ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Università
della
Svizzera
italiana

Stochastic online and offline calibration of a jump-diffusion model for financial bubbles and crashes with non-local self-referencing mispricing and an application to risk measure estimation

Master Thesis

Mattias Franchetto

December 05, 2016

Supervisor: Prof. Patrick Gagliardini
Department of Economics, University of Lugano

Co-supervisors: Prof. Didier Sornette, Prof. Yannick Malevergne
Department of Management, Technology, and Economics, ETH Zürich

Abstract

Extremes and unpredictability are fascinating, but ominous words appearing in many different scientific fields, ranging from mathematics and engineering, to physics and finance. Focusing here on the topic of quantitative finance, this master thesis aims to address and demystify the concepts discussed immediately above in order to give a better understanding of risky financial investments, in the context of unstable and irrational market conditions.

In this research two different calibration approaches are adapted and deployed to the jump-diffusion model developed by Malevergne and Sornette [2014], in order to give statistically significant estimates to its unknown parameters.

To begin with, the model is analysed. Results are indicative of being able to capture empirically observed dynamics of financial time series. Simultaneously, it is revealed that there is inherently complexity in disentangling regular and anomalous activity. One of the most innovative aspects of the model at hand, lies within its non-local self-referencing crash-hazard rate estimate. The aim of this approach is basically to take into account not only instantaneous price deviations of the underlying asset from its fundamental value, but also previously observed unusual and unsustainable dynamics. This is achieved without the need for endogenous factors.

The first calibration approach extends the work that has been done in Berntsen [2015], embedding the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm developed by Spall [1987] within a SIR particle filter implementation. Due to complex dependencies between the hyper-parameters and the latent variables extracted by the filter, no convergence can be guaranteed. More stable estimates are then obtained through an offline profile likelihood maximization.

As a next step, different parameter combinations are analyzed by deploying a Monte Carlo calibration approach, leading to interesting observations about the model itself.

The paper concludes with a comparison to a simpler GARCH(1,1) model, within a practical risk measure estimation context. All code and data are available on Github.

Acknowledgements

First of all I would like to thank Prof. Didier Sornette for the priceless opportunity he gave me to spent those months in the group and for showing me the right direction to follow even when my ideas were just vague thoughts.

A sincere and warn thank you to Prof. Malevergne for the guidance and the daily challenges he helped me to face, but also for the patience he showed while discussing the project together with its extensions and improvements.

I would also like to express my gratitude to Prof. Patrick Gagliardini for the support he gave me from Lugano, despite the complications that arose along the road.

An overall thanks also to all my colleagues at the Chair of Entrepreneurial Risk, with some special remark that has to be made in order to thank Guilherme for showing me how to properly profile a likelihood, Michael for giving me the guidance on the mathematics behind the bubbles, Hermann and Vincent for the interesting discussions on every kind of topic. Some special thanks goes also to my great friends Dimitar, Filippo, Matteo and Riccardo for the advices, motivations and the help they gave me during those several months. But most importantly thanks to all of you for the great time we spent together!

Finally, my sincere gratitude goes to all my other friends, my flatmates Florian and Thomas, my girlfriend Patty and my family. Thanks! Without you all, I would have had no chance to accomplish this amazing and unrepeatable journey.

Contents

Chapter 1

Introduction

In the whole financial market history and especially in the last years we had many clear proofs that the well-known efficient markets hypothesis widely spread in the economic theory, cannot be applied. The presence of phases where stock prices are showing extraordinary growth, known as bubbles, could be considered as one of the most studied example of such a failure.

The idea behind the study of financial bubbles is to understand price deviations from their fundamental value and more precisely to model their dynamics. In chapter 2 the model developed by Malevergne and Sornette [2014] is introduced and positioned within the general jump-diffusion framework. In their novel approach a stock is considered in a bubble if its price is currently too far beyond what would have been its expected growth over a certain time span $[t - \tau, t]$ and a mispricing metric will be introduced to measure such a distance. Obviously the larger the mispricing the larger the probability of a crash at the next time step and this will lead consequently to a larger conditional expected return aiming to introduce positive feedbacks in price dynamics.

The way in which the crash-hazard rate is modeled is used to correct one unrealistic and misleading assumption made in most existing jump-diffusion models and especially in the rational expectation ones. The wrong idea behind them is that the crash-hazard rate is considered to be proportional to the conditional expected return, underestimating the risks whenever the latter is going to zero. Non-arbitrage condition holding true in time of growing bubbles is an additional point that Malevergne and Sornette [2014] model is willing to correct allowing for heterogeneous collection of traders, everyone with his own risk profile.

Due to the (usually) low frequency of jumps in "standard" market conditions, positioning their appearances within rare events, and given the great attention Poisson processes have drawn in modeling extremes, a small part of chapter 2 has also been dedicated to their mathematical construction and simulation.

The present work is mainly interested in estimating the parameters of the aforementioned model. Two different approaches are pursued and they are both described in chapter 3. The first one is basically an extension of Berntsen [2015] where the dual estimation problem is tackled deploying an online version of a stochastic optimisation (SPSA) algorithm included within a Sequential Monte Carlo (SMC) framework. The second approach instead is a modified profile likelihood method implemented following the guidelines in Filimonov, Demos, and Sornette [2016].

