention that $\mathbb{E}[I_t^-] = \frac{1}{2}$, becomes $(\alpha + \beta + \frac{1}{2}\gamma) < 1$.

The asymmetric feature of the GJR-GARCH con be noted in the variance equation, moreover a graphical representation can be given by the News Impact Curve shown in figure 3.4. For this model the NIC becomes

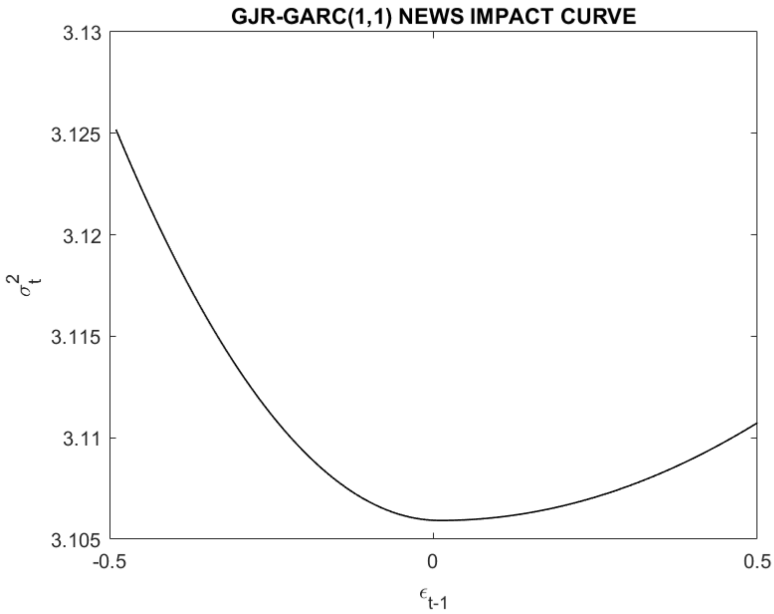


Figure 3.4: News Impact Curve for a GJR-GARCH(1,1)

$$\sigma_t^2 = \begin{cases} A + \alpha\varepsilon_{t-1}^2 & if\ \varepsilon_{t-1} \geq 0 \\ A + (\alpha+\gamma)\varepsilon_{t-1}^2 & if\ \varepsilon_{t-1} < 0 \end{cases}$$

Where as stated for the standard GARCH model, $A := \omega + \beta\sigma^2$. Thus the NIC representation makes clear the greater impact of past negative shocks (i.e. $\varepsilon_{t-1} < 0$) on volatility, this holds true due to the positivity of γ. Likewise the GARCH NIC, also the GJR-GARCH NIC reaches its minimum at $\varepsilon_{t-1} = 0$.

3.2.2 Estimation

The most used, but not unique, procedure for the estimation of $\theta_0 \in \Theta$ in ARCH/GARCH models is the maximum likelihood, a procedure based on the maximization of a likelihood function build upon the assumption of an i.i.d. distribution for the standardized innovations $z_t(\theta)$.
Therefore $f(z_t(\theta);\eta)$ denotes the density function of $z_t(\theta) := \frac{\varepsilon_t(\theta)}{\sqrt{\sigma_t^2(\theta)}}$, with zero mean and unit variance, where η is a nuisance parameter such that $\eta \in H \subseteq \mathbb{R}^k$. Let $(r_1, r_2, ..., r_T)$ be a sample of realizations from a GARCH model as defined in equation 3.3 and $\psi := [\theta \ \eta]'$ the $(m+k) \times 1$ vector of parameters to be estimated for the conditional mean, variance and density function.
Therefore, even though the distribution assumption is done with respect to $z_t(\theta)$ the observable component is $r_t(\psi)$, thus a change of variable in the density function is performed by computing the Jacobian arising in the transformation from the standardized innovations $z_t(\theta)$ to the observable log-returns $r_t(\psi)$, namely $J := \frac{\partial z_t}{\partial r_t} = \frac{1}{\sigma_t(\theta)}$.
Hence, the density function of the observable components becomes

$$f(r_t;\psi) = f(z_t(\theta);\eta) \cdot |J| \tag{3.10}$$

under the assumption that $r_t = g(z_t)$ where $g(\cdot)$ is a monotonic function. The conditional log-likelihood (i.e. the logarithm of the density) for the t-th observation is the following

$$l_t(r_t;\psi) = log(f(z_t(\theta);\eta)) - \frac{1}{2} log(\sigma_t^2(\theta)) \quad t = 1, 2, ..., T. \tag{3.11}$$

The conditional log-likelihood of the full sample is then the sum of the log-likelihood functions in equation 3.11, therefore it becomes

$$l(r_1, r_2, ..., r_T;\psi) = \sum_{t-1}^{T} l_t(r_t;\psi). \tag{3.12}$$

The maximum likelihood estimator for the true parameter $\psi_0 := [\theta_0 \ \eta_0]'$, usually denoted as $\widehat{\psi_T}$, is obtained by means of maximization of equation 3.12. Therefore the following problem has to be solved

$$\widehat{\psi_T} := \underset{\psi \in \Psi}{\text{argmax}} \ l(\psi). \tag{3.13}$$

Under the assumption of differentiable conditional density and $\mu_t(\theta)$ and $\sigma_t(\theta)^2$ to be differentiable functions, then for any $\psi \in \Theta \times H := \Psi$, the maximum likelihood estimator is the solution to the so called likelihood equations [3]

$$S\left(r_1, r_2, ..., r_T; \widehat{\psi_T}\right) = \sum_{t-1}^{T} s_t\left(r_t; \widehat{\psi_T}\right) = \mathbf{0} \tag{3.14}$$

[3]The maximum likelihood estimation in these types of problems behaves well because the distribution is based on an absolutely continuous and smooth functions, moreover the solution to the likelihood equations (besides some cases as the Linear Gaussian Model) are found via numerical optimization routine, given the good behaviour of the problem the most used is the Newton-Raphson method (or Quasi-Newton-Raphson for Hessian correction and avoiding computation, see DFP, BFGS, ecc). This aforementioned method has centuries, nowadays several new methods has been developed, especially during the 80's-90's, for instance methods such as the Particle Swarm, Simulated Annealing, Genetic, etc.

where $s_t := \frac{\partial l_t(r_t;\psi)}{\partial \psi}$ is the so called score vector for the t-th observation (i.e. the gradient).

In order to implement the maximum likelihood estimation, an explicit assumption regarding the conditional density in equation 3.10 has to be made. The most used distribution assumption in the literature is the Gaussian. Therefore the density of the standardized innovation is

$$f(z_t(\theta);\eta) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z_t(\theta)^2}{2}\right). \tag{3.15}$$

Given that the normal distribution is uniquely determined by the first two moments, only the conditional mean and variance parameters enter the log-likelihood function, thus $\psi = \theta$.
The log-likelihood function for the t-th observations is

$$l_t(\theta) = -\frac{1}{2}log(2\pi) - \frac{1}{2}log(\sigma_t^2(\theta)) - \frac{1}{2}\frac{\varepsilon_t(\theta)^2}{\sigma_t^2(\theta)}. \tag{3.16}$$

The conditional Gaussian log-likelihood for the full sample thus becomes

$$l(r_1, r_2, ..., r_T; \theta) = \sum_{t=1}^{T} l_t(\theta) = -\frac{T}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} log(\sigma_t^2(\theta)) - \frac{1}{2}\sum_{t=1}^{T}\frac{\varepsilon_t^2(\theta)}{\sigma_t^2(\theta)}. \tag{3.17}$$

When $\sigma_t^2(\theta)$ is modelled as a GARCH(p,q), starting values for the model parameters ω, $\alpha_i \;\; i = 1,2,...,q$ and $\beta_j \;\; j = 1,2,...,p$ need to be chosen and an initialization of ε_t^2 and σ_t^2 must be supplied.
Usually the starting value of μ_t is set equal to the sample mean, the starting value of σ_t^2 is set to the unconditional variance based on arbitrary initial values of the parameters or it is set equal to the sample variance, whereas the innovations are usually set equal to zero.

Another widely used distribution is the Student-t, which is useful in order to capture the so called fat tail phenomenon, therefore the density of the standardized innovation becomes

$$f(z_t(\theta);\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})}\left(1 + \frac{z_t^2(\theta)}{\nu-2}\right)^{-(\frac{\nu+1}{2})} \tag{3.18}$$

where ν is the number of degrees of freedom, of course in order for the second moment to exists it requires $\nu > 2$, in general the n-th moment exists whenever $\nu > n$.
Unlike the Gaussian distribution, in the Student-t another parameter has to be estimated, namely the number of degrees of freedom, thus $\psi = [\theta \;\nu]'$. The log-likelihood for the t-th observation is

$$l_t(\psi) = log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \frac{1}{2}log(\pi(\nu-2)) - \frac{1}{2}log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{1}{2}log(\sigma_t^2(\theta))$$
$$- \left(\frac{\nu+1}{2}\right) log\left(1 + \frac{\varepsilon_t^2(\theta)}{\sigma_t^2(\theta)(\nu-2)}\right) \tag{3.19}$$

where $\Gamma(k) = \int_0^\infty x^{k-1}e^{-x}dx \;\; k > 0$ is the Gamma function for non-integer positive values, whereas for integer positive values it becomes $\Gamma(n) = (n-1)! \;\; n \in \mathbb{N}$.
As stated above the conditional log-likelihood of the full sample is just the summation of the conditional log-likelihood over the sample, $l(r_1, r_2, ..., r_T; \psi) = \sum_{t=1}^{T} l_t(\psi)$.

In literature other distributions have been used, for instance the GED distribution or, in order to capture also the skewness in the distribution of log-returns, the Skew-t density. For the purposes of this thesis only the Gaussian and the Student-t distributions have been used.

3.3 Realized-GARCH

3.3.1 Description

The GARCH models developed starting from the 1982 until the late 90s are all based on daily data, regardless of all the extension made from the original works of Engle and Bollerslev. Strating from the new century the high frequency trading, the availability of computers and the birth of huge databases allowed the possibility of exploiting intra-daily data at a reasonable "cost".

The GARCH models with daily data have an important drawback, especially in applications such as risk management, because they update the volatility to new shocks very slowly as show in figure 3.5, hence is pretty clear that in the current markets a slow update is a drawback to get rid of. Therefore starting from the 2002 with the work of Robert Engle the first

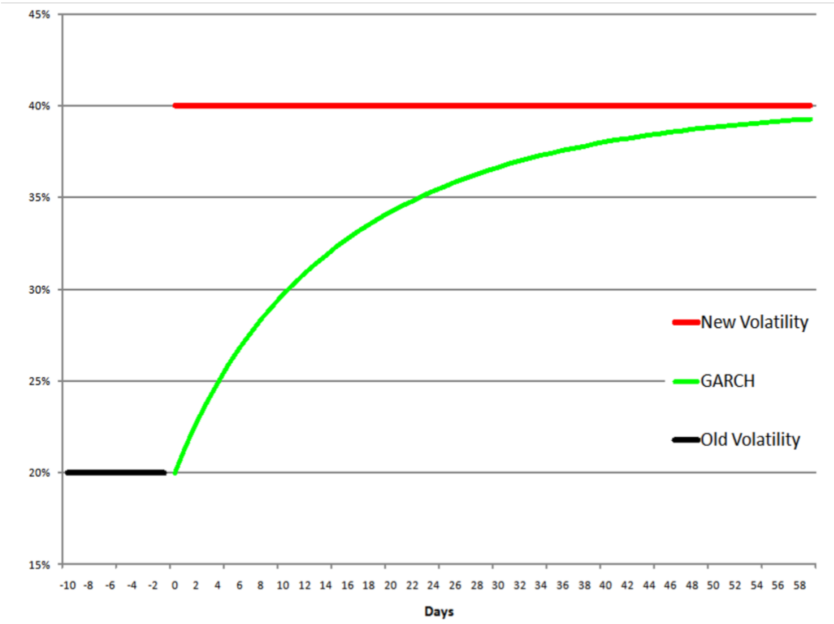


Figure 3.5: Updating behaviour of the standard GARCH model

GARCH model including intra-daily data has been introduced. The first rudimental model was the so called GARCH-X (X because the variable x_t is treated as exogenous)

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma x_{t-1} \tag{3.20}$$

where the variable x_t is the one encompassing the high-frequency data, namely a realized measure of volatility (e.g. Realized Variance, Realized Kernel, MC estimator, etc.).
However the model made by Engle in the 2002 is clearly, as stated by Hansen, incomplete due to the lack of an equation describing the behaviour of x_t, the so called measurement equation. Therefore in the 2012 Hansen, Huang and Shek pubblished a work containing the Realized GARCH, with fixes the problem of the above model by specifying a measurement equation describing x_t, thus making the GARCH model a bivariate one. Moreover the Realized GARCH is able to solve the problem of slow updating of the "daily" GARCH models, indeed in figure 3.6 it can be seen that the Realized GARCH incorporates way faster new

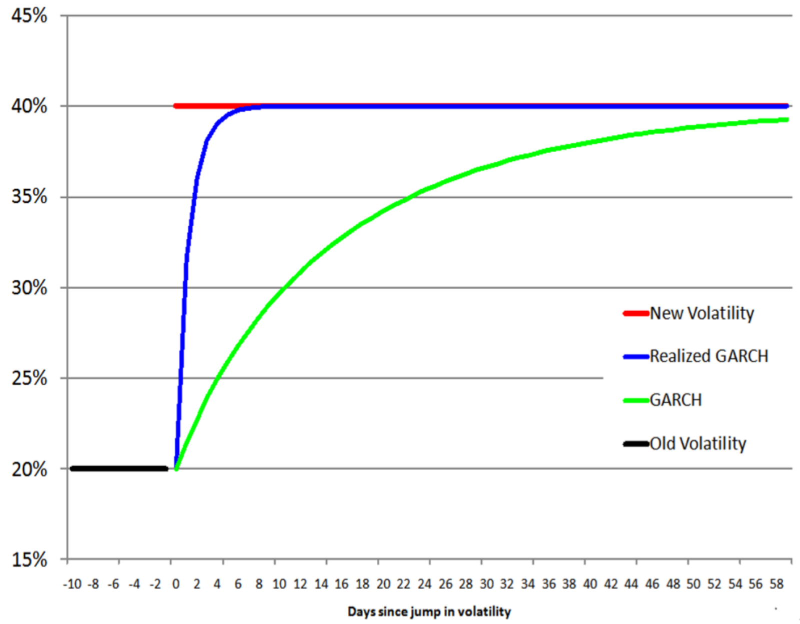


Figure 3.6: Updating behaviour of the Realized GARCH model

informations compared to the standard GARCH.

The canonical version of the Realized GARCH model with a linear specification is given by the following three equations

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t \quad where \;\; \varepsilon_t := \sqrt{\sigma_t^2} z_t \\ \sigma_t^2 &= \omega + \beta \sigma_{t-1}^2 + \gamma x_{t-1} \quad where \\ x_t &= \phi_0 + \phi_1 \sigma_t^2 + \tau(z_t) + u_t \end{aligned} \tag{3.21}$$

where x_t is a realized measure of volatility, usually the realized rolatility or the bipower variation (see part 2.2), $z_t \sim D_z(0,1)$, $u_t \sim D_u(0,\sigma_u^2)$ and as usual $\sigma_t^2 := Var(r_t \mid \mathscr{F}_{t-1})$, where the information set $\mathscr{F}_t := \sigma(r_t, x_t, r_{t-1}, x_{t-1}, \ldots)$.
Usually the first of the equations in 3.21 is referred to as the *return equation*, whereas the second one is the *GARCH equation*, another important remark is that z_t and u_t are mutually independent. The last equation link the observed realized measure to the latent volatility, therefore is called *measurement equation*, the latter thus completes the model which specifies the dynamic properties of both returns and the realized measure (unlike the GARCH-X). This equation if natural when x_t is a consistent estimator of the integrated variance (see part 2.2 for a better understanding), because the integrated variance may be viewed as the conditional variance plus a random innovation, and the latter is absorbed by the $\tau(z_t) + u_t$ part.
The simplest measurement equation is $x_t = \sigma_t^2 + u_t$, the importance of the measurement equation is based also on the fact that it allows a simple way to model the joint dependence between r_t and x_t, which is known to be empirically important.

However the linear specification, alike the log-linear one, allows for many lags in the variable of the GARCH equation, thus the Realized GARCH(p,q) can be defined in the most general setting as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \gamma_j x_{t-j} \quad and \;\; x_t = \phi_0 + \phi_1 \sigma_t^2 + \tau(z_t) + u_t.$$

As in the case for the GARCH(1,1) model the Realized GARCH(1,1) model with linear specification implies that σ_t^2 has an AR(1) representation that is, $\sigma_t^2 = (\omega + \gamma\phi_0) + (\beta + \gamma\phi_1)\sigma_t^2 + \gamma v_{t-1}$ where $v_t := \tau(z_t) + u_t$ is an i.i.d. process and the realized measure x_t is an ARMA(1,1).