Different simulations of the model, Particle Filter and optimisation methods are performed, compared and summarized in chapter 6. Application to financial data is then showed in chapter 7 where Malevergne and Sornette [2014] model is compared to a simpler GARCH(1,1), whose estimation is presented in the Appendix, to estimate VaR and ES leading to quite interesting results.

Chapter 2

Models of Bubbles and Crashes

The idea of the following chapter is to introduce a discrete-time version of a jump-diffusion model developed by Malevergne and Sornette [2014]. Since it is part of the same framework a general introduction to jump-diffusion models and their specific use in modeling bubbles and crashes in financial markets is also given in section 2.1 together with a brief digression on point processes. Section 2.2 is then dedicated to the introduction of rational expectation models. In section 2.3, we propose a model with which we attempt to correct the unrealistic assumptions of the aforementioned class of rational expectation models.

2.1 Jump-Diffusion Models

Over the last few years much attention has been given to jump-diffusion models by academic researchers and practitioners from major banks and financial institutions. Economists like Blanchard and Watson [1982] have tried to explain and understand consistency between rationality and deviation of assets prices from their fundamental values, coming to the conclusion that in many markets, phenomena such as runaway asset prices and crashes have foundation built upon the well-known rational expectations theory. Later in the 80*s* Weil [1987] considered and formalized a psychological phenomenon important during chaotic periods, for example bubbles, that is *trust*. He observed that stochastic bubbles on intrinsically useless assets may exist only if enough faith is pinned on its persistence in the future.

Moving forward to a more mathematical approach the general framework developed by Black and Scholes is widely known to be far too simplistic and not to be able to capture all the features showed by empirical financial time series like leptokurticity, volatility clustering and implied volatility smile just to mention but a few. Its limitations stem from the fact that Black and Scholes were willing to model stock return dynamics deploying only basic

stochastic processes with continuous path. In order to overcome these drawbacks, Merton [1976] extended the diffusion framework introducing jumps in the dynamics aiming to retrieve the discontinuity observed in financial data.

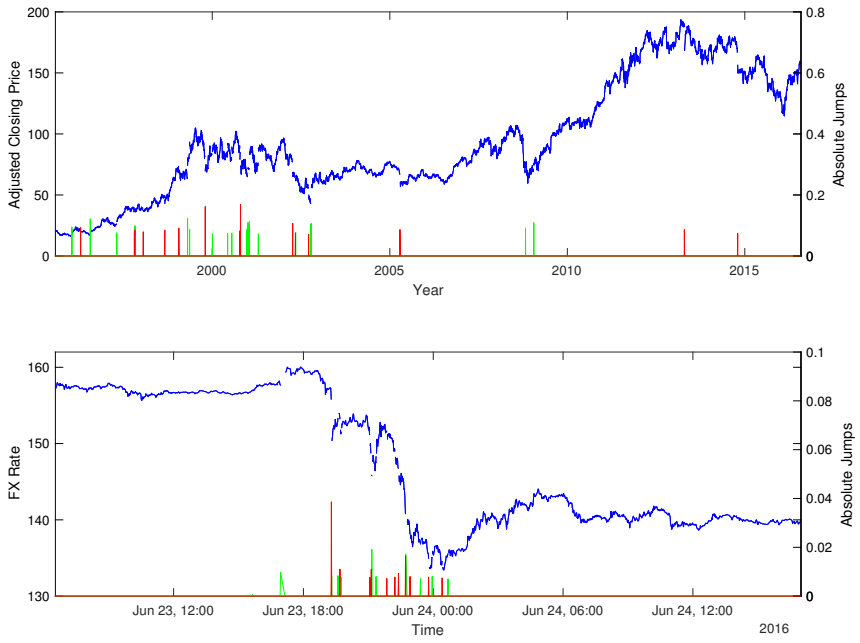

Figure 2.1: Top: IBM daily adjusted closing prices from January 1995 to January 2016. Threshold for (absolute) jumps is set to 7%. Bottom: GBP/JPY FX Rate tick data during Brexit week, 23-24 June 2016. Threshold for jump set to 0.7%. Green and red bars denote positive and negative jumps respectively. Reference for jumps always on the right y axis.

In the general jump-diffusion framework log-prices are indeed described by a mixture of a Brownian motion and a Poisson point process, named respectively, diffusion and jump part. Although during unstable historical periods and especially after the last financial crisis jumps are appearing with more and more frequency, they are still considered in research as part of the branch dealing with rare events and point processes are pretty well-known means used to model them. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ a measurable state-space where the points exist. For a sequence $\{Z_i\}_{i\geq 1} \in E$ and $\forall A \in \mathcal{E}$ a counting measure on $\mathcal{E}$ is defined by

$$m(A) = \sum_{i:Z_i \in A} 1 = \#\{i : Z_i \in A\}. \tag{2.1}$$

Additionally if for all compact sets $K \subset E$ the value $m(K)$ is finite, eq. 2.1 defines a point-measure. Let now $M_p(E)$ be the space of all point-measures on E equipped with an appropriate $\sigma-$algebra $\mathcal{M}_p(E)$. We can now give the followings:

Definition 2.1 *A point-process on E is a measurable map*

$$N : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow [M_p(E), \mathcal{M}_p(E)] .$$

$$\omega \longmapsto N(\omega) = m_\omega(\cdot) .$$

Definition 2.2 *A point-process N is called Poisson Point Process (or Poisson Random Measure) on $\mathcal{E}$ with intensity measure (or mean measure) Λ if:*