As stated many times when dealing with some asset classes, especially stocks and stock indexes the leverage effect becomes relevant, therefore is crucial to plug into the model a component able to account for the sign of the shocks. In the Realized GARCH that role is played by the function $\tau(z_t)$ which is called not by chance leverage function, indeed it is able to captures the dependence between returns and future volatility.
The leverage function is normalized by considering $\mathbb{E}[\tau(z_t)] = 0$. In the paper of Hansen et al., the leverage functions are constructed from Hermite polynomials, namely

$$\tau(z) = \tau_1(z) + \tau_2(z^2 - 1) + \tau_3(z^3 - 3z) + \tau_4(z^4 - 6z^2 + 3) + ...$$

the baseline choice of the leverage function is a simple quadratic form $\tau(z_t) = \tau_1 z_t + \tau_2(z_t^2 - 1)$, where of course τ_1 and τ_2 are two parameters to be estimated. This choice is convenient because it ensures that $\mathbb{E}[\tau(z_t)] = 0$, for any distribution such that $\mathbb{E}[z_t] = 0$ and $Var(z_t) = 1$, which is the usual choice in each model (because z_t is a standardized innovation).
It is easy to find in applications also the log-linear specification for the Realized GARCH (likewise HAR-RV models), because it has an useful advantage over the linear specification when it comes to the mutually independence between z_t and u_t, the general expression is

$$r_t = \mu_t + \varepsilon_t \quad where \;\; \varepsilon_t := \sqrt{\sigma_t^2} z_t$$

$$log(\sigma_t^2) = \omega \sum_{i=1}^{p} \beta_i log(\sigma_{t-i}^2) + \sum_{j=1}^{q} \gamma_j log(x_{t-j}) \tag{3.22}$$

$$log(x_t) = \phi_0 + \phi_1 log(\sigma_t^2) + \tau(z_t) + u_t.$$

The conditional variance σ_t^2 is adapted to $\mathscr{F}_{t-1}$, therefore $\mathscr{F}_t$ has to be such that $x_t \in \mathscr{F}_t$. This requirement is satisfied by $\mathscr{F}_t = \sigma(r_t, x_t, r_{t-1}, x_{t-1}, ...)$ but $\mathscr{F}_t$ could in principle be an even richer σ-field, moreover the measurement equation does not require x_t to be an unbiased measure of σ_t^2. An attractive feature of the log-linear Realized GARCH model is that it preserves the ARMA structure that characterizes some of the standard GARCH models.
Moreover, the measurement equation induces a GARCH structure similar to an E-GARCH with a stochastic volatility component, thus the Realized GARCH can induce a flexible stochastic volatility structure, but does in fact have a GARCH structure because u_{t-1} is $\mathscr{F}_{t-1}$-measurable. This is and interesting feature, especially when performing forecasts beyond one-step-ahead. Another advantage of the log-linear specification is that it ensure positivity. Finally the log-linear specification is less at odds with the data, resulting in a way lower misspecification than the linear model, which present greater heteroskedasticity as it can be seen in figure 3.7, the heteroskedasticity causes Quasi-maximum-likelihood to be inefficient.

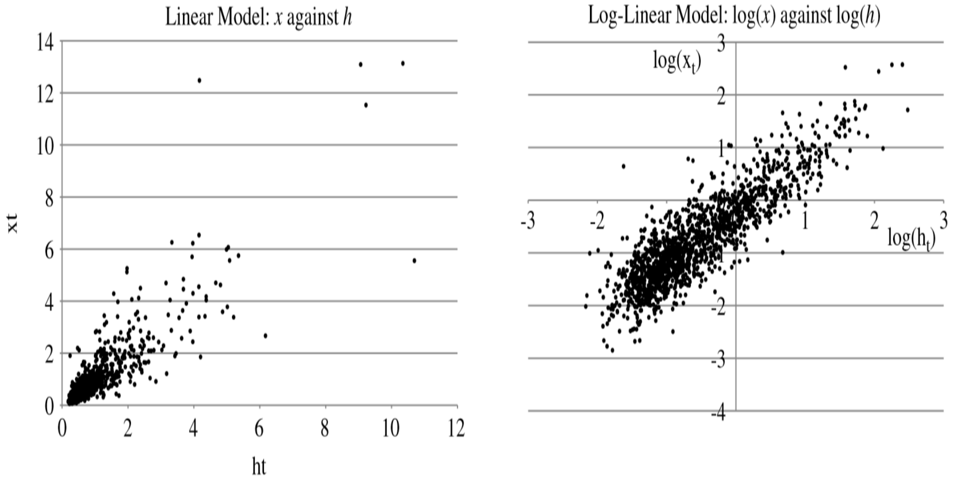


Figure 3.7: Heteroskedasticity in measurement equation, where $h_t := \sigma_t^2$

Moreover the log-linear specification, with the presence of the leverage function, offers a much better agreement with the underlying assumption $z_t \perp u_t \ \forall t = 1, 2, ...T$, a graphical explanation is given in figure 3.8.

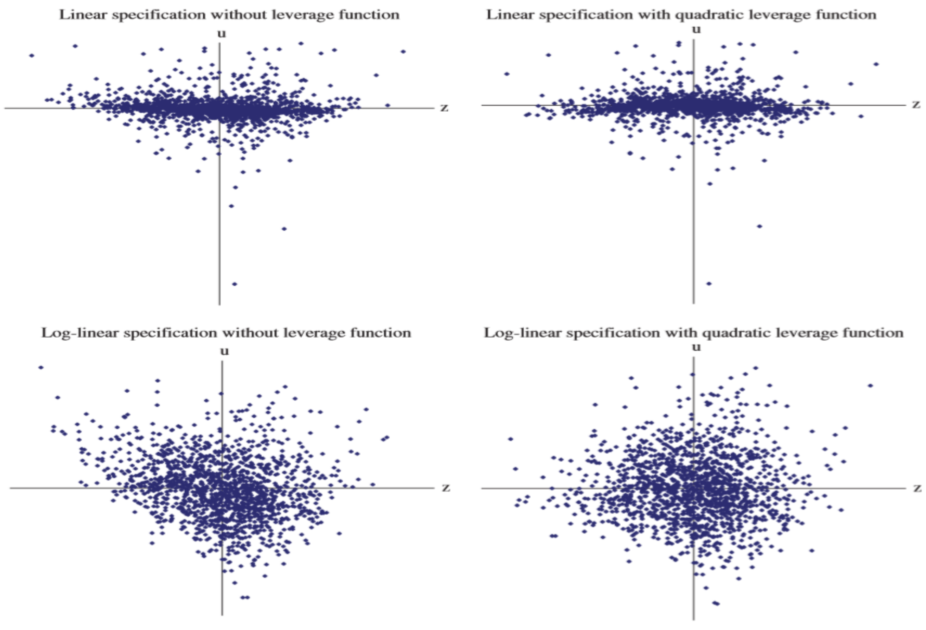


Figure 3.8: Scatter plots of the residuals $(\hat{z}_t, \hat{u}_t)$ for different setups

The leverage function is closely related to the News Impact Curve, which maps out how positive and negative shocks to the price affect future volatility (see part 3.2.1).

High-frequency data facilitate a more detailed study of the news impact curve than is possible with daily returns, Chen and Ghysels (2011) studied the NIC in this context. Moreover the Hermite polynomial specification for the leverage function presents a very flexible framework for estimating this effect. The presence of the function $\tau(z_t)$ allows, as it can be seen from figure 3.9, an asymmetric NIC capturing the leverage effect.

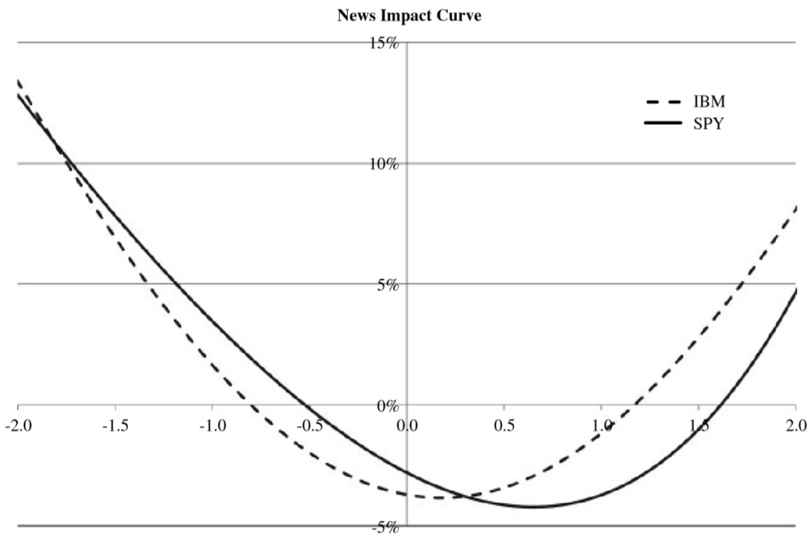


Figure 3.9: News Impact Curve for IBM and SPY

The NIC in the Real GARCH(1,1) with log-linear specification is $\gamma\tau(z_t)$, moreover it worth noting that unlike the standard GARCH and the GJR the Real GARCH NIC hits its minimum not in zero.

3.3.2 Estimation

The Realized GARCH can be estimated via conditional log-likelihood (see part 3.2.2), it is actually a Quasi-loglikelihood due to the heteroskedasticity structure in the innovations. Under the assumptions that $z_t \sim NID(0,1) \vee \frac{u_t}{\sigma_u} \sim NID(0,1)$ with the condition that $(z_t, \frac{u_t}{\sigma_u}) \sim N(\mathbf{0},\mathbf{I})$, then the conditional densities of interest are

$$f(r_t \mid \mathscr{F}_{t-1};\theta_1) = \frac{1}{\sqrt{2\pi\sigma_t^2(\theta_1)}} exp\left(-\frac{\varepsilon_t^2(\theta_1)}{2\sigma_t^2(\theta_1)}\right)$$

$$f(x_t \mid \mathscr{F}_{t-1}, r_t;\psi) = \frac{1}{\sqrt{2\pi\sigma_u^2}} exp\left(-\frac{u_t^2(\psi)}{2\sigma_u^2}\right) \tag{3.23}$$

where $\mathscr{F}_t = \sigma(r_t, x_t, r_{t-1}, x_{t-1}, ...)$ is the information set on which the conditioning takes place, $\theta_1 := [\omega\ \beta\ \gamma]'$ is the vector of parameters of the variance equation, $\theta_2 := [\phi_0\ \phi_1\ \tau_1\ \tau_2\ \sigma_u^2]'$ is the vector of parameters of the measurement equation and $\psi := [\theta_1\ \theta_2]'$ is the vector of all the parameters of the bivariate model.

Actually the log-likelihood function comes from a joint density, $l(\psi) = \sum_{t=1}^T log(f(r_t, x_t \mid \mathscr{F}_{t-1};\psi))$. However, the joint density can be factorized (due to the Bayes formula) into the product between the marginal density and the conditional density, namely

$$f(r_t, x_t \mid \mathscr{F}_{t-1};\psi) = f(r_t \mid \mathscr{F}_{t-1};\theta_1) f(x_t \mid \mathscr{F}_{t-1}, r_t;\psi) \tag{3.24}$$

given the distribution assumption on z_t and u_t and their mutual independence, is then possible to write the joint log-likelihood function at the t-th observation as

$$l_t(\psi) = \left[-\frac{1}{2}log(2\pi) - \frac{1}{2}log(\sigma_t^2(\theta_1)) - \frac{1}{2}\frac{\varepsilon_t^2(\theta_1)}{\sigma_t^2(\theta_1)}\right] + \left[-\frac{1}{2}log(2\pi) - \frac{1}{2}log(\sigma_u^2) - \frac{1}{2}\frac{u_t^2(\psi)}{\sigma_u^2}\right].$$

As usual, the joint log-likelihood function of the full sample is the sum over all the observations, and it is written as $l(r_1, ..., r_T, x_1, ..., x_T;\psi) := \sum_{t=1}^T l_t(\psi)$.
It worth noting that in order to compare via the log-likelihood functions the Realized GARCH with daily GARCH models, the first has to be considered only with respect to the so called partial log-likelihood function. The partial log-likelihood function is the log-likelihood of the log-return that is, $l_t(\theta_1) = -\frac{1}{2}log(2\pi) - \frac{1}{2}log(\sigma_t^2(\theta_1)) - \frac{1}{2}\frac{\varepsilon^2(\theta_1)}{\sigma_t^2(\theta_1)}$.

Another distribution specification that is useful in modelling financial returns is the Student-t, which means $z_t \sim t_\nu \vee u_t \sim N(0,\sigma_u^2)$ and assuming mutual independence between both z_t and u_t. Then the conditional densities of interest are

$$f(r_t \mid \mathscr{F}_{t-1};\theta_1) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\sigma_t^2(\theta_1)\pi(\nu-2)}\Gamma(\frac{\nu}{2})}\left[1 + \frac{\varepsilon_t^2(\theta_1)}{\sigma_t^2(\theta_1)(\nu-2)}\right]^{-\frac{\nu+1}{2}}$$

$$f(x_t \mid \mathscr{F}_{t-1}, r_t;\psi) = \frac{1}{\sqrt{2\pi\sigma_u^2}} exp\left(-\frac{u_t^2(\psi)}{2\sigma_u^2}\right) \tag{3.25}$$

where ν represents the number of degrees of freedom, thus $\theta_1 := [\omega\ \beta\ \gamma\ \nu]'$ and $\Gamma(\cdot)$ is the Gamma function. Therefore by exploiting the same procedure used for the Gaussian specification, the joint log-likelihood function at the t-th observation is

$$l_t(\psi) = log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \frac{1}{2}log(\pi(\nu-2)) - \frac{1}{2}log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{1}{2}log(\sigma_t^2(\theta)_1) -$$
$$-\left(\frac{\nu+1}{2}\right)log\left(1+\frac{\varepsilon_t^2(\theta_1)}{\sigma_t^2(\theta_1)(\nu-2)}\right) + \left[-\frac{1}{2}log(2\pi) - \frac{1}{2}log(\sigma_u^2) - \frac{1}{2}\frac{u_t^2(\psi)}{\sigma_u^2}\right]$$

where the first part is the log-likelihood of the log-return (i.e. the partial log-likelihood) and the second part is the log-likelihood of the realized measure of volatility (e.g. Realized Variance), and $\psi := [\theta_1\ \theta_2]'$.

3.4 HAR-RV

3.4.1 Description

The Heterogeneous Autoregressive model (HAR) was developed by the Italian econometrician Fulvio Corsi during his PhD at the Univeristy of Lugano, the model was then published in the 2009 in the Journal of Financial Econometrics.
The HAR is an additive cascade model, based on the idea of the so called Heterogeneous Market Hypothesis (HMH) presented by Muller et al. in the 1993. The HMH recognizes the presence of heterogeneity across traders (this view of financial markets can be also related to the fractal market hypothesis of Peters), this heterogeneity might arises because of differences in agents' endowments, time horizons, institutional constraints, risk profile, geographical locations, and so on.

The HAR model is focused on the heterogeneity originating from differences in the time horizons. Therefore agents with different time horizons perceive, react to, and cause different types of volatility components.
Indeed, is possible to discern three primary volatility components based respectively on: the short-term traders, with daily or higher trading frequency, the medium-term investors who typically rebalance their positions weekly, and the long-term agents with a characteristic time of one or more months.

Moreover, it worth pointing out that it has been observed that volatility over longer time intervals has a stronger influence on volatility over shorter time intervals than vice versa.
The overall pattern that emerges is a volatility cascade from low frequencies to high frequencies. Economically the explanation is that for short-term traders the level of long-term volatility matters because it determines the expected future size of trends and risk.
Thus, on the one hand, short-term traders react to changes in long-term volatility by reviewing their trading behaviour and thus causing short-term volatility. On the other hand, the level of short-term volatility does not affect the trading strategies of long-term traders.

This hierarchical structure induced some researchers to propose an analogy between FX dynamics and the motion of turbulent fluid, in which a multiplicative energy cascade from large to small spatial scales is present [4].

Let $\widetilde{\sigma}_t^{(\cdot)}$ be the *latent partial volatility*, which is the volatility generated by a certain market component, the HAR is an additive cascade of partial volatilities each having an AR(1) structure. Actually the process is not a true AR(1), because on the right-hand side there is not the lagged latent volatility, but rather the corresponding realized volatility thus it belongs to the class of hidden Markov models, however the realized volatility is a close proxy of the latent one therefore the process is close to an AR(1).
The hierarchical model considered has only three components corresponding to one day, one week and one month horizons, respectively denoted as $\widetilde{\sigma}_t^{(d)}$, $\widetilde{\sigma}_t^{(w)}$, $\widetilde{\sigma}_t^{(m)}$, with $\widetilde{\sigma}_t^{(d)} = \sigma_t^{(d)}$ which is the daily integrated volatility. Therefore the return process is, as stated for the GARCH family models,

$$r_t = \sigma_t^{(d)} z_t$$

with $z_t \sim D(0,1)$ be the i.i.d. standardized innovation.

[4] Inspired by the Kolmogorov model of hydrodynamic turbulence, multiplicative cascade processes for volatility have been proposed by Ghashaghaie et a. (1996) and Breymann et al. (2000). Albeit these models are able, in theory, to reproduce the main feature of financial data, their empirical estimation is still an open question

The model for the unobserved volatility process $\widetilde{\sigma}_t^{(\cdot)}$ at each level of the cascade is a function of the past realized volatility occurred at the same time scale (i.e. the AR(1) component) and, because of the asymmetric propagation of volatility, it is also function of the expectation of the next-period values of the longer-term partial volatilities. Therefore the model reads

$$\widetilde{\sigma}_{t+1m}^{(m)} = c^{(m)} + \phi^{(m)} RV_t^{(m)} + \widetilde{\omega}_{t+1m}^{(m)}$$

$$\widetilde{\sigma}_{t+1w}^{(w)} = c^{(w)} + \phi^{(w)} RV_t^{(w)} + \gamma^{(w)} \mathbb{E}_t[\widetilde{\sigma}_{t+1m}^{(m)}] + \widetilde{\omega}_{t+1w}^{(w)}$$

$$\widetilde{\sigma}_{t+1d}^{(d)} = c^{(d)} + \phi^{(d)} RV_t^{(d)} + \gamma^{(d)} \mathbb{E}_t[\widetilde{\sigma}_{t+1w}^{(w)}] + \widetilde{\omega}_{t+1d}^{(d)}$$

where $RV_t^{(d)}$, $RV_t^{(w)} = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i+1}^{(d)}$, $RV_t^{(m)} = \frac{1}{22}\sum_{i=1}^{21} RV_{t-i+1}^{(d)}$ are respectively the daily, weekly and monthly ex-post realized volatilities, whereas $\widetilde{\omega}_{t+1m}^{(m)}$, $\widetilde{\omega}_{t+1w}^{(w)}$ and $\widetilde{\omega}_{t+1d}^{(d)}$ are the contemporaneously and serially independent zero-mean nuisance variates, with truncated left tail to guarantee the positivity of partial volatilities.

Each volatility component in the cascade corresponds to a market component that forms expectations for the next period's volatility based on the observation of the current realized volatility and on the expectation for the longer horizon volatility.
By recursive substitution of the partial volatilities, and recalling that $\widetilde{\sigma}_t^{(d)} = \sigma_t^{(d)}$, the cascade model can be expressed as

$$\sigma_{t+1d}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \widetilde{\omega}_{t+1d}^{(d)}. \tag{3.26}$$

The equation 3.26 can be seen as a three-factor stochastic volatility model, in which the factors are past realized volatilities observed at different frequencies. Moreover, the latent daily volatility $\sigma_{t+1d}^{(d)}$ can be written as

$$\sigma_{t+1d}^{(d)} = RV_{t+1d}^{(d)} + \omega_{t+1d}^{(d)} \tag{3.27}$$

where $\omega_{t+1d}^{(d)}$ represents both latent daily volatility and estimation error.
The last equation links the latent daily volatility $\sigma_{t+1d}^{(d)}$ with the ex-post volatility measure $RV_{t+1d}^{(d)}$. Combining together equation 3.26 and 3.27, the cascade model has an elegant representation and noting that $1d = 1$ in usual notation it can be written as

$$RV_{t+1}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+1} \tag{3.28}$$

where $\omega_{t+1} := \widetilde{\omega}_{t+1d}^{(d)} - \omega_{t+1d}^{(d)}$.

Equation 3.28 has a true autoregressive structure, it is an AR(22) restricted to only four parameters. It is important to precise that even though the HAR theoretically belongs to the class of short memory models (unlike for instance the FIGARCH), it is able to empirical reproduce the long memory behaviour of volatility, moreover the HAR is also able to reproduce the stylized facts of log-returns (i.e. excess kurtosis, cross-over from fat tail to thin tail).

Usually the HAR model is specified in logarithmic transformation for two reasons, first because the log-RV is approximately Gaussian and also because the log-transformation ensures positivity, therefore equation 3.28 becomes

$$logRV_{t+1}^{(d)} = c + \beta^{(d)} logRV_t^{(d)} + \beta^{(w)} logRV_t^{(w)} + \beta^{(m)} logRV_t^{(m)} + \omega_{t+1}. \tag{3.29}$$

However as discussed previously for some asset classes the leverage effect is crucial, and an advantage of the HAR model is that its structure is very flexible and allows to add in a straightforward fashion some components able to account for the leverage effect.
By following the approach of Corsi and Renò (2012) the LHAR in log-linear specification can be written as

$$logRV_{t+1}^{(d)} = c + \beta^{(d)} logRV_t^{(d)} + \beta^{(w)} logRV_t^{(w)} + \beta^{(m)} logRV_t^{(m)} +$$

$$+\gamma^{(d)} r_t^{-} + \gamma^{(w)} r_t^{(w)-} + \gamma^{(m)} r_t^{(m)-} + \omega_{t+1} \tag{3.30}$$

where $r_t^{(h)-} := min(r_t^{(h)}, 0)$ and $r_t^{(h)} = \frac{1}{h} \sum_{i=1}^{h} r_{t-i+1}$, and of course for the weekly component $h = 5$ and for the monthly one $h = 22$.
The parameters γ are negative and the parameters β are positive, in order to guarantee a proper behaviour of the model (the signs are met in empirical estimation).

3.4.2 Estimation

The HAR model belongs to the autoregressive structure processes, therefore there are several ways in which it can be estimated. However a robust and straightforward approach is the ordinary least square (OLS), where the linear model $Y = X\beta + \varepsilon$ for the plain logarithmic HAR is specified as

$$\begin{bmatrix} logRV_{p+1}^{(d)} \\ logRV_{p+2}^{(d)} \\ \vdots \\ logRV_{T}^{(d)} \end{bmatrix} = \begin{bmatrix} 1 & logRV_{p}^{(d)} & logRV_{p}^{(w)} & logRV_{p}^{(m)} \\ 1 & logRV_{p+1}^{(d)} & logRV_{p+1}^{(w)} & logRV_{p+1}^{(m)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & logRV_{T-1}^{(d)} & logRV_{T-1}^{(w)} & logRV_{T-1}^{(m)} \end{bmatrix} \cdot \begin{bmatrix} c \\ \beta^{(d)} \\ \beta^{(w)} \\ \beta^{(m)} \end{bmatrix} \cdot \begin{bmatrix} \omega_{p+1} \\ \omega_{p+2} \\ \vdots \\ \omega_{T} \end{bmatrix} \tag{3.31}$$

where Y is the vector of dependent variables, X is the matrix of covariates, β is the vector of parameters to be estimated and ε is the vector of errors.
For the HAR model the number of lags is $p = 22$, and given that the log-RV is nearly Gaussian the error term is assumed to be Normal, $\omega_t \sim N(0, \sigma_{\omega}^2)\ \forall t$.

The OLS problem is an unconstrained minimization in which the objective function is a distance function, thus the problem can be written as

$$\widehat{\beta}_{OLS} := \underset{\beta \in B}{\text{argmin}} \parallel Y - X\beta \parallel_2^2 \tag{3.32}$$

where the objective function is the $l2$ norm squared, therefore the rational is to find that vector of coefficients β for which the linear combination $X\beta$ belongs to the column space of X, $col(X)$, and is at the same time as close as possible to Y, that vector turns out to be $\widehat{\beta}_{OLS}$. Moreover the vector $X\widehat{\beta}_{OLS}$ is the orthogonal projection of Y onto $col(X)$, usually denoted as $\widehat{Y}$.

Then from equation 3.32 the first order conditions (FOC) becomes

$$\nabla f_d(\beta) = X'(Y - X\beta) = \mathbf{0} \tag{3.33}$$

where $\nabla f_d(\beta) := \frac{\partial f_d(\beta)}{\partial \beta}$ is the gradient vector of the distance function $f_d :=\parallel Y - X\beta \parallel_2^2$.
The set of equations in 3.33 forms a square system of linear equations also known as Normal Equations. Moreover it worth noting that the hessian matrix of the distance function $\nabla^2 f_d(\beta) := \frac{\partial^2 f_d(\beta)}{\partial \beta \partial \beta'} = 2(X'X)$, where the second moment matrix $(X'X)$ is a positive definite matrix[5] therefore the objective function is globally convex and is easy to handle.
The solution of the set of Normal Equations that gives the parameters estimate is the linear projection, therefore the expression is

$$\widehat{\beta}_{OLS} = (X'X)^{-1}X'Y$$

The estimation is the same for the LHAR model, the only difference is the number of parameters and the matrix of data (X), because the lagged returns and average returns have to be considered, augmenting the dimensions of X and β but the result is exactly the same.

[5]The matrix $(X'X)$ is a positive definite matrix whenever $rank(X) = k$ where k is the number of regressors (i.e. the number of columns), if X has a rank lower than k then there is a problem of multicollinearity (i.e. some column vectors are linearly dependent) thus an identification issue arises because the second moment matrix is no longer invertible.

3.5 Testing framework

The current Basel Regulation allows financial institutions to develop their own internal models of risk management and computation of Value-at-Risk. Therefore, nowadays plays a pivotal role not only the estimation approach, but also the backtesting procedures of VaR forecasts (see Jorion, 2001).
However, the Regulatory Authorities provide no particular backtesting technique.

The two main approaches for evaluating VaR forecasts are respectively focused on assessing the accuracy of the projections and their efficiency via a loss function (which is different form the concept of conditional efficiency defined below). The accuracy approach is based on testing two fundamental hypotheses regarding the process of VaR violations for a given coverage rate, the hypothesis of unconditional coverage and the hypothesis of independence (see Christoffersen, 1998). It worth noting that a violation is a situation in which the ex-post portfolio returns are lower that the VaR forecasts.
Therefore the hypothesis of unconditional coverage means, for instance, that a 95% VaR is valid whenever the expected frequency of observed violations is equal to 5% (i.e. the coverage rate). Whereas the independence hypothesis means that the VaR is valid whenever the violations are distributed independently, thus there must be no violation cluster that is, the occurrence of a loss violation does not contain informations useful to forecast future violations.

The aforementioned two hypothesis have been synthesize by Berkowitz et al. (2005), by showing that the violation process centred on the coverage rate, known as Hit function, is a martingale difference sequence (MDS), namely $\mathbb{E}[I_t(\alpha) - \alpha \mid \mathscr{F}_{t-1}] = 0$.

Let denote with r_t the return of an asset or a portfolio at generic time t, and with $VaR_{t+1|t}(\alpha)$ the forecasted VaR for an $\alpha\%$ coverage rate conditioned on an information set $\mathscr{F}_t$.
Then the violation process associated to the ex-post observation of an $\alpha\%$ VaR at time t and labeled as $I_t(\alpha)$ is defined in the following way

$$I_t(\alpha) = \begin{cases} 1 & \text{if } r_t < VaR_{t|t-1}(\alpha) \\ 0 & \text{if } r_t \geq VaR_{t|t-1}(\alpha) \end{cases} \tag{3.34}$$

Thus the problem of the validity of the Value-at-Risk becomes, under the accuracy standpoint, knowing whether the violation sequence $\{I_t\}_{t=1}^{T}$ is in line with the two hypotheses.
The unconditional coverage hypothesis is that the probability of an ex-post loss exceeding VaR forecasts has to be equal to the coverage rate, namely

$$\mathbb{P}(I_t(\alpha) = 1) = \mathbb{E}[I_t(\alpha)] = \alpha. \tag{3.35}$$

The independence hypothesis is based on the fact that VaR violations observed at two different points in time and for the same coverage rate have to be distributed independently, which means

$$I_t(\alpha) \perp I_s(\alpha) \quad \forall\, s \neq k. \tag{3.36}$$

this property is valid for any variable in the information set, and for any other coverage rate $\beta \in]0,1[$.

The first hypothesis is intuitive, indeed if the probability associated to the event $I_t(\alpha) = 1$

denoted as $\pi_t := \mathbb{P}(I_t(\alpha) = 1)$ and assessed by the frequency of violations observed over a time T, namely $fr := T^{-1}\sum_{t=1}^{T} I_t(\alpha)$ is significantly lower or higher than the α nominal coverage rate then it shows an overestimation or underestimation of the VaR and hence too few or too many violations.
The first test of unconditional coverage has been developed by Paul Kupiec (1995), the aim is to test whether the frequency of violations (fr) is statistically equal to the coverage rate. The violation is assumed to follow an i.i.d. Bernoulli distribution, $I_t \sim i.i.d.Ber(\alpha)$ with $I_t \in \{0,1\}$ and $\alpha \in [0,1]$ is the probability of success (i.e. the coverage rate), therefore the sequence of violations is a Bernoulli trial [6] and its mass function over the full sample is the Binomial one $\mathbb{P}(k) = \binom{T}{k}\alpha^k(1-\alpha)^{(T-k)}$, where T is the sample size and k is the number of violations over the sample.
Thus the Kupiec test is naturally in the class of likelihood ratio tests, where the unrestricted and the restricted log-likelihood functions (i.e. under the H_1 and the H_0) are, without considering the binomial coefficient, respectively

$$\begin{aligned} L_U &= log fr^k(1-fr)^{(T-k)} \\ L_R &= log \alpha^k(1-\alpha)^{(T-k)} \end{aligned} \tag{3.37}$$

therefore the likelihood ratio statistic is $LR = 2*[L_U - L_R] \xrightarrow{d} \chi^2_{(1)}$.
The hypotheses in the Kupiec test are

$$\begin{aligned} H_0 &: fr = \alpha \\ H_1 &: fr \neq \alpha \end{aligned} \tag{3.38}$$

where the null means that the unconditional coverage is met, thus the model is correctly specified, whilst the alternative implies that there is no unconditional coverage and therefore the model is misleading.

The Kupiec test is clearly incomplete because it does not take into consideration the independence hypothesis, indeed it asses only the unconditional coverage and not the so called conditional efficiency which is met when also independence in the violation process occurs, therefore the conditional efficiency is met when the violation process is a martingale difference, namely whenever $\mathbb{E}[I_t(\alpha) - \alpha \mid \mathscr{F}_{t-1}] = 0$ where the information set includes past violations but in case also other variables such as past VaR levels, returns, etc. The two tests for the conditional efficiency considered in this work are the Christoffersen test (1998) and the Engle and Manganelli test (2004), actually there are several other tests in literature.

The Christoffersen test is based on a peculiar alternative hypothesis of VaR inefficiency, indeed is assumed that the process of $I_t(\alpha)$ violations is a first order Markov chain whose probabilities transition matrix is defined as

$$\Pi = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}$$

where $\pi_{ij} := \mathbb{P}(I_t(\alpha) = j \mid I_{t-1}(\alpha) = i)$, thus for instance π_{00} is the probability of getting no

[6] Bernoulli trial is an experiment where a certain action is repeated many times. Each time the process has two possible outcomes, either success or failure. The probabilities of the outcomes are the same in every trial, i.e. the repeated actions must be independent of each other.

violation given that in the previous time step no violations occurred.
The Markov chain reflects the existence of an order one memory in the process of violations, indeed the probability of getting a violation in the current period depends upon the occurrence or not of a violation in the previous period.

The null hypothesis of conditional efficiency is therefore defined as

$$H_0 : \Pi = \Pi_\alpha = \begin{bmatrix} 1-\alpha & \alpha \\ 1-\alpha & \alpha \end{bmatrix}$$

Thus if the test is not rejected, then both unconditional coverage and one period independence are met. Likewise the Kupiec test, also the Christoffersen one is defined in terms of a likelihood ratio test, namely $LR = 2 * [L_U - L_R] \xrightarrow{d} \chi^2_{(2)}$ (in this case there are two restrictions), where the log-likelihood functions are

$$L_U = log[(1-\widehat{\pi}_{01})^{(T_{00})}(\widehat{\pi}_{01})^{(T_{01})} \cdot (1-\widehat{\pi}_{11})^{(T_{10})}(\widehat{\pi}_{11})^{(T_{11})}]$$

$$L_R = log[(1-\alpha)^{(T_{00}+T_{10})}\alpha^{(T_{01}+T_{11})}]$$

where, T_{ij} is the number of times $I_t(\alpha) = j$ and $I_{t-1}(\alpha) = i$ occurred (for instance, T_{00} is the number of times no violations occurred given that in the previous period no violation has occurred), and $\widehat{\pi}_{ij}$ is the maximum likelihood estimator of the transition probability.
Therefore the hypotheses of the Christoffersen test are formally defined as

$$H_0 : \mathbb{E}[I_t(\alpha) \mid \mathscr{F}_{t-1}] = \alpha$$
$$H_1 : \mathbb{E}[I_t(\alpha) \mid \mathscr{F}_{t-1}] \neq \alpha$$

where the null represents the conditional efficiency (i.e. the violation process is a MDS) and the alternative represents the lack of efficiency, hence if the test is not rejected then the model is correctly specified and the other way around. The main drawback of the Christoffersen test is that the independence is tested against a first order Markov chain, therefore the test does not take into account higher order forms of dependence.

Engle and Manganelli in 2004 proposed a test called Dynamic Quantile (DQ) able to cope with the Christoffersen test drawback, the DQ test is based on a linear regression model linking current violations to past violations so as to test the conditional efficiency hypothesis. Therefore let $Hit(\alpha)$ be the de-meanded violation process on α, thus specified as

$$Hit_t(\alpha) = \begin{cases} 1-\alpha & \text{if } r_t < VaR_{t|t-1}(\alpha) \\ -\alpha & \text{if } r_t \geq VaR_{t|t-1}(\alpha) \end{cases}$$

Then the $Hit_t(\alpha)$ is modelled via a linear regression model, therefore it can be expressed in the following way

$$Hit(\alpha) = c + \sum_{k=1}^{K} \beta_k Hit_{t-k}(\alpha) + \sum_{k=1}^{K} \gamma_k g\left(Hit_{t-k}(\alpha), Hit_{t-k-1}(\alpha), ..., z_{t-k}, z_{t-k-1}, ...\right) + \varepsilon_t \tag{3.39}$$

where ε_t is an i.i.d. process playing the role of error and $g(\cdot)$ is a function of past violations and a set of variables z_{t-k} belonging to the information set $\mathscr{F}_{t-1}$.