- *for $A \in \mathcal{E}$ and $K \geq 0$,*

$$\mathbb{P}(N(A) = K) = \begin{cases} \frac{\Lambda(A)^K}{K!} \cdot e^{-\Lambda(A)} & \text{if } \Lambda(A) < \infty , \\ 0 & \text{if } \Lambda(A) = \infty ; \end{cases} \tag{2.2}$$

- *for any $m \geq 1$ if $A_1, \cdots, A_m$ are mutually disjoint sets in $\mathcal{E}$ then the random variables $N(A_1), \cdots, N(A_m)$ are iid.*

Poisson processes have piecewise constant trajectories, are right-continuous with left limits and constant jumps size equal to 1 where the intervals between them are exponentially distributed.
Three (intrinsic) properties follow immediately from the above definitions

1. points occur independently of one another;
2. the occurrence of a point $x \in E$ neither encourages nor inhibits the occurrence of other points;
3. *complete randomness*, that is lack of interaction between different regions and the points in general, driven by the variation of the intensity measure Λ.

Since the main focus of this thesis is in finance, it makes more sense to consider a process with not only a single possible jump size. A generalization is then given by the compound Poisson process where the waiting times between jumps are exponential, but the jump sizes can have an arbitrary distribution. More precisely, let N be a Poisson process with parameter λ

and $\{Y_i\}_{i\geq 1}$ be a sequence of independent random variables with law f. The process

$$X_t = \sum_{i=1}^{N_t} Y_i \ ,$$

is called compound point process.

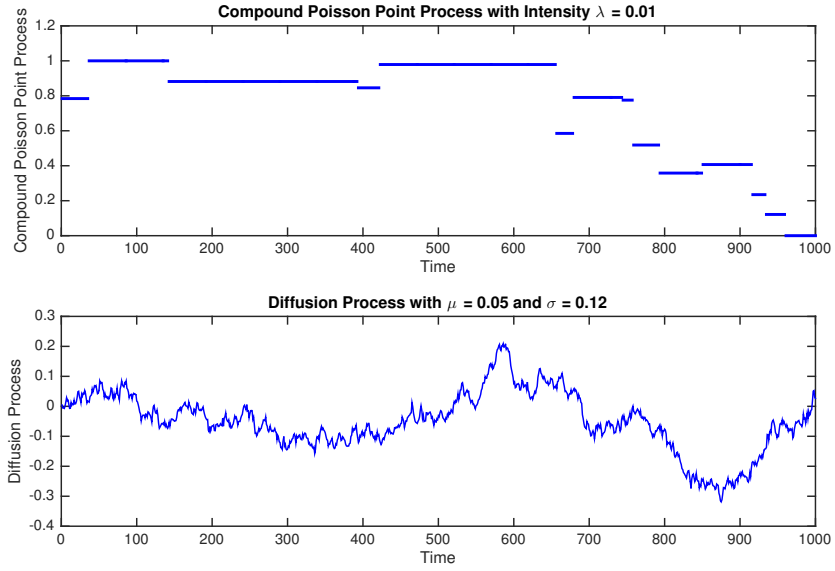

Figure 2.2: Top: sample path of a compound Poisson point process with normally distributed jump sizes. Bottom: simulated Brownian motion trajectory.

As we have already specified in order to obtain a simple jump-diffusion process it is enough to combine a Brownian motion with drift and a compound Poisson process. In finance one of the best known model of this type is the Merton model [Merton, 1976] where the stock price dynamics is given by

$$\frac{dS_t}{S_{t-}} = \mu dt + \sigma dW_t + d\left(\sum_{i=1}^{N_t} Y_i\right), \tag{2.3}$$

where $S_{t-} = \lim_{s\uparrow t} S_s$ and W_t is a standard poisson process. As described in Eraker, Johannes, and Polson [2003] and Christoffersen, Jacobs, and Mimouni [2007] a more general definition of a jump-diffusion model is assuming that the logarithm of stock prices, $X_t = \log S_t$ and the underlying variance, V_t, jointly solve the following system of stochastic differential equations:

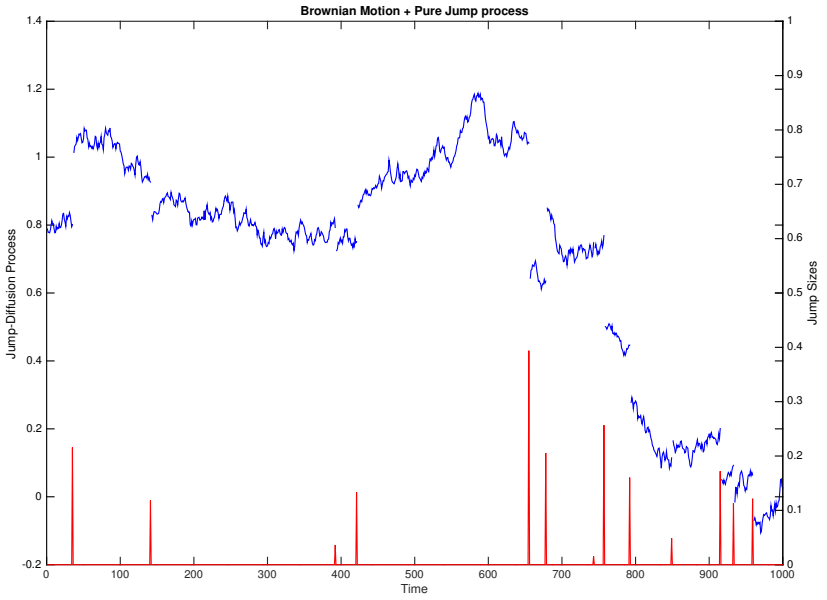

Figure 2.3: Blue line (left scale): simulated sample path of a jump-diffusion process (Brownian motion + compound Poisson). Red bars (right scale): jumps generated by the compound Poisson point process.