For instance $g(\cdot)$ can be function of past returns r_{t-k}, squared past returns r_{t-k}^2, past values of VaR forecasts $VaR_{t-k|t-k-1}(\alpha)$, etc. Regardless of the specification of the set of regressors in the right-hand side of $Hit_t(\alpha)$ the null hypothesis test of conditional efficiency corresponds to testing the joint nullity of coefficients β_k and γ_k and the constant c (i.e. all the covariates have no explanatory power) , namely

$$H_0 : c = 0 \cap \beta_1 = 0 \cap ... \cap \beta_k = 0 \cap \gamma_1 = 0 \cap ... \cap \gamma_k = 0$$

$$H_1 : c \neq 0 \cup \beta_1 \neq 0 \cup ... \cup \beta_k \neq 0 \cup \gamma_1 \neq 0 \cup ... \cup \gamma_k \neq 0$$

Therefore the unconditional coverage hypothesis is verified when $c = 0$, and the current VaR violations are uncorrelated to past violations whenever $\beta_1 = ... = \beta_k = \gamma_1 = ... = \gamma_k = 0$. Indeed, under the null hypothesis $\mathbb{E}[Hit_t(\alpha)] = \mathbb{E}[\varepsilon_t] = 0$ which implies by construction that $\mathbb{P}(I_t(\alpha) = 1) = \mathbb{E}[I_t(\alpha)] = \alpha$, thus the joint nullity test of all coefficients, including the constant, corresponds to a conditional efficiency test.

Let $\Psi := [c\ \beta_1\ ...\ \beta_k\ \gamma_1\ ...\ \gamma_k]'$ be the $2K+1$ vector of parameters and Z the matrix of covariates, and noting that under the null hypothesis the $Hit_t(\alpha)$ process follows a Bernoulli distribution, $Hit_t(\alpha) \sim Ber(\alpha)$, then the variance of Hit_t is $Var(Hit_t(\alpha)) = \alpha(1-\alpha)$. Thus a Wald statistic can be used to test the simultaneous nullity of the coefficients (i.e. $H_0 :\ Z\Psi = \mathbf{0}$), so the statistic is an $l2$ norm squared and written around the inverse of the variance of $Hit_t(\alpha)$, namely

$$DQ = \| Z\widehat{\Psi} \|^2_{2,Var(Hit_t(\alpha))^{-1}} := \left[\frac{\widehat{\Psi}' Z' Z \widehat{\Psi}}{\alpha(1-\alpha)} \right] \xrightarrow{d} \chi^2_{(2K+1)}$$

where $2K+1$ is not only the number of covariates, but also the number of restrictions.
A natural extension of the Engle and Manganelli test is to consider, instead of a linear regression model, a binary model (e.g. Logit or Probit) due to the dichotomic nature of the dependent variable. Indeed, Andrew Patton in his PhD thesis at the Univeristy of California San Diego proposed a LR test based on the Logit model linking the violation probability at time t to the set of explanatory variables Z_t. However diving deeper into the testing framework of the conditional efficiency hypothesis is not in the aim of this work.

The other approach to evaluate the VaR forecasts is the one based on assessing the efficiency of the forecasts in terms of a loss function, that of course has to be as small as possible. Therefore the interest is posed not only on the number of failures but also on the magnitude of these failures.
Thus this approach has a nice economic interpretation, because the value of the VaR enters into the computation of the regulatory Market Risk Capital, hence the institution has the incentive of getting a VaR model with not too many violations, because of the traffic light system [7], but at the same time the institution does not want a model delivering to large VaR values because the market risk capital is proportional to these values. Thus a good trade-off between accuracy and efficiency is usually searched, indeed the grater the market risk capital, the greater is the opportunity cost because the capital is frozen for risk management purposes.

In this thesis two different loss functions are used, first the Quadratic Loss Function (QLF)

[7] The traffic light system is based on add a multiplier to the VaR value, based on the number of violations, therefore given the coverage rate (usually 1%) the more violations the greater is the multiplier and the greater is the market risk capital.

devised by Lopez (1999) which considers both the number of violations and their magnitude, computed as the square of the distance between the returns and the VaR estimate, namely

$$QLF_t(\alpha) = \begin{cases} 1 + (r_t - VaR_{t|t-1}(\alpha))^2 & \text{if } r_t < VaR_{t|t-1}(\alpha) \\ 0 & \text{if } r_t \geq VaR_{t|t-1}(\alpha) \end{cases}$$

Therefore large violations are penalized, however albeit the model with the smallest average QLF is deemed to be the most accurate one, the metric tend to favor models that are too conservative and their number of violations is well below the prescribed coverage rate.

Sarma et al. (2003) introduced the so called Firm Loss Function (FLF), in which the non-violation days are penalized as well according to the opportunity cost of the reserved capital held by the firm for risk management purposes, thus the FLF is defined as

$$FLF_t(\alpha) = \begin{cases} 1 + (r_t - VaR_{t|t-1}(\alpha))^2 & \text{if } r_t < VaR_{t|t-1}(\alpha) \\ -kVaR_{t|t-1}(\alpha) & \text{if } r_t \geq VaR_{t|t-1}(\alpha) \end{cases}$$

where k is the firm's cost of capital. Therefore, an otherwise accurate model producing a limited number of small magnitude violations might be highly inefficient as high daily VaR estimates entail additional opportunity cost, thus the FLF penalizes too conservative models.

It is important to point out that there is not a perfect way to approach the VaR computation, unless unfeasibility cases, therefore the institution chooses the approach and the specification that finds the best (subjective) midpoint between statistical accuracy and efficiency, however a question could be whether using high-frequency data can improve the overall performance of VaR forecasts compared to using daily data only, of course the approach to address this question is the variance-covariance approach with several models and specifications.

Chapter 4

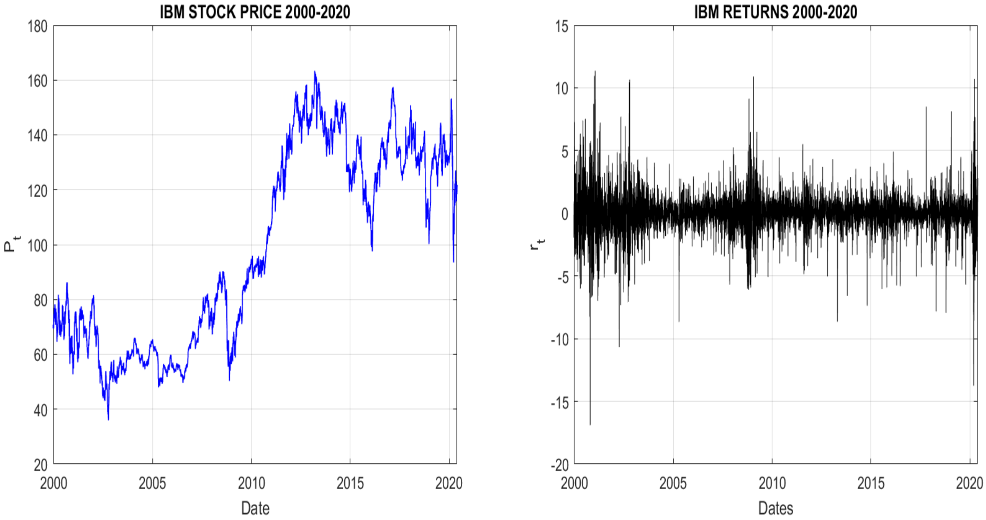
Empirical Analysis

In this chapter an attempt to reply to the question at the end of chapter 3 is carried out, therefore several models using either daily data or daily and high-frequency data are taken into consideration, and for each of these models two distribution assumptions with respect to the log-returns are made. The distribution assumptions are the Gaussianity of log-returns, namely the baseline assumption, and the Student-t which allows to capture the fat tail phenomenon of returns, moreover also accounting for the leverage effect is evaluated.

4.1 Data Description

The data used for the empirical analysis regards the American Blue Chip IBM. The International Business Machines Corporation (IBM or Big Blue) founded by Charles Flint, is one of the oldest and biggest companies of the tech sector. It has more than 352,000 employees, 70% of them are based outside the USA. The Big Blue employees have been awarded five Nobel Prizes, ten National Medals of technology, six Turing Awards and five National Medal of Science.
The IBM produces and sells many types of products, such as software, hardware, middleware, services as hosting, cloud computing and artificial intelligence. The Big Blue was and still is a key player in the innovation game, it created the first modern calculator, namely the Mark 1, the first Personal Computer with the DOS system (made by Microsoft), the first floppy disk, the magnetic stripe card, the programming languages FORTRAN and SQL, the UPC barcode and the list could go on.

IBM has ever had a strategic view, thus shifting business operations focusing on higher-value markets. Indeed, IBM sold the personal computer and X86-based server business to Lenovo (in 2005 and 2014 respectively), it acquired companies such as Pwc Consulting (2002), SPSS (2009), The Weather Company (2016) and Red Hat (2019). Therefore is interesting to take a look to the data of the Big Blue, the left panel in figure 4.1 shows the price over time and it can be noted a systematic increase starting from 2009 (with the acquisition of SPSS). Moreover a large fall in the price is occurred recently due to the Coronavirus, indeed the price is now around the level it had at the late 2010, from an econometric point of view it has also to be stated that the price is a non-stationary stochastic process (as assessed by the PP test). The right panel in figure 4.1 instead shows the log-returns chart, it can be seen that returns are stationary and there are clusters of volatility with large peaks in 2000-2003 dur-



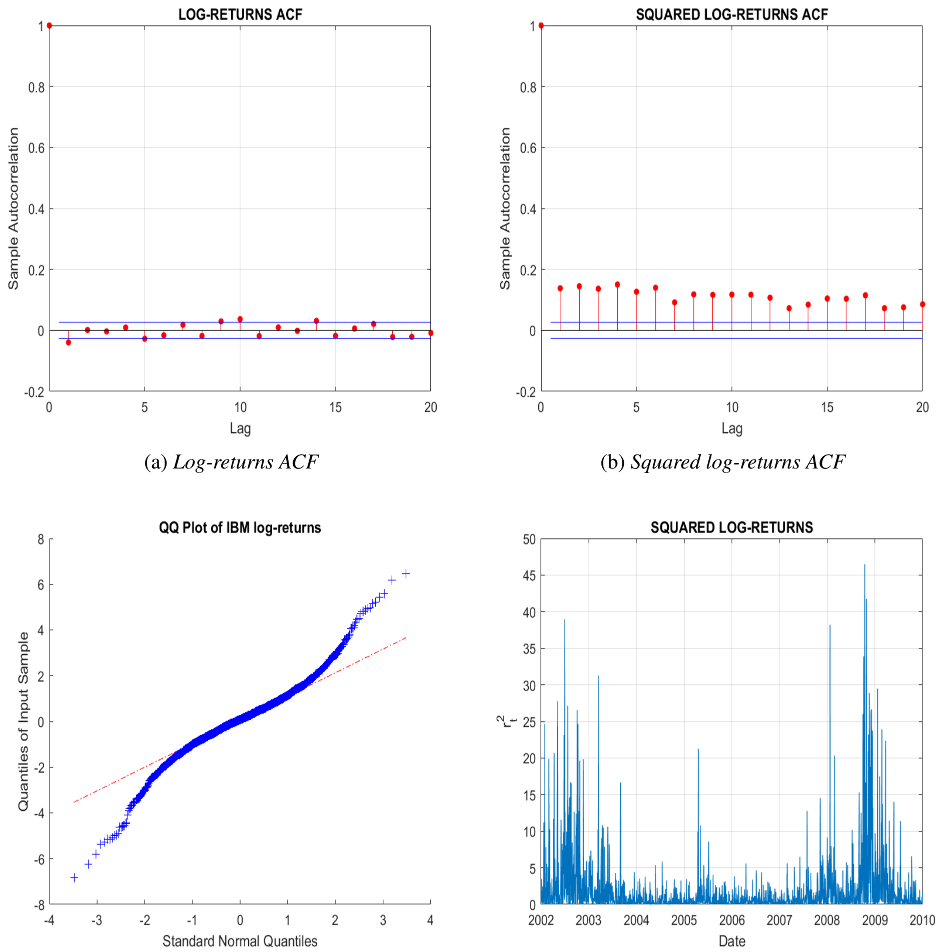
(a) *IBM stock price 2000-2020* (b) *IBM log-return 2000-2020*

Figure 4.1: IBM stock price and logarithmic return from 2000 to 2020

ing the dot-com bubble and the irrational exuberance (see the book of Robert Shiller called ”Irrational Exuberance”), clusters and peaks arose also in the 2008-2009 during the financial crisis and also at the beginning of the 2020 during the Coronavirus.

Proceeding in the data investigation the panels in the first row of figure 4.2 show two important stylized facts regarding log-returns, the left panel shows the lack of autocorrelation in the returns series, whilst the right panel displays the presence of serial correlation in squared log-returns, this is the reason why trying to forecast returns volatility makes sense. Moreover the panels in the second row of figure 4.2 show two other important features, the bottom left panel clearly shows that log-returns are not Gaussian because of the presence of fat tails, and the bottom right panel shows the series of squared log returns that can be thought as a first noisy realized measure of volatility.

Regarding the Realized Variance figure 4.3 shows the RV of the IBM stock over time, and as stated above peaks occur during the period 2002-2003 and of course during the financial crisis, these periods are thus characterized by an high investment risk. Furthermore, figure 4.4 shows some interesting stylized facts of the RV, the top left panel shows the QQ plot for the log-returns standardized by the ex-post RV, they are very close to be Gaussian as endorsed by their histogram. While the bottom left panel in figure 4.3 shows the QQ plot of the log-RV, which turns out to be approximately Gaussian as endorsed by the corresponding histogram. The dataset used for the empirical analysis goes from 1 February 2002 to 31 December 2009, therefore there are 2014 observations in the dataset. The Realized Variance is computed by using 5-minutes prices, the choice of the frequency is an heuristic choice in line with a cut-off able to reduce the microstructure noises (see section 2.2).



(a) *Log-returns ACF*

(b) *Squared log-returns ACF*

(c) *Log-returns QQ plot*

(d) *Squared log-returns*

Figure 4.2: Different features of the IBM stock returns

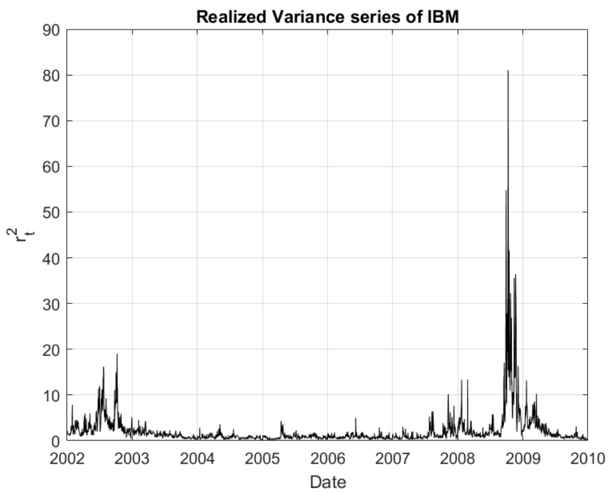
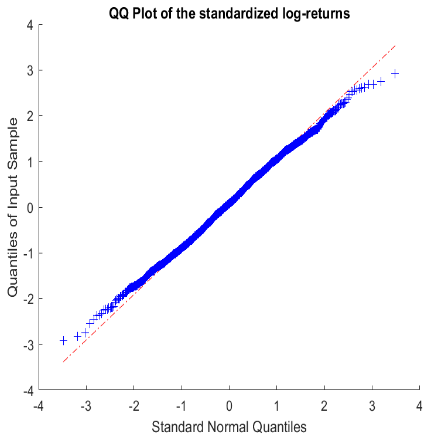
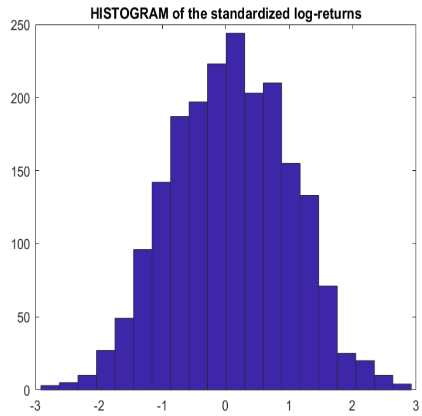


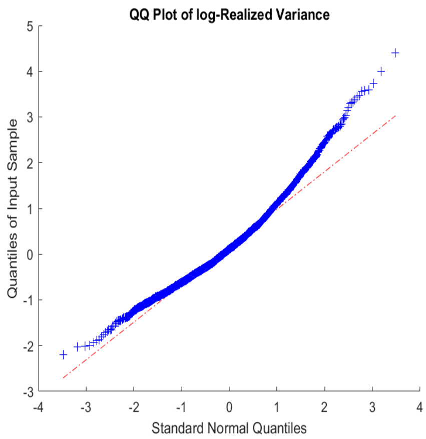
Figure 4.3: IBM Realized Variance from 2002 to 2009



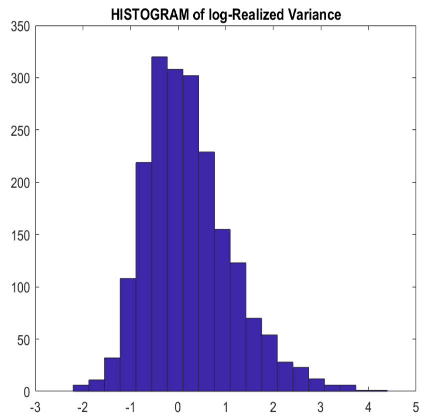
(a) *Standardized log-returns QQ plot*



(b) *Standardized log-returns Histogram*



(c) *Log-RV QQ plot*



(d) *Log-RV Histogram*

Figure 4.4: Different features of the IBM stock RV

4.2 Models Set Up

Before setting up the models two important statistical tests have been carried out on the log-returns. First the Ljung-Box test in order to asses whether log-returns are correlated over time and they turn out not to be correlated, while their squared values are. Then the ARCH test has been performed in order to asses whether there is conditional heteroskedasticity in log-returns, and they turn out to be heteroskedastic. Thus, the statistical tests endorsed exactly some of the stylized facts described previously.

The dataset is composed of 2014 observations and the estimation is performed in a rolling window basis with a window length of 1000 observations, thus approximately four trading years.

The HAR-RV models need 22 observations (the number of lags) in order to be initialized, therefore for the other models 22 observations are dropped from the in-sample dataset and at the end the one-step ahead forecasts composing the out-of sample dataset are 992.

Four models are analyzed: GARCH, GJR-GARCH, HAR-RV, Realized-GARCH, for each model two distributional assumptions are made, first the Gaussian distribution and then in order to capture the fat tail phenomenon the Student-t distribution.

For the GARCH family models a model selection procedure based on the BIC information criterion has been carried out, where $BIC = -2l(\widehat{\psi}) + klogT$ and $l(\widehat{\psi})$ is the log-likelihood function of the model in the estimated parameters, k represents its number of parameters and T is the sample size.

The model selection procedure for the return equation showed an optimal lag length of zero for both the autoregressive and the moving average parts, therefore the return equation has no conditional mean and is thus represented by a weak White Noise only (as stated below). While the model selection procedure for the variance equation showed an optimal lag length of one for the ARCH and GARCH parts for all the models GARCH, GJR-GARCH and RealizedGARCH.

The first model to be estimated is a GARCH(1,1) with the following specification

$$r_t = \varepsilon_t \quad where \;\; \varepsilon_t := \sqrt{\sigma_t^2} z_t$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{4.1}$$

The conditional densities assumed are first the Gaussian and then the Student-t, namely

$$f(r_t \mid \mathscr{F}_{t-1}; \psi) = \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \quad given \;\; z_t \sim N(0,1)$$

$$f(r_t \mid \mathscr{F}_{t-1}; \psi) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma_t \sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})} \left(1 + \frac{\varepsilon_t^2}{\sigma_t^2(\nu-2)}\right)^{-\frac{\nu+1}{2}} \quad given \;\; z_t \sim t_\nu(0,1). \tag{4.2}$$

Where for the Gaussian specification the vector of parameters is $\psi := [\omega \; \alpha \; \beta]'$, whilst for the Student-t specification also the number of degrees of freedom has to be estimated imposing $\nu > 2$ for the existence of the second moment, hence the parameters vector becomes $\psi := [\omega \; \alpha \; \beta \; \nu]'$.

For the GJR-GARCH(1,1) what changes is only the functional form of the variance equation in order to account for the leverage effect in the stock returns, therefore the conditional variance becomes

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma I_{t-1}^- \varepsilon_{t-1}^2$$

where the function I_t^- is an indicator function taking the value 1 if $r_t < 0$ and 0 otherwise. For the Gaussian GJR-GARCH(1,1) the vector of parameters is $\psi := [\omega\ \alpha\ \beta\ \gamma]'$, whereas for the Student-t specification also the parameter ν enters into the vector ψ in the same fashion as the GARCH(1,1).

For the GARCH family models the one-step ahead forecast is immediate because all the variables needed are $\mathscr{F}_t$-measurable, therefore it is

$$\sigma_{t+1|t}^2 = \widehat{\omega} + \widehat{\alpha}\varepsilon_t^2 + \widehat{\beta}\sigma_t^2$$

$$\sigma_{t+1|t}^2 = \widehat{\omega} + \widehat{\alpha}\varepsilon_t^2 + \widehat{\beta}\sigma_t^2 + \widehat{\gamma} I_t^- \varepsilon_t^2$$

where the parameters are estimated in each rolling window based on the in-sample data. Initial assumptions on the values of σ_1^2 and ε_1 are needed in order to use the Maximum Likelihood. A convenient choice is to set the conditional variance equal to its unconditional value, given dummy starting values for the variance equation parameters (necessary to initialize the Newton-Raphson algorithm), and zero for the value of ε.

For what concern the HAR-RV model a two step estimation[1] is carried out and four different specifications are taken into consideration, a Gaussian and Student-t HAR and a Gaussian and Student-t LHAR are estimated in log-linear form. Therefore the variance equations are

$$logRV_t^{(d)} = c + \beta^{(d)} logRV_{t-1}^{(d)} + \beta^{(w)} logRV_{t-1}^{(w)} + \beta^{(m)} logRV_{t-1}^{(m)} + \omega_t$$

$$logRV_t^{(d)} = c + \beta^{(d)} logRV_{t-1}^{(d)} + \beta^{(w)} logRV_{t-1}^{(w)} + \beta^{(m)} logRV_{t-1}^{(m)} + \\ + \gamma^{(d)} r_{t-1}^- + \gamma^{(w)} r_{t-1}^{(w)-} + \gamma^{(m)} r_{t-1}^{(m)-} + \omega_t$$

respectively for the HAR and LHAR. The first step consists in estimating the vector of parameters $\psi := [c\ \beta^{(d)}\ \beta^{(w)}\ \beta^{(m)}]'$ for the HAR and $\psi := [c\ \beta^{(d)}\ \beta^{(w)}\ \beta^{(m)}\ \gamma^{(d)}\ \gamma^{(w)}\ \gamma^{(m)}]'$ for the LHAR model. The first step is performed via OLS.
The second step is necessary in order to allow the log-return to have other distributional assumption than the Gaussian one, therefore first the conditional realized variance is computed as

$$RV_{t|t-1} = exp(\mathbb{E}[logRV_t \mid \mathscr{F}_{t-1}])$$

where the vector $\widehat{\psi}$ is estimated in the first step.

Then in the second step the following model is estimated

$$r_t = \varepsilon_t \quad where\ \ \varepsilon_t := \sqrt{\sigma_t^2} z_t$$

$$\sigma_t^2 = \phi RV_{t|t-1} \tag{4.3}$$

where ϕ is an additional parameter to be estimated, thus from equation 4.3 it can be seen that specification is similar to the one in equation 4.1. However, in the second step of the HAR-RV models the variance equation shows a proportional relationship between the conditional

[1]The procedure is based on the Giot and Laurent work (2004), however they applied this method to the ARFIMAX model instead of the HAR-RV, hence some adjustments have been made.

variance and the conditional realized variance.
The estimation in the second step is performed via Maximum Likelihood, allowing the same distributional assumptions for the log-return in equations 4.2.
The one-step ahead forecast is obtained as

$$\sigma^2_{t+1|t} = \widehat{\phi} RV_{t+1|t}$$

where $\widehat{\phi}$ is obtained in the second step (it is usually close to one), and $RV_{t+1|t}$ is obtained in the first step as $RV_{t+1|t} = exp(\mathbb{E}[logRV_{t+1} \mid \mathscr{F}_t])$, where of course $logRV_{t+1}$ is $\mathscr{F}_t$-measurable.

The RealizedGARCH estimation is developed in a similar fashion as the GARCH models, thus a RealizedGARCH(1,1) specification without the leverage function and one with the leverage function are considered and the realized measure used is the RV, namely

$$r_t = \varepsilon_t \quad where \;\; \varepsilon_t := \sqrt{\sigma_t^2} z_t \quad and \;\; \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \gamma RV_{t-1}$$

$$RV_t = \phi_0 + \phi_1 \sigma_t^2 + u_t$$

$$RV_t = \phi_0 + \phi_1 \sigma_t^2 + \tau(z_t) + u_t \quad where \;\; \tau(z_t) := \tau_1 z_t + \tau_2(z_t^2 - 1). \tag{4.4}$$

where the first measurement equation does not contain the leverage function, whereas the second one does.
The distributional assumptions for the log-returns are the same of the equations 4.1, whilst the shock to the measurement equation is assumed to be a Gaussian White Noise, $u_t \sim NID(0, \sigma_u^2)$ independent of z_t. The estimation of the RealizedGARCH is performed via Maximum Likelihood, for the Gussianity assumption the vector of parameters is $\psi := [\omega\, \beta\, \gamma\, \phi_0\, \phi_1\, \sigma_u^2]'$ with no leverage and $\psi := [\omega\;\beta\;\gamma\;\phi_0\;\phi_1\;\sigma_u^2\;\tau_1\;\tau_2]'$ when the leverage function is added, as usual when the Student-t distribution is used also the number of degrees of freedom (ν) has to be estimated.
The one-step ahead forecast is obtained in the same way as the GARCH models because all the variables needed are $\mathscr{F}_t$-measurable,

$$\sigma^2_{t+1|t} = \widehat{\omega} + \widehat{\beta}\sigma_t^2 + \widehat{\gamma} RV_t$$

where the parameters are estimated, for each forecast, in a rolling window basis.

After estimating all the necessary parameters the Value-at-Risk forecast for a specific confidence level is obtained, depending on the distributional assumption, respectively via the Delta-Normal and Delta-Student formulas

$$VaR^{\alpha}_{t+1|t} = -\sigma_{t+1|t}\Phi^{-1}(\alpha)$$

$$VaR^{\alpha}_{t+1|t} = -\sigma_{t+1|t} t_{\nu}^{-1}(\alpha) \tag{4.5}$$

where the functions $\Phi^{-1}(\cdot)$ and $t_\nu^{-1}(\cdot)$ are respectively the quantile functions (in this case the inverse of the CDF) of the Standard Normal and Standard Student-t distributions.

4.3 Results

The models have first been estimated [2] entirely on the eight years dataset in order to look at the estimates at hand. Then the VaR forecasts are computed in a rolling window basis with a length of 1000 observations (in-sample) covering the period from 2002 to 2006, hence almost 1000 forecasts (out-of-sample) are computed and in particular they cover the period 2006-2009. Hence, the out-of-sample period includes both a turbulent period, namely the financial crisis and a calm period from 2006 to the end of 2007. Therefore, the VaR models are tested with respect to a calm and a turbulent period.
Table 5.1 shows the estimation results for the GARCH model and the GJR-GARCH with Gaussian and Student-t distribution assumption.

Par.	Gaus. GARCH(1,1)	Stud. GARCH(1,1)	Gaus. GJR(1,1)	Stud. GJR(1,1)
ω	0.0098 (0.0083)	0.0088 (0.0062)	0.0121 (0.0049)	0.0117 (0.0049)
α	0.0663 (0.0182)	0.0619 (0.0145)	0.0088 (0.0105)	0.0106 (0.0097)
β	0.9298 (0.0221)	0.9344 (0.0170)	0.9486 (0.0171)	0.9480 (0.0166)
ν	–	10.3997 (1.0719)	–	13.6048 (0.0173)
γ	–	–	0.0744 (0.173)	0.0722 (0.0173)

Table 4.1: Estimation results of the GARCH and GJR for both Gaussian and Student-t distributional assumptions on the IBM stock, the values in parenthesis are the standard errors.

From the estimation of the GJR it can be seen that the value of γ is always positive, thus allowing to properly capture the leverage effect. Moreover the p-values of γ make it significant, thus endorsing the relevance of the leverage effects.

In table 5.2 instead are shown the parameters estimate for the various HAR-RV models, it can be seen that all the parameters $\gamma^{(\cdot)}$ are negative in order to properly account for the leverage effect. Furthermore the parameters $\gamma^{(\cdot)}$ are, besides the monthly one, significant and therefore proving the relevance of the leverage effect.

Par.	HAR	LHAR
c	0.0051 (0.0105)	-0.0540 (0.0120)
$\beta^{(d)}$	0.2984 (0.0291)	0.2500 (0.0286)
$\beta^{(w)}$	0.4847 (0.0473)	0.4406 (0.0454)
$\beta^{(m)}$	0.1766 (0.0348)	0.2161 (0.0352)
$\gamma^{(d)}$	–	-0.0813 (0.0153)
$\gamma^{(w)}$	–	-0.1581 (0.0386)
$\gamma^{(m)}$	–	-0.0318 (0.0637)

Table 4.2: Estimation results of the HAR and LHAR on the IBM stock, the values in parenthesis are the standard errors.

Moreover, when performing the second step estimation the parameter ϕ is significant and

[2] All the estimates, the plots and the statistical tests in this chapter are performed by using the software MatLab. Moreover all the estimation procedures and the vast majority of tests have been implemented from scratch, without using any built-in function. The codes are available upon request.

close to 1, it is actually $\widehat{\phi} = 1.0068$ for the HAR and $\widehat{\phi} = 1.0036$ for the LHAR.

Regarding the RealizedGARCH(1,1) models table 5.3 shows the estimation results.

Par.	Gaus. Real(1,1)*	Gaus. Real(1,1)	Stud. Real(1,1)*	Stud. Real(1,1)
ω	0.0127 (0.0146)	0.0210 (0.0159)	0.0122 (0.0140)	0.0220 (0.0162)
β	0.7299 (0.0551)	0.7330 (0.0551)	0.7307 (0.0554)	0.7341 (0.0554)
γ	0.2395 (0.0489)	0.2289 (0.0483)	0.2391 (0.0499)	0.2266 (0.0490)
ϕ_0	0.1324 (0.1438)	0.0949 (0.1350)	0.1291 (0.1515)	0.0851 (0.1364)
ϕ_1	1.0306 (0.1106)	1.0687 (0.1030)	1.0299 (0.1152)	1.0754 (0.0984)
σ_u^2	6.6277 (2.2353)	6.5204 (2.2935)	6.6278 (2.3455)	6.5200 (2.2821)
τ_1	–	−0.0438 (0.0534)	–	−0.0435 (0.0535)
τ_2	–	0.2034 (0.0229)	–	0.2036 (0.0346)
ν	–	–	15.2199 (6.3777)	15.2370 (1.9470)

Table 4.3: Estimation results of the RealizedGARCH for both Gaussian and Student-t distributional assumptions on the IBM stock, the values in parenthesis are the standard errors and the * means the RealizedGARCH without leverage function.

It can be noted that the value of γ is greater than the value of α in the GARCH models, this shows how the RV is a way better realized measure compared to the squared log-return, thus capturing more weight in the estimation. Moreover the value of the parameter ϕ_1 is very close to one, suggesting a good proportional relationship between the realized variance and the conditional variance.

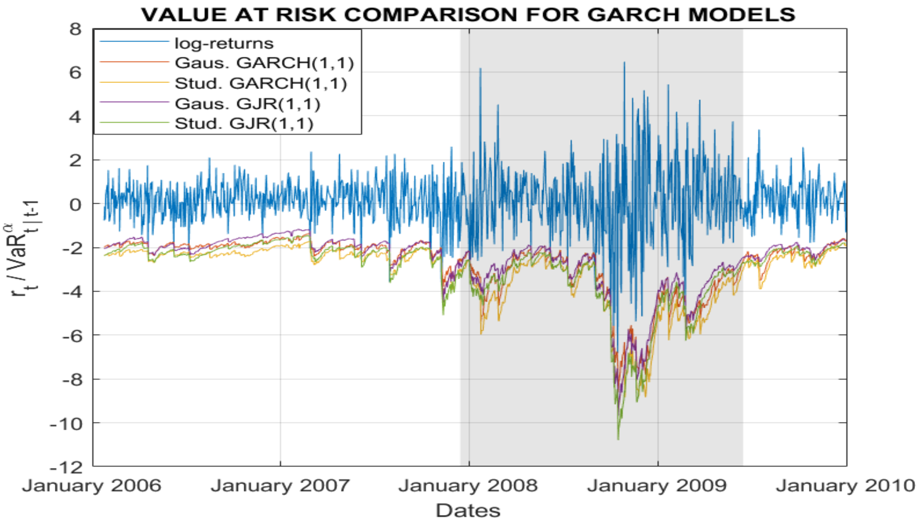
Regarding the Value-at-Risk computation a confidence level of 99% is used, because it bears the greatest practical interest due to the Basel requirements. The evaluation of the models forecasting power is a two step procedure, first the models are assessed via statistical accuracy tests and in particular the Kupiec, Christoffersen and Engle and Manganelli test, the rejection rule is at 5% in order to reduce the probability of a type I error (in this context rejecting a true null is more "dangerous" than the opposite).

Only the models that go through the first stage are then evaluated in terms of loss functions, taking into consideration the loss functions of Lopez and Sarma.