$$\begin{pmatrix} dX_t \\ dV_t \end{pmatrix} = \begin{pmatrix} \mu - \frac{1}{2}V_{t-} \\ \alpha(V_{t-}) \end{pmatrix} dt + \sqrt{V_{t-}} \begin{pmatrix} 1 & 0 \\ \rho\sigma_v & \sqrt{(1-\rho^2)} \cdot \sigma_v \end{pmatrix} dW_t + \begin{pmatrix} \xi^x dN_t^x \\ \xi^v dN_t^v \end{pmatrix}, \tag{2.4}$$

where $V_{t-} = \lim_{s \uparrow t} V_s$, W_t is a standard Brownian motion in $\mathbb{R}^2$ where the two components dW_t^x and dW_t^v have correlation ρdt, N_t^x and N_t^v are Poisson processes with constant intensities λ_x and λ_v, and ξ^x and ξ^v are the jump sizes in returns and volatility, respectively. Just to be a bit more clear ρ controls the asymmetric correlation in the leverage effect, that is the empirical evidence showed in Black [1976] that leverage of firms increase proportionally to the decrease in the share values. Most of the popular model used in finance for option pricing and hedging like Heston, pure stochastic volatility (SV), SVJ that incorporate jumps in returns and also the more complex SVIJ that include correlated jumps within both variance and return dynamics, are included in the specification (2.4).

The following Euler discretization scheme with interval $\Delta = 1$ is used in order to solve the equation (2.4) with $r_{t+1} = X_{t+1} - X_t$ describing asset log-returns

$$
\begin{aligned}
r_{t+1} &= \mu - \frac{1}{2}V_t + \sqrt{V_t}\varepsilon^x_{t+1} + \xi^x_{t+1}J^x_{t+1} \,, \\
V_{t+1} &= V_t + \alpha(V_t) + \sigma_v\sqrt{V_t}\varepsilon^v_{t+1} + \xi^v_{t+1}J^v_{t+1} \,,
\end{aligned}
\tag{2.5}
$$

where the two shocks $\varepsilon^x_{t+1} = W^x_{t+1} - W^x_t$ and $\varepsilon^v_{t+1} = W^v_{t+1} - W^v_t$ are bivariate normally distributed with zero mean, unit variance and correlation ρ, and J^x_{t+1} and J^v_{t+1} indicate jump arrivals following Bernoulli random laws with success probabilities λ_x and λ_v respectively.

2.2 Rational-Expectation models of bubbles

Following rational expectation theory, the cumulative effect of people's forecast about any future behavior is considered optimal, since it has been using all the available data and information in the market. Might those forecasts be limited by any extent? Are market participants "smart" in a "rational-expectation way"?

In the discrete-time economy considered here, only one speculative asset is traded and for simplicity interest rate, information asymmetry, risk aversion and frictions are ignored, while all efficient market hypothesis hold true. Within this simple and stylized framework, rational expectations are matching the following martingale condition:

$$
\forall T > t \,, \quad \mathbb{E}_t[p(T)] = p(t) \,,
\tag{2.6}
$$

where $p(t)$ stands for the price of the asset at time t and $\mathbb{E}_t[\cdot]$ denotes the conditional expectation. Since risk aversion is ignored, there is no distinction between physical and risk neutral probability. No dividends are paid by the asset. Its fundamental value at time t is then equal to 0 leading to a speculative bubble for every positive value of $p(t)$. We can now introduce a probability of a crash simply adding a supplementary term j denoting a jump process

$$
j = \begin{cases} 0 & \text{before the crash} \,, \\ 1 & \text{afterwards} \,. \end{cases}
\tag{2.7}
$$

Following the guidelines in Johansen, Ledoit, and Sornette [2000] the cdf and the pdf of the time of the crash are called $Q(t)$ and $q(t) = \frac{dQ}{dt}$ respectively. The hazard rate is then defined as

$$
h(t) = \frac{q(t)}{1 - Q(t)} \,.
\tag{2.8}
$$

Eq. (2.8) describes the probability per unit of time that the crash will happen in the next instant if it has not happened yet. Assuming that in case of a crash, the price drops by a percentage $\kappa \in (0,1)$, asset price dynamics before the crash are given by:

$$dp = \mu(t)p(t)dt - \kappa p(t)dj \ . \tag{2.9}$$

As already pointed out at the beginning of the section, martingale condition (2.6) must hold and so $\mu(t)$ has to be chosen accordingly. More precisely taking the conditional expectation on both sides of eq. (2.9) and observing that

$$\mathbb{E}_t[dj] = 1 \cdot h(t)dt + 0 \cdot (1 - h(t)dt) \ , \tag{2.10}$$

we get to the following equality

$$\mathbb{E}_t \left[\frac{dp/dt}{p(t)} \right] = \mu(t) - \kappa h(t) = 0 \quad \Longrightarrow \quad \mu(t) = \kappa h(t) \ . \tag{2.11}$$

Assuming no jump between t_0 and t and plugging the previous equation into eq. (2.9), we get an ODE whose solution is given by

$$\log \left[\frac{p(t)}{p(t_o)} \right] = \kappa \int_{t_0}^{t} h(s)ds \ . \tag{2.12}$$

A big drawback concerning the whole rational-expectation theory is that, as mentioned above, it imposes proportionality between the conditional expected return and the contemporaneous crash hazard rate. An immediate implication is that when the asset price stabilizes, this relationship may seriously underestimate the underlying risks faced by the investors. In order to correct this deficiency, the new jump-diffusion model developed in Malevergne and Sornette [2014] is introduced in the next section. It argues indeed that the unrealistic condition matching instantaneously return and risk does not hold true especially in uncertain times of growing bubbles in which perfect markets and no friction are assumptions that should be avoided.

2.3 Non-local behavior models

In the simplified discrete-time economy considered here only a risk free and a risky asset are traded. The former has of course a rate of return equal to r_f while the risk of the latter one is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ enriched with the filtration $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$. The log-return of the risky asset at time t is described by the following dynamic

$$\log\left(S_t/S_{t-1}\right) = r_t = \mu_t + \sigma_t \cdot \varepsilon_t - \underbrace{\kappa \cdot J_t \cdot I_t}_{\text{correction term}} , \tag{2.13}$$

where

- S_t is the price of the risky asset;
- μ_t and σ_t are drift and volatility at time t, respectively;
- $\varepsilon_t \sim \mathcal{N}(0,1)$ independent of $\mathcal{F}_{t-1}$;
- $\kappa > 0$ is the average jump size;
- $J_t \sim \text{Exp}(1)$ independent of $\mathcal{F}_{t-1}$;
- $I_t \sim \text{Bern}(\lambda_t)$, where λ_t denotes the conditional success probability and is equal to the log-logistic (or Fisk) function evaluated in a time-changing mispricing index;
- random variables ε_t, J_t and I_t are independent conditional on $\mathcal{F}_{t-1}$.

The correction term in eq. (2.13) avoids underestimating the downside risk that may be faced by investors when regime changes and asset price is stabilized. As specified also in Malevergne and Sornette [2014] different parameter settings within this framework, may lead to some well-known models such as

μ_t	σ_t	κ	λ_t	Model
μ	σ	0		Geometric Random Walk
μ	GARCH(p,q)	0		GARCH model
μ	σ		λ	most Rational-Exp Bubbles models

Under the assumption that the expected return at time t conditional on the available information at $t-1$ is constant and equal to $\bar{r}$ and that r_t admits stationary distribution with finite second moment as described in Malevergne and Sornette [2014] we get

$$\begin{aligned} \bar{r} &= \mathbb{E}_{t-1}[r_t] , \\ &= \mathbb{E}_{t-1}[\mu_t + \sigma_t \cdot \varepsilon_t - \kappa \cdot J_t \cdot I_t] , \\ &= \mu_t - \kappa \cdot \mathbb{E}_{t-1}[J_t] \cdot \mathbb{E}_{t-1}[I_t] , \\ &= \mu_t - \kappa \cdot \lambda_t , \end{aligned} \tag{2.14}$$

where $\mathbb{E}_{t-1}[J_t] = 1$ by standardization. Recalling that $\mu_t = \bar{r} + \kappa\lambda_t$, it is easy to see that the return required by the investors namely μ_t, increases

together with the time varying crash-hazard rate λ_t. This first conclusion completely agrees with the well-known economic theory which claims that higher risk implies higher return. An additional and maybe more important point within the financial bubble framework is also implicitly sustained by the above equation. Due to the high return and since there is no deterministic evidence about the end of the bubble or in general about a change of regime, for every investor it makes more sense to keep the risky position open, inflating the speculative bubble even further.

As previously pointed out, volatility clustering is a widely accepted stylized fact about empirical data, giving us a bit of predictability power. Within our model volatility is described by σ_t and the process we chose to model its dynamics and capture its most important characteristics without deploying too complicated means, is a GARCH(1,1). Following its standard specifications, the dynamic of the variance of the risky asset is described by

$$\sigma_t^2 = \omega + \alpha \cdot (r_{t-1} - \bar{r})^2 + \beta \cdot \sigma_{t-1}^2 \,, \quad \text{where } \omega, \alpha, \beta \geq 0 \,. \tag{2.15}$$

In this case $\sigma_t \cdot \varepsilon_t$ is said to follow a GARCH(1,1) process. This model is able to capture volatility persistency, since $|r_t|$ could be large if either $|r_{t-1}|$ or σ_{t-1} is large. Covariance-stationarity for such a model is guaranteed by *Proposition 4.21* in McNeil, Frey, and Embrechts [2015] that is

Proposition 2.3 *The GARCH(1,1) process defined in 2.15 is a covariance-stationary white noise process if and only if $\alpha + \beta < 1$. Its variance is then given by*

$$\bar{\sigma}^2 = \frac{\omega}{1 - \alpha - \beta} \,. \tag{2.16}$$

Plugging in eq. 2.16 in 2.15 will lead to the following reparametrization of the process

$$\sigma_t^2 = \bar{\sigma}^2 \cdot (1 - \alpha - \beta) + \alpha \cdot (r_{t-1} - \bar{r})^2 + \beta \cdot \sigma_{t-1}^2 \,. \tag{2.17}$$

Moving from $\theta = (\omega, \alpha, \beta)$ to $\bar{\theta} = (\bar{\sigma}^2, \alpha, \beta)$ is pretty useful especially from an estimation stability point of view. Every optimization routine is indeed really sensitive to the starting values and considering $\bar{\theta}$, sample variance could be taken as a starting point for the parameter $\bar{\sigma}^2$. Additionally as shown in Appendix A, when computing the gradient with respect to the parameter vector $\bar{\theta}$ we will get also components with more similar order of magnitude compared to the ones taken with respect to θ. A big difference within the components of the gradient could lead to optimization problems when the minimization routine is deployed, since too less importance is given to the smaller values.