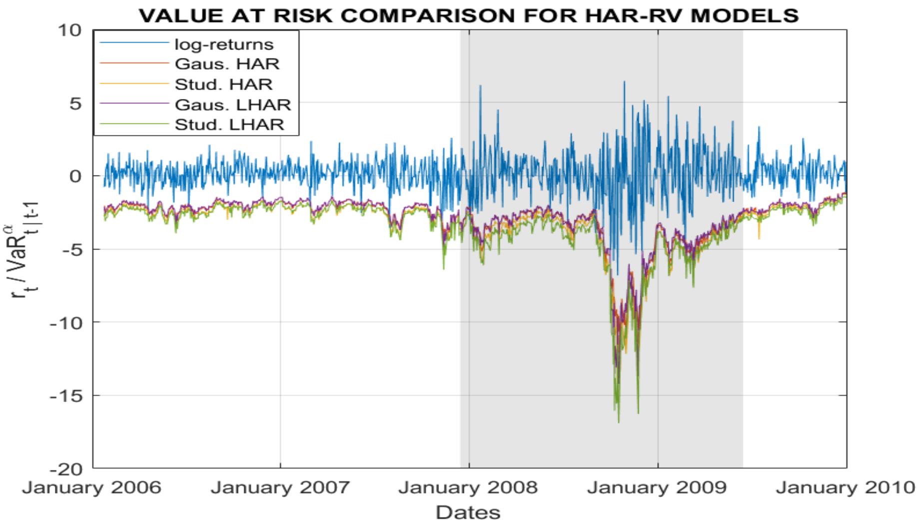
Figure 5.1 shows the Value-at-Risk one-step ahead forecasts for the three class of models, it has to be noted that the yellow and the green lines tend to be below the others. Thus, the Student-t density allows to generate in absolute value greater VaR forecasts, and thus fits better periods of turbulent markets.

Moreover is interesting to note how the VaR is able to follow the market behaviour, indeed during the 2008 and 2009, a period of a well known financial crisis, the forecasts sharply decreases. The financial explanation is because during turbulent market periods the volatility increases due to a greater uncertainty and this feature can be also noted by looking at the VIX (also known as "fear index") plot in figure 4.6. The VIX clearly skyrocket during turbulent periods. The mathematical explanation is instead that the volatility enters into the VaR forecasts with a negative magnitude, therefore the greater the volatility in absolute value the greater the VaR forecast in absolute value as well.

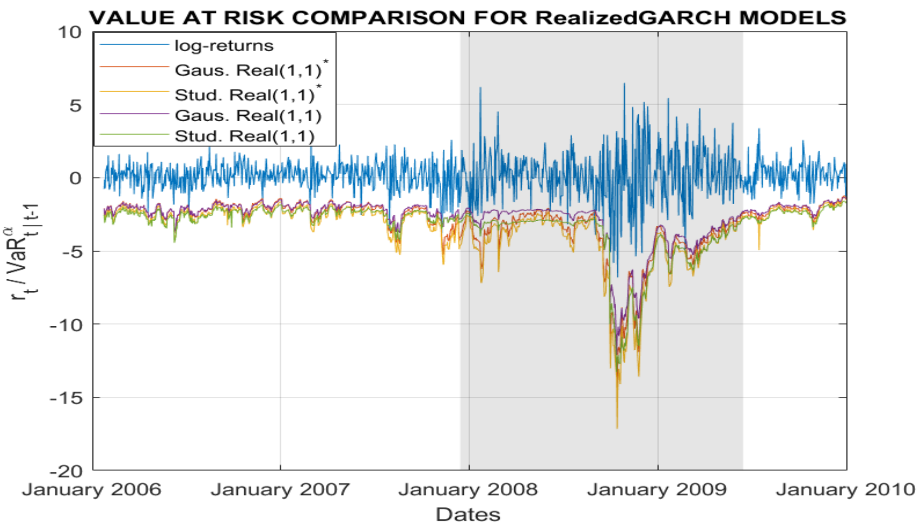
After discussing the graphical intuitions of the VaR forecasts, is interesting to look at the results of the statistical accuracy tests. Indeed, the first step of the evaluation process is based



(a) *GARCH models VaR forecasts*



(b) *HAR models VaR forecasts*



(c) *RealizedGARCH models VaR forecasts*

Figure 4.5: Value-at-Risk forecasts comparison, the grey shaded area represents the period of the financial crisis.

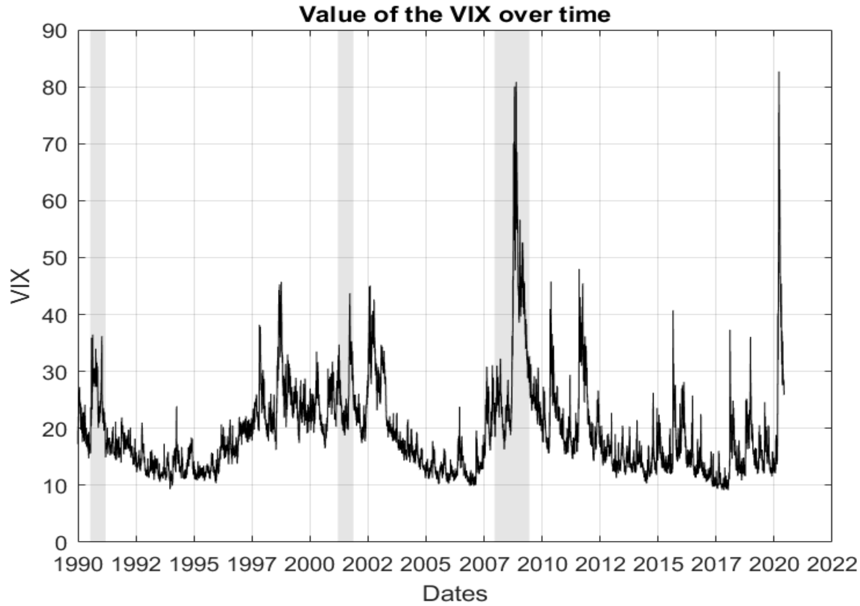


Figure 4.6: VIX value over time.

on the Christoffersen and the Engle and Manganelli test, the Kupiec test is shown for the sake of presentation but is not considered due to its incompleteness. Therefore, only the models able to pass both the aforementioned tests are then evaluated in terms of the Lopez and Sarma loss functions.
Starting with the GARCH models, table 5.4 shows the p-values of the Kupiec, Christoffersen and Engle and Manganelli test.

Test	Ga. GARCH(1,1)	St. GARCH(1,1)	Ga. GJR(1,1)	St. GJR(1,1)
Kupiec	0.0206	0.9797	0.0748	0.3482
Christoffersen	0.0425	0.2277	0.1063	0.2375
Engle and Mang.	0.0323	0.0705	0.0832	0.1377

Table 4.4: P-values of different accuracy tests for the GARCH models.

Therefore, for the GARCH models only the standard GARCH(1,1) with Gaussian density specification does not pass the first step check. Indeed for the Gaussian GARCH(1,1) all the three tests are rejected, thus implying a strong lack of conditional efficiency.

Proceeding with the HAR-RV models, table 5.5 shows the p-values of the tests used.

Test	Gaus. HAR	Stud. HAR	Gaus. LHAR	Stud. LHAR
Kupiec	0.7348	0.3253	0.9797	0.0826
Christoffersen	0.2354	0.0587	0.2056	0.2172
Engle and Mang.	0.0708	0.0019	0.0440	0.6479

Table 4.5: P-values of different accuracy tests for the HAR models.

Thus, for the HAR-RV models both the Student HAR and the Gaussian LHAR fail to pass the tests. In particular the Student LHAR fails both the Christoffersen and the Engle and

Manganelli test, whilst the Gaussian LHAR fails the DQ test only.

Regarding the RealizedGARCH models, table 5.6 shows the results of the accuracy tests.

Test	Ga. Real(1,1)*	Ga. Real(1,1)	St. Real(1,1)*	St. Real(1,1)
Kupiec	0.7655	0.5205	0.0096	0.3253
Christoffersen	0.8816	0.2602	0.0647	0.0690
Engle and Mang.	0.9890	0.0132	0.3023	0.2160

Table 4.6: P-values of different accuracy tests for the RealizedGARCH models, the * means that the model does not account for leverage effect.

Hence, for the RealizedGARCH models only the RealizedGARCH(1,1) with Gaussian density and leverage function is rejected, all the others pass both the tests.

A first important result emerging form the empirical analysis performed so far is that the statistical accuracy is considerably improved when a log-return density more close to the stylized facts is adopted. Indeed, only one model out of six with Student-t specification has been rejected (i.e. 83% of models have been accepted), whereas for the Gaussian specification three models out of six have been rejected (i.e. 50% of models have been accepted).
This result has a nice financial implication, indeed a regulator which is more interested in preserving the stability of the system or a conservative risk manager should focus more on the log-return density than on the volatility model selection. It is widely common in practice to use a Gaussian assumption, which is far from adequately describe log-returns density, thus this is not a proper setting. Indeed a widespread model such as the Gaussian GARCH(1,1) has been totally rejected, showing a lack of conditional efficiency (the results would not be different if another popular model as the RiskMetrics had been employed).

Regarding the second step of the evaluation process, the QLF and the FLF loss functions are taken into consideration and a 10% cost of capital is assumed for the Sarma loss function (it can be whatever because is only a change of scale and not of order).
In table 5.7 the values of the loss functions are shown for the selected models, of course given that the measure is a loss function the purpose is to get a value as small as possible.

Model	Lopez (QLF)	Sarma (FLF)
Stud. GARCH(1,1)	0.0144	0.3330
Gaus. GJR(1,1)	0.0267	0.2851
Stud. GJR(1,1)	0.0177	0.3136
Gaus. HAR	0.0166	0.3037
Stud. LHAR	0.0064	0.3553
Gaus. Real(1,1)*	0.0129	0.3186
Stud. Real(1,1)*	0.0051	0.3544
Stud. Real(1,1)	0.0091	0.3358

Table 4.7: Lopez and Sarma loss functions results for the models passing the first step.

From the results in table 5.7 it can be noted that the decrease of the QLF when going from models using daily data only to models using also intra-daily data is pretty evident, showing a greater efficiency of these newer models (for instance the Student GARCH(1,1) is completely dominated by the Gaussian RealizedGARCH(1,1)). Moreover, given the same model the Student-t density helps in lowering the QLF as it can be seen for instance for the GJR-GARCH(1,1) . However, the models using intra-daily data experience a slightly higher FLF, but the increase in the FLF is marginal when compared to the decrease in the QLF which is two or even three times lower, this trade-off effect is probably due to the more conservative behaviour of the HAR-RV and RealizedGARCH models. Indeed the Sarma loss function severely penalizes conservative models.

Furthermore, is interesting to see how the most commonly used method performs in the testing framework built. Indeed, the Historical Simulation is still the benchmark in the industry [3]. Figure 5.2 shows the behaviour of the VaR computed via HS and WHS. After many

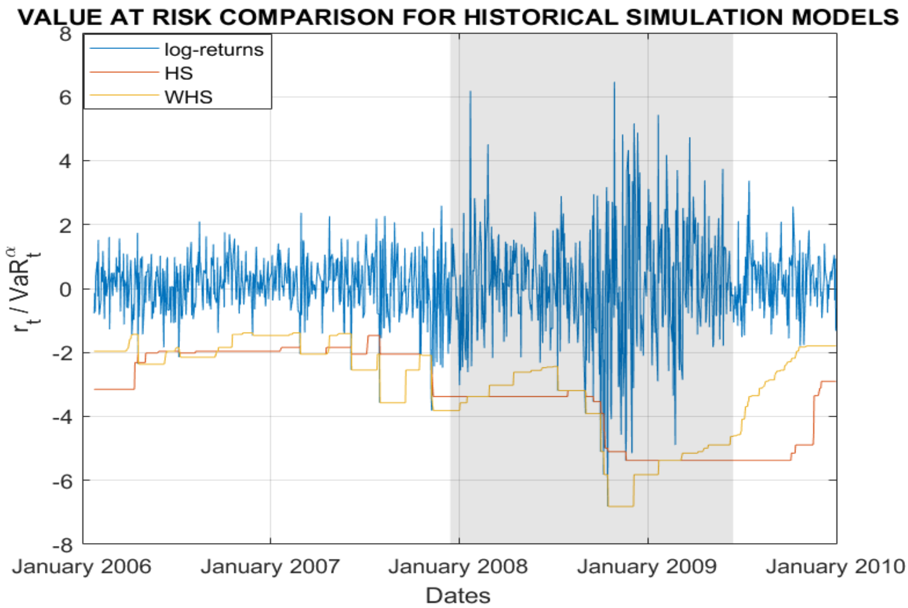


Figure 4.7: Historical Simulation and Weighted Historical Simulation VaR estimates

trials the best setting for the HS turned out to be a window length of 250 observations (i.e. one trading year). While for the WHS the best setting turned out to be a window length of 750 observations and a decay factor of 0.98. It is easy to note how the WHS reacts much faster to changes in volatility compared to the plain Historical Simulation. Regarding the statistical accuracy the HS is not able to pass any of the tests at 5% significance level, whilst the WHS is able to go through the Kupiec and Christoffersen tests but not thorough the Engle and Manganelli test. Thus the WHS is way more accurate than the simpler HS. Regarding

Model	Lopez (QLF)	Sarma (FLF)
Historical Simulation	0.0364	0.3317
Weighted Historical Simulation	0.0261	0.3014

Table 4.8: Lopez and Sarma loss functions results for the HS and WHS

[3]Perignon and Smith (2010) reported that from the 64.5% of the banks that revealed their VaR methodology in their survey, 73% uses the Historical Simulation

the efficiency table 5.8 shows both the Lopez and Sarma loss functions. The HS is clearly dominated by the WHS and considering the trade-off between the QLF and the FLF the HS is dominated also by the models in table 5.7, whereas the WHS puts a decent competition, in terms of loss function efficiency, against the parametric models. However, even though the WHS has a decent performance in terms of loss functions it fails to achieve the conditional efficiency feature. Therefore, is important to stress that when choosing non-parametric models as the HS and the WHS there is for sure a gain in terms of ease of implementation and room of applicability, but the costs are a loss of accuracy and a not so bright efficiency compared to parametric models especially for the HS.

Now is possible to reply to the question posed at the end of chapter 3, in fact models using intra-daily data are more efficient than models using daily data only, thus using more advanced models in computing the VaR allows institutions to be more careful in monitoring their market risk and at the same time increase, or at least not reduce, the efficiency (for other works on the same topic see Giot and Laurent 2004, or Louzis et al. 2014).
A possible explanation for the finding that models using intra-daily data improve the overall performance of VaR forecasts is that realized variance is a way more efficient volatility estimator compared to the squared log-return utilized by the GARCH models. Therefore, models based on the realized variance are able to give back less noisy VaR forecasts.

Furthermore, for the sake of completeness is interesting to look at the forecasting power of the models considered so far not only in terms of VaR forecast, but also in terms of volatility forecast per se. Table 5.9 shows the Mean Squared Error (MSE) for all the models developed in this work, where the true variance (it is a second moment so unobservable) is approximated by the RV, hence computing the MSE in this way has been criticized because is somewhat biased in favour of models using the RV itself.

Model	*MSE*
Gaussian GARCH(1,1)	17.2349
Student GARCH(1,1)	17.5916
Gaussian GJR(1,1)	15.5375
Student GJR(1,1)	15.7767
Gaussian HAR	13.1513
Gaussian LHAR	11.9539
Student HAR	13.1515
Student LHAR	11.9522
Gaussian RealGARCH(1,1)*	12.5470
Gaussian RealGARCH(1,1)	14.2274
Student RealGARCH(1,1)*	12.4337
Student RealGARCH(1,1)	14.3623

Table 4.9: MSE values for all the volatility models.

As shown in the above table, the MSE of models using intra-daily data is always lower than the one of models using daily data only. Hence, another possible explanation of the better performance of HAR-RV and RealizedGARCH models is that they return back a better variance forecast, which directly enters into the VaR formula and in turn improve the forecast of the risk measure itself.

However, it is important to point out that the better performance of models using intra-daily data comes at some costs, because they require the availability of high frequency data for a broad range of asset classes, storage and real-time processing of the data which are not trivial and have to be accounted before deploying these VaR models to cover all the activities of a financial institution.

Chapter 5

Conclusions

The main purpose of this thesis has been to assess the forecasting ability of two classes of volatility models, a class of traditional models using daily data only and a class of new models using also high-frequency data.
However, the forecasting power of the models has not been assessed in a traditional fashion by using widespread measures as the MSE or the MAE or the R^2. Indeed, the forecasts have been then utilized to compute the Value-at-Risk projections and then these projections have been evaluated in order to assess the forecasting power of the models.

Using the VaR as a different way to evaluate volatility models is not only an elegant alternative to traditional methods, but also an approach that bears a lot of practical interest due to the current Basel regulation.
The question posed in the introduction was whether models using also high-frequency data improve the VaR forecasts, and as shown in chapter 4 the new models improve the overall quality of the VaR estimates compared to the traditional models. Moreover, a kind of serendipity has occurred as often in science (let's think for instance about Alexander Fleming or Cristoforo Colombo), because the results showed that in computing VaR forecasts is of course relevant the chosen volatility model, but is maybe even more important the log-returns density used.
Indeed, in chapter 4 is clear how using a density which better fits returns stylized facts drastically improves the statistical accuracy and the efficiency of the VaR estimates.

These results are an important bottom line in financial literature, proving that research lead to useful improvements. In fact these findings are of paramount relevance for both authorities and practitioners. From a regulator's point of view a greater accuracy of risk management models could help mitigate systemic risk in periods of high market volatility, thus contributing to the overall stability of the financial system.
Furthermore, a risk manager willing to adopt a conservative approach should go for intra-daily models. However, it has to be considered that using intra-daily models requires availability of high-frequency data, storage and processing efforts that are non trivial elements to be accounted. Anyway, whenever high-frequency data are available using intra-daily models is strongly suggested, the HAR-RV and the RealizedGARCH have similar performance but the time needed for the computation is lower for the former model.
Moreover, it has also been shown how the plain HS method lacks of statistical accuracy, whilst the WHS performs better both in terms of accuracy, failing the DQ test only, and in

terms of efficiency. Anyway the historical simulation methods have the great advantage to be model-free and are thus easy to implement. Hence, as usual, there is a trade-off to be accounted for when choosing the VaR computation approach.