The last term in eq. 2.13 is the one we called *correction term* and it is probably the most innovative part of the whole model together with the *mispricing metric* we are going to introduce later on in the section. This correction part is dealing with the jumps and their arrivals times described by a Bernoulli random variable I_t, whose success probability λ_t, also known as *crash-hazard rate* is time-varying and asymptotically follows a logistic-normal distribution. A non-local estimation procedure is also included in the way in which the term λ_t is deduced. As we have already specified in the introduction this novel approach is considering a given stock in a bubble if its price is currently too far beyond what would have been its expected growth over a certain time span $[t-\tau, t]$. To measure such a distance an investor may consider the following metric

$$\delta_{t,\tau} = \frac{S_t}{S_{t-\tau}} \cdot e^{-\tau \bar{r}} \,, \tag{2.18}$$

where S_t is the price of the asset a time t and using eq. 2.14 together with the *tower property* of the conditional expectation we can define $\bar{r}$ as the unconditional expected return of the risky asset, that is $\bar{r} = \mathbb{E}[r_t]$. The metric above is essentially a discounted increment over the considered time span and it is known as an *anchoring on price*. Simply considering $\delta_{t,\tau}^{1/\tau}$ instead of $\delta_{t,\tau}$ *anchoring on return* could also be considered following the same idea.
Once more, the above construction is putting emphasis on the "positive-feedback circle" that is shown in real price dynamics. Abnormal growth in price implies higher risk and higher risk generates a larger probability of a crash at next time step. Then higher risk subsequently leads to a higher expected return required by investors given that no crash will happen. Until prices go back to their fundamental values the loop does not stop pushing stock prices up with a super-exponential growth.

Within our framework mispricing is determined on a time interval $[t-\tau, t]$. That is for different choices of τ we will get a different point of view matching the real world situation in which every trader has his own perception and could use different time scales to define his strategy, from high-frequency to buy-and-hold. In order to assess those differences, starting from eq. (2.18) Malevergne and Sornette [2014] noticed that

$$\log\left[\delta_{t,\tau}^{1/\tau} e^{\bar{r}}\right] = \frac{1}{\tau}\sum_{k=0}^{\tau-1} \log\frac{S_{t-k}}{S_{t-k-1}} \,, \tag{2.19}$$

is essentially a moving average over the rolling window $[0, \tau]$ and they further observed that a similar result could be obtained using an exponentially-weighted moving average

$$\begin{aligned}\log(\delta_{t,a} \cdot e^{\bar{r}}) &= (1-a)\sum_{k=0}^{\infty} a^k \cdot \log\frac{S_{t-k}}{S_{t-k-1}} \ , \\ &= (1-a)\cdot\log\frac{S_t}{S_{t-1}} + a\cdot\log(\delta_{t-1,a}\cdot e^{\bar{r}}) \ ,\end{aligned} \tag{2.20}$$

hence

$$\log\delta_{t,a} = (1-a)\cdot\left(\log\frac{S_t}{S_{t-1}} - \bar{r}\right) + a\cdot\log\delta_{t-1,a} \ . \tag{2.21}$$

Parameter a, that in Berntsen [2015] was shown to be one of the most difficult to calibrate, is merging traders heterogeneity, trying to describe the point of view of the whole market participants about a specific mispricing.

Now it is finally possible to introduce the *mispricing index*

$$X_t := \frac{1}{s}\log\frac{\delta_{t-1,a}}{d} \ , \tag{2.22}$$

and using eq. (2.21) we get

$$X_t = \underbrace{-\frac{\log d}{s}}_{\bar{X}}\cdot(1-a) + a\cdot X_{t-1} + \underbrace{\frac{1-a}{s}}_{\eta}\cdot(r_{t-1}-\bar{r}) \ , \tag{2.23}$$

where in the two previous equations d and s define respectively the mispricing threshold above which a stock is considered "too" far from its fair value and the uncertainty about it. Formally instead, d is the scale parameter while s is the shape parameter of the log-logistic function

$$\begin{aligned}F(\delta) &= \frac{1}{1+\left(\frac{\delta}{d}\right)^{-\frac{1}{s}}} \ , \\ &= \frac{1}{1+\exp(-\frac{1}{s}\log\frac{\delta}{d})} \ , \\ &= L\left(\frac{1}{s}\log\frac{\delta}{d}\right) \ ,\end{aligned} \tag{2.24}$$

where $L(\cdot)$ is the standard logistic function given by $L(x) = \frac{1}{1+e^{-x}}$. Moreover, defining $\lambda_t := L(X_t)$ we finally specify one of the most important

relationship of the whole model, that is the link between mispricing index and crash-hazard rate. As an additional remark Aitchison and Shen [1980] showed that λ_t thus defined, follows asymptotically the logistic-normal distribution, whose density reads

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma x(1-x)} \cdot \exp\left[-\frac{1}{2\sigma^2}\left(\ln\frac{x}{1-x} - \mu\right)^2\right] , \tag{2.25}$$

where μ and σ are respectively, the mean and the variance of the mispricing index

$$X_t \sim \mathcal{N}\left(\bar{X}, \frac{\eta^2}{1-a^2} \cdot \mathrm{Var}(r_t)\right) . \tag{2.26}$$

To conclude the section the dynamics of the jump-diffusion model introduced by Malevergne and Sornette [2014] are the following

$$\begin{aligned} r_t &= \bar{r} + \kappa \cdot L(X_t) + \sqrt{V_t} \cdot \varepsilon_t - \kappa \cdot J_t \cdot I_t , \\ X_t &= (1-a) \cdot \bar{X} + a \cdot X_{t-1} + \eta \cdot (r_{t-1} - \bar{r}) , \\ V_t &= \bar{\sigma}^2 \cdot (1-\alpha-\beta) + \alpha \cdot (r_{t-1} - \bar{r})^2 + \beta \cdot V_{t-1} , \end{aligned} \tag{2.27}$$

where $\lambda_t = L(X_t) = \frac{1}{1+e^{-X_t}}$ while the three stochastic processes $J_t \sim \mathrm{Exp}(1)$, $I_t \sim \mathrm{Bern}(\lambda_t)$ and $\varepsilon_t \sim \mathcal{N}(0,1)$ are independent conditional on the available information at time $t-1$.

The aim of this master thesis is to calibrate the model described by the above dynamics. In order to clarify once more the meaning of the 8 parameters we want to estimate, here is a brief description of each one of them:

- $\bar{r}$ is the unconditional expected return of the risky asset;
- κ is the average jump size;
- a is the constant smoothing factor of the exponential moving average. It represents the degree of weighting decrease, meaning a higher a discounts older observations faster;
- $\bar{X}$ is the average value for mispricing. It is proportional to the threshold d and inversely proportional to the uncertainty s;
- η depends on both the smoothing factor a and the fuzziness parameter s. It regulates the variance dependency of the mispricing index;
- $\bar{\sigma}^2$ is the long-time variance of the GARCH(1,1) process;

- α and β are two parameters of the GARCH(1,1) model driving the persistency of the volatility through time. The former describes this dependency on past squared returns while the latter on past volatilities.

Chapter 3

Dual estimation

Let's say we have a model involving unobservable variables and unknown parameters. How could we estimate the parameters and extract the hidden states? Such a problem, called *dual estimation problem*, is well-known in mathematics and engineering and an extensive literature from many fields has been dealing with it since decades. In the following chapter the problem is introduced and an online estimation procedure is adapted for the model at hand. The first section presents a Sequential Monte Carlo (SMC) method implemented by Berntsen [2015] and used to extract hidden states, precisely jump times and sizes, from discretely realised stock prices. Section 3.2 first explains the difference between online and offline parameter estimation and then illustrate how the Simultaneous Perturbation Stochastic Approximation algorithm developed by Spall [1998] is structured and how it is tailored to the filtering framework given in section 3.1

3.1 State space filtering

State-space models, also known as hidden Markov models, are used in different fields of research like econometrics, biology, engineering and many others with applications spanning from stochastic volatility models to neuroscience or biochemical network models. In a general framework, a state-space model with an unknown parameter $\theta \in \Theta$, is formally defined by an observable stochastic process, say $\{r_t\}_{t\in\mathbb{N}_{>0}}$ and an unobserved one, say $\{L_t\}_{t\in\mathbb{N}_{>0}}$ with Markov transition density $g_\theta(L_t|L_{t-1})$. Hence their two dynamics are defined as

$$r_t = f_\theta(L_t, \varepsilon_t^y) \,, \tag{3.1}$$

$$L_t = g_\theta(L_{t-1}, \varepsilon_t^x) \,, \tag{3.2}$$

where ε_t^y and ε_t^x describe respectively, the noise in the observed data and the state shock. In the stochastic volatility models for example we can identify volatility with the unobservable state L_t and define r_t as the asset's observed log-return. The idea of such models is to extract latent information from the observed dynamics using some statistical (Bayesian) model. Using those information, we would then be able to capture unobserved changes in the underlying (hidden) process and use them to infer and follow a different and more precise dynamic. In this first section we assume that the parameter $\theta \in \Theta$ defining the model is known, while section 3.2 will show how to deal with its estimation when the unkown case is analyzed.

3.1.1 Optimal Filtering

Filtering is basically the process of estimating the state vector at time t, given all the (noisy) observations collected up to that time, namely $\{r_i\}_{i=1}^t$. In a Bayesian framework, it means that $p(L_t|r_1, \cdots, r_t)$, which to ease the notation will in future be specified as $p(L_t|r_{1:t})$, can be computed by the following two-step recursion.

(i) *Prediction step:* $p(L_t|r_{1:t-1})$ is computed via

$$p(L_t|r_{1:t-1}) = \int p(L_t|L_{t-1})p(L_{t-1}|r_{1:t-1})dL_{t-1} \,, \tag{3.3}$$

where the distribution $p(L_t|r_{1:t-1})$ can be seen as the prior over L_t when the most recent measurement r_t is not yet available.