Finally, new paths for further investigation are open since other conditional densities such as the GED or the skewed-Student can be used. Regarding high-frequency models, also more robust realized measures of volatility as the Bipower Variation or the Realized Kernel can be used. Furthermore, other interesting models can be taken into consideration, for instance the MEM, the APARCH or the ARFIMAX-RV. Lastly, a comparison amongst different methods for computing the VaR can be performed. For instance would be of practical relevance to compare the fully parametric variance-covariance approach with approaches as the Filtered Historical Simulation proposed by Hull and White (1998), Barone-Adesi et al. (1998) or the Peaks Over Threshold EVT.

Bibliography

[1] Acerbi, O., Tasche, D., 2002. *On the coherence of expected shortfall.* Journal of Banking and Finance 26, 1505-1518.

[2] Andersen, T.G., Bollerslev, T., 1998. *Answering the skeptics: yes, standard volatility models do provide accurate forecasts.* International Economic Review 39, 885-905.

[3] Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001. *The distribution of realized stock returns volatility.* Journal of Financial Economics 61, 43-76.

[4] Andersen, T.G., Bollerslev. T., Diebold, F.X., Labys, P., 2001. *The distribution of exchange rate volatility.* Journal of American Statistical Association 96, 42-55.

[5] Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. *Roughing it up: including jump components in the measurement, modelling and forecasting of return volatility.* Review of Economics and Statistics 89, 701-720.

[6] Barndorff-Nielsen, O.E., Shephard, N., 2002. *Econometric analysis of realized volatility and its use in estimating stochastic volatility models.* Journal of Royal Statistical Society B 64, 253-280.

[7] Barndorff-Nielsen, O.E., Shephard, N., 2004. *Power and Bipower variation with stochastic volatility and jumps.* Journal of Financial Econometrics 2, 1-37.

[8] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. *Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise.* Econometrica 76, 1481-1536.

[9] Barone-Adesi, G., Bourgoin, F., Giannopoulos, K., 1998. *Don't look back.* Risk 11, 100-103.

[10] Black, F., Scholes, M., 1973. *The pricing of options and corporate liabilities.* Journal of Political Economy 81, 637-654.

[11] Black, F., 1976. *Studies of stock market volatility changes.* Proceeding of the American Statistical Association, Business and Economic Statistic Section, 177-181.

[12] Bollerslev, T., 1986. *Generalized autoregressive conditional heteroskedasticity.* Journal of Econometrics 31, 307-327.

[13] Caporin, M., Rossi, E., Santucci de Magistris, P., 2016. *Volatility Jumps and their economic determinants.* Journal of Financial Econometrics, 29-80.

[14] Christoffersen, P., 1998. *Evaluating interval forecasts.* International Economic Review 39, 841-862.

[15] Christoffersen, P., 2003. *Elements of Financial Risk Management.* Academic Press, Boston.

[16] Cox, J.C., Ross, S. A., 1976. *The valuation of options for alternative stochastic processes.* Journal of Financial Economics 3, 145-166.

[17] Corsi, F., Pirino, D., Renò, R., 2008. *Volatility forecasting: the jumps do matter.* Series Department of Economics University of Siena 534.

[18] Corsi, F., 2009. *A simple approximate long memory model of realized volatility.* Journal of Financial Econometrics 7, 174-196.

[19] Coris, F., Pirino, D., Renò, R., 2010. *Threshold bipower variation and the impact of jumps on volatility forecasting.* Journal of Econometrics 159, 276-288.

[20] Corsi, F., Renò, R., 2012. *Discrete-time volatility forecasting with persistent leverage effect and link with continuous-time volatility modelling.* Journal of Business and Economic Statistics 30, 368-380.

[21] de la Vega, J., 1688. *Confusion de Confusiones.*

[22] Dunbar, N., 2000. *Investing Money.* Wiley

[23] Engle, R.F., 1982. *Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation.* Econometrica 50, 987-1008.

[24] Engle, R.F., Bollerslev, T., 1986. *Modeling the persistence of conditional variances.* Econometric Review 5, 1-50.

[25] Engle, R.F., Ng, V., 1993. *Measuring and testing the impact of news on volatility.* Journal of Finance 48, 1747-1778.

[26] Engle, R.F., 2002. *New frontiers of ARCH models.* Journal of Applied Econometrics 17, 425-446.

[27] Engle, R.F., Manganelli, S., 2004. *CaViaR: conditional autoregressive value at risk by regression quantiles.* Journal of Business and Economic Statistic 22, 367-381.

[28] Fama, E., 1970. *Efficient Capital Markets: A review of theory and empirical work.* Journal of Finance 25, 383-417.

[29] Giot, P., Laurent, S., 2004. *Modeling daily value-at-risk using realized volatility and ARCH type models.* Journal of Empirical Finance 11, 379-398.

[30] Glosten, L., Jagannathan, R., Runkle, D., 1993. *Relationship between the expected value and the volatility of the nominal excess return on stocks.* Journal of Finance 48, 1779-1801.

[31] Hamilton, J., 1994. *Time Series Analysis.* Princeton University Press.

[32] Hansen, P.R., Lunde, A., 2006, *Realized volatility and market microstructure noise.* Journal of Business and Economic Statistics 24, 127-161.

[33] Hansen, P.R., Huang, Z., Shek, H.H., 2012. *Realized GARCH: a joint model for returns and realized measures of volatility.* Journal of Applied Econometrics 27, 877-906.

[34] Hull, J., White, A., 1998. *Incorporating volatility updating into the historical simulation method for VaR.* Journal of Risk 1, 5-19.

[35] Jorion, P., 2001. *Value-at-Risk: The New Benchmark for Managing Financial Risk.* McGraw-Hill.

[36] Koopman, S.J., Jungbacher, B., Hal, E., 2005. *Forecasting daily variability of the S&P100 stock index using historical, realized and implied volatility measurements.* Journal of Empirical Finance 12, 445-475.

[37] Kolmogorov, A.N., 1933. *Foundations of the Theory of Probability.* Springer.

[38] Kupiec, P., 1995. *Techniques for verifying the accuracy of risk measurement models.* Journal of Derivatieves 2, 173-184

[39] Lopez, J.A., 1999. *Methods for evaluating Value-at-Risk estimates.* Fed. Reserv. Bank San Franc. Economic Review 2, 3-17.

[40] Markowitz, H.M., 1952. *Portfolio Selection.* Journal of Finance 7, 77-91.

[41] Muller, U., Dacorogna, M., Dav, R., Olsen, R., Pictet, O., Ward, J., 1997. *Fractals and intrinsic time- A challenge to econometricians.* 39th International AEA Conference on Real Time Econometrics.

[42] Pagan, A., Schwert, G., 1990. *Alternative models for conditional stock volatility.* Journal of Econometrics 45, 267-290.

[43] Perignon, C., Smith, D., 2010. *The level and quality of Value-at-Risk disclosure by commercial banks.* Journal of Banking and Finance 34, 362-377.

[44] Protter, P., 1992. *Stochastic Integration and Differential Equations.* A new approach. Springer.

[45] Sarma, M., Thomas, S., Shah, A., 2003. *Selection of Value-at-Risk models.* Journal of Forecasting 22, 337-358.

[46] Zhang, L., Mykland, P.A., Ait-Sahalia, Y., 2005. *A tale of two time scales: determining integrated volatility with noisy high frequency data.* Journal of American Statistical Association 100, 1394-1411.

SUMMARY

The recent financial crisis in 2007-2009 demonstrated that the financial institutions risk management systems were not as adept as thought in tracking and anticipating extreme price movements during that highly volatile period.
Almost all financial institutions experienced multiple consecutive violations. Therefore, several doubts was cast and many questions were raised regarding the accuracy and the reliability of the implemented VaR models, procedures and systems.

The foundation of financial econometrics and modern risk management were laid with the seminal works of Engle in 1982 and Bollerslev in 1986, introducing ARCH-type models. However, further research achievements allowed the formal development of theory of quadratic variation as backbone of realized measures of volatility, and thus a lot of things has changed since the 80s.
Moreover, with technological developments such as cheap computational power and high-frequency databases several new volatility models have been proposed and their implementation became easier.

Therefore, the main purpose of the thesis is to assess whether the newer models using also intra-daily data are able to give back better volatility forecasts compared to the commonly used traditional models based on daily data only. However, several authors already demonstrated the superiority of realized volatility models over ARHC-type models (e.g. Koopman et al., 2005). This is the reason why in this work the assessment of the models is performed under the framework of the Value-at-Risk.

The assessment of volatility models under the VaR is a new approach initiated with the work of Giot and Laurent (2004) and is not only a valid alternative to traditional approaches (e.g. those based on measures as the MAE, MSE, etc.), but is also of practical importance due to the widespread application of the VaR as a risk measure.

The empirical results focus on comparing the traditional models GARCH and GJR-GARCH with high-frequency models such as the HAR-RV and the RealizedGARCH. In particular, the comparison is based on a two step procedure. The first step is based on the evaluation of the statistical accuracy of the models by using two tests for conditional efficiency, the Christoffersen test and the Engle and Manganelli test.
The second step is instead based on the assessment of the models in terms of a loss function, thus meaningful measures of efficiency are used and in particular the loss functions of Sarma and Lopez.

The results are shown with respect to the IBM stock for a period going from 2002 to 2009, thus considering also the turbulent period of the financial crisis.
The choice of the models has a precise rationale, the GARCH is chosen because is the most

used discrete time volatility model. The GJR-GARCH is used in order to allow to capture the leverage effect, thus it is an extension of the standard GARCH. The RealizedGARCH shares the same structure of the ARCH-type models, but instead of using squared returns it utilizes realized volatility hence it exploits informations in high-frequency data. Moreover, the RealizedGARCH is able to handle the leverage effects but in a different way compared to the GJR and in general to ARCH-type models. Finally the HAR-RV model is chosen because it takes into account some important stylized facts of realized volatility such as the long memory property.

The dataset used for the empirical analysis goes from 1 February 2002 to 31 December 2009. The Realized Variance is computed by using 5-minutes prices, the choice of the frequency is an heuristic choice in line with a cut-off able to reduce the microstructure noises.

Regarding the empirical analysis two important statistical tests have been carried out on the log-returns. First the Ljung-Box test in order to asses whether log-returns are correlated over time and they turn out not to be correlated, while their squared values are. Then the ARCH test has been performed in order to asses whether there is conditional heteroskedasticity in log-returns, and they turn out to be heteroskedastic. Thus, the statistical tests endorsed exactly some of the stylized facts of log-returns.
The dataset is composed of 2014 observations and the estimation is performed in a rolling window basis with a window length of 1000 observations, thus approximately four trading years.
The HAR-RV models need 22 observations (the number of lags) in order to be initialized, therefore for the other models 22 observations are dropped from the in-sample dataset and at the end the one-step ahead forecasts composing the out-of sample dataset are 992.

Four models are analyzed: GARCH, GJR-GARCH, HAR-RV, Realized-GARCH, for each model two distributional assumptions are made, first the Gaussian distribution and then in order to capture the fat tail phenomenon the Student-t distribution.
For the GARCH family models a model selection procedure based on the BIC information criterion has been carried out, where $BIC = -2l(\widehat{\psi}) + klogT$ and $l(\widehat{\psi})$ is the log-likelihood function of the model in the estimated parameters, k represents its number of parameters and T is the sample size.
The model selection procedure for the return equation showed an optimal lag length of zero for both the autoregressive and the moving average parts, therefore the return equation has no conditional mean and is thus represented by a weak White Noise only (as stated below). While the model selection procedure for the variance equation showed an optimal lag length of one for the ARCH and GARCH parts for all the models GARCH, GJR-GARCH and RealizedGARCH.

The models have first been estimated [1] entirely on the eight years dataset in order to look at the estimates at hand. Then the VaR forecasts are computed in a rolling window basis with a length of 1000 observations (in-sample) covering the period from 2002 to 2006, hence almost 1000 forecasts (out-of-sample) are computed and in particular they cover the period 2006-2009. Hence, the out-of-sample period includes both a turbulent period, namely the financial crisis and a calm period from 2006 to the end of 2007. Therefore, the VaR models

[1] All the estimates, the plots and the statistical tests in this chapter are performed by using the software MatLab. Moreover all the estimation procedures and the vast majority of tests have been implemented from scratch, without using any built-in function. The codes are available upon request.

are tested with respect to a calm and a turbulent period.
Table 5.1 shows the estimation results for the GARCH model and the GJR-GARCH with Gaussian and Student-t distribution assumption.

Par.	Gaus. GARCH(1,1)	Stud. GARCH(1,1)	Gaus. GJR(1,1)	Stud. GJR(1,1)
ω	0.0098 (0.0083)	0.0088 (0.0062)	0.0121 (0.0049)	0.0117 (0.0049)
α	0.0663 (0.0182)	0.0619 (0.0145)	0.0088 (0.0105)	0.0106 (0.0097)
β	0.9298 (0.0221)	0.9344 (0.0170)	0.9486 (0.0171)	0.9480 (0.0166)
ν	–	10.3997 (1.0719)	–	13.6048 (0.0173)
γ	–	–	0.0744 (0.173)	0.0722 (0.0173)

Table 5.1: Estimation results of the GARCH and GJR for both Gaussian and Student-t distributional assumptions on the IBM stock, the values in parenthesis are the standard errors.

From the estimation of the GJR it can be seen that the value of γ is always positive, thus allowing to properly capture the leverage effect. Moreover the p-values of γ make it significant, thus endorsing the relevance of the leverage effects.

In table 5.2 instead are shown the parameters estimate for the various HAR-RV models, it can be seen that all the parameters $\gamma^{(\cdot)}$ are negative in order to properly account for the leverage effect. Furthermore the parameters $\gamma^{(\cdot)}$ are, besides the monthly one, significant and therefore proving the relevance of the leverage effect.

Par.	HAR	LHAR
c	0.0051 (0.0105)	−0.0540 (0.0120)
$\beta^{(d)}$	0.2984 (0.0291)	0.2500 (0.0286)
$\beta^{(w)}$	0.4847 (0.0473)	0.4406 (0.0454)
$\beta^{(m)}$	0.1766 (0.0348)	0.2161 (0.0352)
$\gamma^{(d)}$	–	−0.0813 (0.0153)
$\gamma^{(w)}$	–	−0.1581 (0.0386)
$\gamma^{(m)}$	–	−0.0318 (0.0637)

Table 5.2: Estimation results of the HAR and LHAR on the IBM stock, the values in parenthesis are the standard errors.

Moreover, when performing the second step estimation the parameter ϕ is significant and close to 1, it is actually $\widehat{\phi} = 1.0068$ for the HAR and $\widehat{\phi} = 1.0036$ for the LHAR.

Regarding the RealizedGARCH(1,1) models table 5.3 shows the estimation results.
It can be noted that the value of γ is greater than the value of α in the GARCH models, this shows how the RV is a way better realized measure compared to the squared log-return, thus capturing more weight in the estimation. Moreover the value of the parameter ϕ_1 is very close to one, suggesting a good proportional relationship between the realized variance and the conditional variance.

After estimating all the necessary parameters the Value-at-Risk forecast for a specific confidence level is obtained, depending on the distributional assumption, respectively via the

Par.	Gaus. Real(1,1)*	Gaus. Real(1,1)	Stud. Real(1,1)*	Stud. Real(1,1)
ω	0.0127 (0.0146)	0.0210 (0.0159)	0.0122 (0.0140)	0.0220 (0.0162)
β	0.7299 (0.0551)	0.7330 (0.0551)	0.7307 (0.0554)	0.7341 (0.0554)
γ	0.2395 (0.0489)	0.2289 (0.0483)	0.2391 (0.0499)	0.2266 (0.0490)
ϕ_0	0.1324 (0.1438)	0.0949 (0.1350)	0.1291 (0.1515)	0.0851 (0.1364)
ϕ_1	1.0306 (0.1106)	1.0687 (0.1030)	1.0299 (0.1152)	1.0754 (0.0984)
σ_u^2	6.6277 (2.2353)	6.5204 (2.2935)	6.6278 (2.3455)	6.5200 (2.2821)
τ_1	–	−0.0438 (0.0534)	–	−0.0435 (0.0535)
τ_2	–	0.2034 (0.0229)	–	0.2036 (0.0346)
ν	–	–	15.2199 (6.3777)	15.2370 (1.9470)

Table 5.3: Estimation results of the RealizedGARCH for both Gaussian and Student-t distributional assumptions on the IBM stock, the values in parenthesis are the standard errors and the * means the RealizedGARCH without leverage function.

Delta-Normal and Delta-Student formulas

$$VaR^{\alpha}_{t+1|t} = -\sigma_{t+1|t}\Phi^{-1}(\alpha)$$

$$VaR^{\alpha}_{t+1|t} = -\sigma_{t+1|t}t_{\nu}^{-1}(\alpha) \tag{5.1}$$

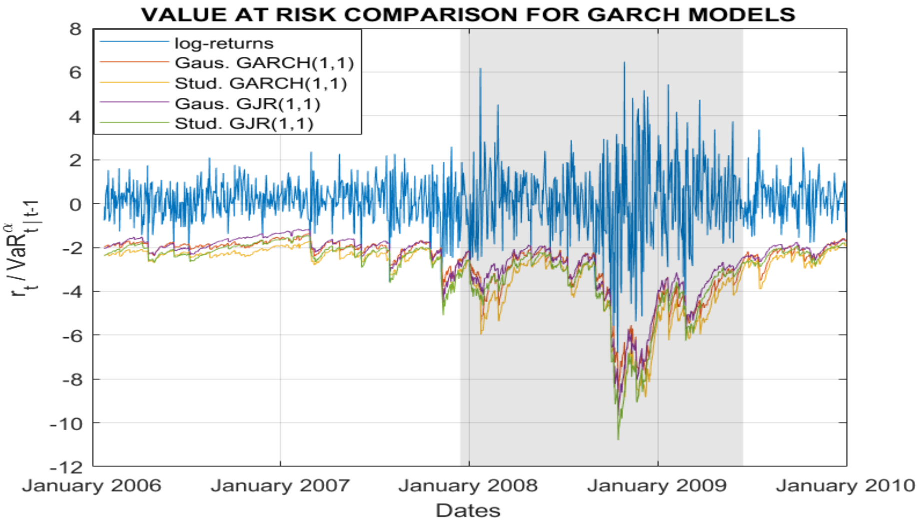
where the functions $\Phi^{-1}(\cdot)$ and $t_{\nu}^{-1}(\cdot)$ are respectively the quantile functions (in this case the inverse of the CDF) of the Standard Normal and Standard Student-t distributions. Regarding the Value-at-Risk computation a confidence level of 99% is used, because it bears the greatest practical interest due to the Basel requirements. The evaluation of the models forecasting power is a two step procedure, first the models are assessed via statistical accuracy tests and in particular the Kupiec, Christoffersen and Engle and Manganelli test, the rejection rule is at 5% in order to reduce the probability of a type I error (in this context rejecting a true null is more ”dangerous” than the opposite).
Only the models that go through the first stage are then evaluated in terms of loss functions, taking into consideration the loss functions of Lopez and Sarma.

Figure 5.1 shows the Value-at-Risk one-step ahead forecasts for the three class of models, it has to be noted that the yellow and the green lines tend to be below the others. Thus, the Student-t density allows to generate in absolute value greater VaR forecasts, and thus fits better periods of turbulent markets.

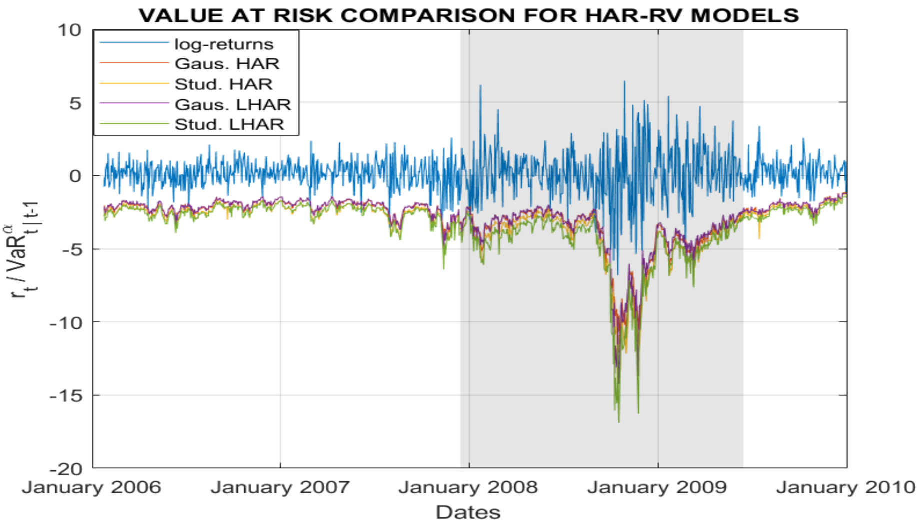
Moreover is interesting to note how the VaR is able to follow the market behaviour, indeed during the 2008 and 2009, a period of a well known financial crisis, the forecasts sharply decreases. The financial explanation is because during turbulent market periods the volatility increases due to a greater uncertainty.

After discussing the graphical intuitions of the VaR forecasts, is interesting to look at the results of the statistical accuracy tests on the VaR violation processes. Starting with the GARCH models, table 5.4 shows the p-values of the Kupiec, Christoffersen and Engle and Manganelli test.

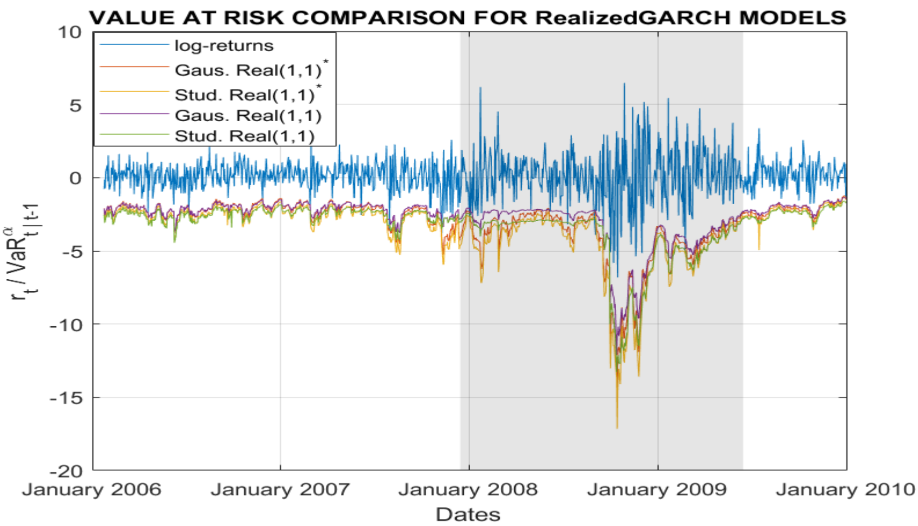
Therefore, for the GARCH models only the standard GARCH(1,1) with Gaussian density



(a) *GARCH models VaR forecasts*



(b) *HAR models VaR forecasts*



(c) *RealizedGARCH models VaR forecasts*

Figure 5.1: Value-at-Risk forecasts comparison, the grey shaded area represents the period of the financial crisis.

Test	Ga. GARCH(1,1)	St. GARCH(1,1)	Ga. GJR(1,1)	St. GJR(1,1)
Kupiec	0.0206	0.9797	0.0748	0.3482
Christoffersen	0.0425	0.2277	0.1063	0.2375
Engle and Mang.	0.0323	0.0705	0.0832	0.1377

Table 5.4: P-values of different accuracy tests for the GARCH models.

specification does not pass the first step check. Indeed for the Gaussian GARCH(1,1) all the three tests are rejected, thus implying a strong lack of conditional efficiency.

Proceeding with the HAR-RV models, table 5.5 shows the p-values of the tests used.

Test	Gaus. HAR	Stud. HAR	Gaus. LHAR	Stud. LHAR
Kupiec	0.7348	0.3253	0.9797	0.0826
Christoffersen	0.2354	0.0587	0.2056	0.2172
Engle and Mang.	0.0708	0.0019	0.0440	0.6479

Table 5.5: P-values of different accuracy tests for the HAR models.

Thus, for the HAR-RV models both the Student HAR and the Gaussian LHAR fail to pass the tests. In particular the Student LHAR fails both the Christoffersen and the Engle and Manganelli test, whilst the Gaussian LHAR fails the DQ test only.

Regarding the RealizedGARCH models, table 5.6 shows the results of the accuracy tests.

Test	Ga. Real(1,1)*	Ga. Real(1,1)	St. Real(1,1)*	St. Real(1,1)
Kupiec	0.7655	0.5205	0.0096	0.3253
Christoffersen	0.8816	0.2602	0.0647	0.0690
Engle and Mang.	0.9890	0.0132	0.3023	0.2160

Table 5.6: P-values of different accuracy tests for the RealizedGARCH models, the * means that the model does not account for leverage effect.

Hence, for the RealizedGARCH models only the RealizedGARCH(1,1) with Gaussian density and leverage function is rejected, all the others pass both the tests.

A first important result emerging form the empirical analysis performed so far is that the statistical accuracy is considerably improved when a log-return density more close to the stylized facts is adopted. Indeed, only one model out of six with Student-t specification has been rejected (i.e. 83% of models have been accepted), whereas for the Gaussian specification three models out of six have been rejected (i.e. 50% of models have been accepted).
This result has a nice financial implication, indeed a regulator which is more interested in preserving the stability of the system or a conservative risk manager should focus more on the log-return density than on the volatility model selection. It is widely common in practice to use a Gaussian assumption, which is far from adequately describe log-returns density, thus this is not a proper setting. Indeed a widespread model such as the Gaussian GARCH(1,1) has been totally rejected, showing a lack of conditional efficiency (the results would not be different if another popular model as the RiskMetrics had been employed).

Regarding the second step of the evaluation process, the QLF and the FLF loss functions are taken into consideration and a 10% cost of capital is assumed for the Sarma loss function

(it can be whatever because is only a change of scale and not of order).
In table 5.7 the values of the loss functions are shown for the selected models, of course given that the measure is a loss function the purpose is to get a value as small as possible.

Model	Lopez (QLF)	Sarma (FLF)
Stud. GARCH(1,1)	0.0144	0.3330
Gaus. GJR(1,1)	0.0267	0.2851
Stud. GJR(1,1)	0.0177	0.3136
Gaus. HAR	0.0166	0.3037
Stud. LHAR	0.0064	0.3553
Gaus. Real(1,1)*	0.0129	0.3186
Stud. Real(1,1)*	0.0051	0.3544
Stud. Real(1,1)	0.0091	0.3358

Table 5.7: Lopez and Sarma loss functions results for the models passing the first step.

From the results in table 5.7 it can be noted that the decrease of the QLF when going from models using daily data only to models using also intra-daily data is pretty evident, showing a greater efficiency of these newer models (for instance the Student GARCH(1,1) is completely dominated by the Gaussian RealizedGARCH(1,1)). Moreover, given the same model the Student-t density helps in lowering the QLF as it can be seen for instance for the GJR-GARCH(1,1) . However, the models using intra-daily data experience a slightly higher FLF, but the increase in the FLF is marginal when compared to the decrease in the QLF which is two or even three times lower, this trade-off effect is probably due to the more conservative behaviour of the HAR-RV and RealizedGARCH models. Indeed the Sarma loss function severely penalizes conservative models.

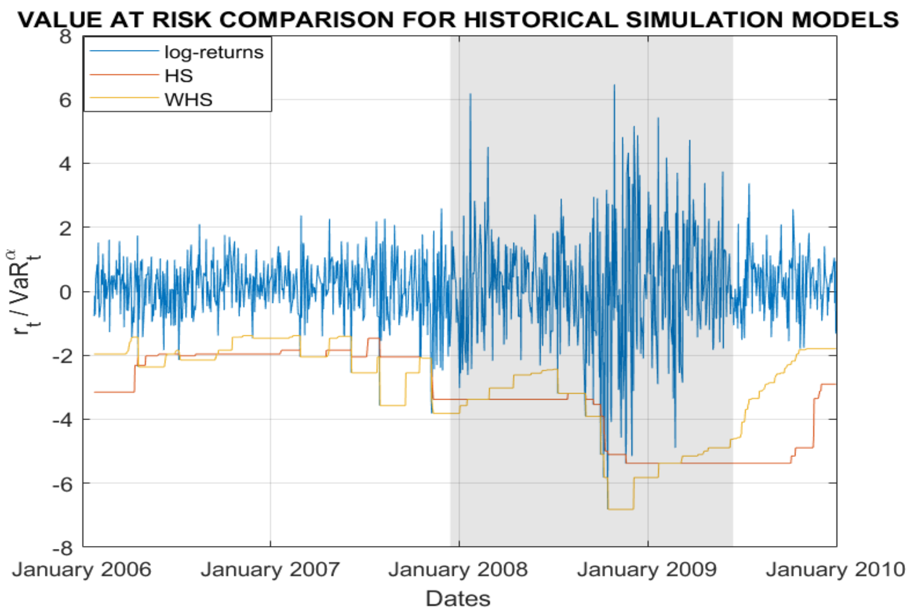


Figure 5.2: Historical Simulation and Weighted Historical Simulation VaR estimates

Furthermore, is interesting to see how the most commonly used method performs in the testing framework built. Indeed, the Historical Simulation is still the benchmark in the industry [2]. Figure 5.2 shows the behaviour of the VaR computed via HS and WHS. After many trials the best setting for the HS turned out to be a window length of 250 observations (i.e. one trading year). While for the WHS the best setting turned out to be a window length of 750 observations and a decay factor of 0.98. It is easy to note how the WHS reacts much faster to changes in volatility compared to the plain Historical Simulation. Regarding the statistical accuracy the HS is not able to pass any of the tests, whilst the WHS is able to go through the Kupiec and Christoffersen tests but not thorough the Engle and Manganelli test. Thus the WHS is way more accurate than the simpler HS. Regarding the efficiency table 5.8 shows both the Lopez and Sarma loss functions. The HS is clearly dominated by the WHS and considering the trade-off between the QLF and the FLF the HS is dominated also by the models in table 5.7, whereas the WHS puts a decent competition, in terms of loss function efficiency, against the parametric models. However, even though the WHS has a decent performance in terms of loss functions it fails to achieve the conditional efficiency feature. Therefore, is important to stress that when choosing non-parametric models as the HS and the WHS there is for sure a gain in terms of ease of implementation and room of applicability, but the costs are a loss of accuracy and a not so bright efficiency compared to parametric models especially for the HS.

Model	Lopez (QLF)	Sarma (FLF)
Historical Simulation	0.0761	0.3105
Weighted Historical Simulation	0.0261	0.3014

Table 5.8: Lopez and Sarma loss functions results for the HS and WHS

Now is possible to reply to the core question of the work, in fact models using intra-daily data are more efficient than models using daily data only, thus using more advanced models in computing the VaR allows institutions to be more careful in monitoring their market risk and at the same time increase, or at least not reduce, the efficiency (for other works on the same topic see Giot and Laurent 2004, or Louzis et al. 2014).
A possible explanation for the finding that models using intra-daily data improve the overall performance of VaR forecasts is that realized variance is a way more efficient volatility estimator compared to the squared log-return utilized by the GARCH models. Therefore, models based on the realized variance are able to give back less noisy VaR forecasts.

Furthermore, for the sake of completeness is interesting to look at the forecasting power of the models considered so far not only in terms of VaR forecast, but also in terms of volatility forecast per se. Table 5.9 shows the Mean Squared Error (MSE) for all the models developed in this work, where the true variance (it is a second moment so unobservable) is approximated by the RV, hence computing the MSE in this way has been criticized because is somewhat biased in favour of models using the RV itself.
As shown in the above table, the MSE of models using intra-daily data is always lower than the one of models using daily data only. Hence, another possible explanation of the better

[2]Perignon and Smith (2010) reported that from the 64.5% of the banks that revealed their VaR methodology in their survey, 73% uses the Historical Simulation

Model	*MSE*
Gaussian GARCH(1,1)	17.2349
Student GARCH(1,1)	17.5916
Gaussian GJR(1,1)	15.5375
Student GJR(1,1)	15.7767
Gaussian HAR	13.1513
Gaussian LHAR	11.9539
Student HAR	13.1515
Student LHAR	11.9522
Gaussian RealGARCH(1,1)*	12.5470
Gaussian RealGARCH(1,1)	14.2274
Student RealGARCH(1,1)*	12.4337
Student RealGARCH(1,1)	14.3623

Table 5.9: MSE values for all the volatility models.

performance of HAR-RV and RealizedGARCH models is that they return back a better variance forecast, which directly enters into the VaR formula and in turn improve the forecast of the risk measure itself.

These results are an important bottom line in financial literature, proving that research lead to useful improvements. In fact these findings are of paramount relevance for both authorities and practitioners. From a regulator's point of view a greater accuracy of risk management models could help mitigate systemic risk in periods of high market volatility, thus contributing to the overall stability of the financial system.
Furthermore, a risk manager willing to adopt a conservative approach should go for intra-daily models. However, it has to be considered that using intra-daily models requires availability of high-frequency data, storage and processing efforts that are non trivial elements to be accounted. Anyway, whenever high-frequency data are available using intra-daily models is strongly suggested, the HAR-RV and the RealizedGARCH have similar performance but the time needed for the computation is lower for the former model.

Moreover, it has also been shown how the plain HS method lacks of statistical accuracy, whilst the WHS performs better both in terms of accuracy, failing the DQ test only, and in terms of efficiency. Anyway the historical simulation methods have the great advantage to be model-free and are thus easy to implement. Hence, as usual, there is a trade-off to be accounted for when choosing the VaR computation approach.

Finally, new paths for further investigation are open since other conditional densities such as the GED or the skewed-Student can be used. Regarding high-frequency models, also more robust realized measures of volatility as the Bipower Variation or the Realized Kernel can be used. Furthermore, other interesting models can be taken into consideration, for instance the MEM, the APARCH or the ARFIMAX-RV. Lastly, a comparison amongst different methods for computing the VaR can be performed. For instance would be of practical relevance to compare the parametric variance-covariance approach with approaches as the Filtered Historical Simulation proposed by Barone-Adesi et al. (1998) or the Peaks Over Threshold EVT.