(ii) *Update step:* the posterior over L_t is obtained using Bayes' rule to update the prior with the new measurement r_t, that is

$$p(L_t|r_{1:t}) \propto p(r_t|L_t)p(L_t|r_{1:t-1}) \,. \tag{3.4}$$

As pointed out in Kantas, Doucet, Singh, Maciejowski, Chopin, et al. [2015] both online and offline analytical inference on the state process $\{L_t\}$ based on the observed dynamic $\{r_t\}$, are feasible only for simple linear Gaussian state-space models. When non-linear non-Gaussian scenarios are analyzed instead, approaches like Extended Kalman Filter and Gaussian sum filter together with Markov Chain Monte Carlo (MCMC) methods have been proposed, for example in Alspach and Sorenson [1972]. As described in Berntsen [2015] for example, when dealing with more complex situations where the filtering density is not known analytically, we may rely on sequential Monte Carlo algorithms, also known as *particle filters*. Their wide

popularity stems from their pretty simple implementation and high parallelizable structure and especially due to their higher precision in the estimates compared to the standard alternatives [Cappé, Moulines, and Rydén, 2009], [Del Moral, 2004].

3.1.2 Particle Filter

Sequential importance sampling (SIS) is the basic Monte Carlo method used when prediction and update steps are not analytically tractable. To describe how SIS particle filter works and how it is adapted to our framework we will rely on the guidelines given in Berntsen [2015]. In this framework is now useful to consider the full posterior distribution $p(L_{0:t-1}|r_{1:t-1})$, rather than the filtering distribution, $p(L_{t-1}|r_{1:t-1})$. The idea of this method is to give a discrete approximation $\hat{p}(L_{0:t-1}|r_{1:t-1})$ to $p(L_{0:t-1}|r_{1:t-1})$, with a (normalized) weighted set of samples $\{\pi_{t-1}^{(n)}, L_{t-1}^{(n)}\}_{n=1}^{N}$ called *particles*, and recursively update these particles to obtain an approximation to $p(L_t|r_{1:t})$. The filtering density is then approximated via

$$p(L_{0:t-1}|r_{1:t-1}) \approx \sum_{n=1}^{N} \pi_{t-1}^{(n)} \delta_{L_{0:t-1}^{(n)}}, \tag{3.5}$$

where $\pi_{t-1}^{(n)} = p(L_{t-1}^{(n)}|r_{1:t-1})$ and $\delta_{L_{t-1}^{(n)}}$ is the Dirac function centered in $L_{t-1}^{(n)}$. As already stated the idea will be now to recursively update the particles in order to derive an approximation to $p(L_{0:t}|r_{1:t})$. To do that, let's now define a proposal density $q(L_{0:t-1}|r_{1:t-1})$ way easier to sample from, compared to $p(L_{0:t-1}|r_{1:t-1})$ and let's assume that the importance density $q(L_{0:t}|r_{1:t})$ can be factorized as

$$q(L_{0:t}|r_{1:t}) = q(L_{0:t-1}|r_{1:t-1})q(L_t|L_{0:t-1}, r_{1:t}) \ .$$

Then remembering that the state process is Markov and that the observations are conditionally independent, we can use the following result from Berntsen [2015] in order to define a recursive formula for the importance weights

$$\pi_t = \pi_{t-1} \frac{p(r_t|L_t)p(L_t|L_{t-1})}{q(L_t|L_{t-1}, r_t)} \ . \tag{3.6}$$

We finally obtained a discrete approximation $p^N(L_t|r_{1:t})$ of the filtering density $p(L_t|r_{1:t})$ via the following equations

$$L_t^{(n)} \sim q(L_t|L_{t-1}^{(n)}, r_t) , \tag{3.7}$$

$$\pi_t^{(n)} = \pi_{t-1}^{(n)} \frac{p(r_t|L_t^{(n)})p(L_t^{(n)}|L_{t-1}^{(n)})}{q(L_t^{(n)}|L_{t-1}^{(n)}, r_t)} . \tag{3.8}$$

As mentioned already, the algorithm we have just introduced is returning a discrete approximation $\{L_t^{(n)}, \pi_t^{(n)}\}_{n=1}^N$ of the posterior $p(L_t|r_{1:t})$, solving basically our particle filtering problem. Unfortunately this solution has an evident drawback, named *degeneracy*, as soon as the time t starts to increase. Degeneracy is typically measured by the value

$$\mathrm{N_{eff}} = \frac{1}{\sum_{i=1}^N (\pi_t^{(i)})^2} ,$$

said *effective sample size*, where a smaller $\mathrm{N_{eff}}$ means a larger variance for the weights. As the time t increase indeed, only few particles will have a significant weight, where all the others will have a weight close to 0. Sequential importance resampling (SIR) algorithm corrects this shortcoming with two simple refinements:

1. the proposal density $q(L_t|L_{t-1}^{(n)}, r_t)$ is taken to be the state transition distribution $p(L_t|L_{t-1}^{(n)})$;
2. at every iteration step resampling has to be applied.

In our implementation, step 2 is active at every iteration, but to increase efficiency it could be set dynamically in order to take place only whenever the value $\mathrm{N_{eff}}$ plunges below a pre-defined threshold. Given that resampling is done with replacement particles with large weigths are more likely to be drawn multiple times while the small-weighted ones tend to be "forgotten". A weight equal to $1/N$ is then assigned to all the newly resampled particles solving the original degeneracy problem. This result unfortunately comes with a price, cause *sample impoverishment* problem is then added to the method. Diversity of the particles will indeed tend to decrease after a resampling step, with the extreme case, that all particles might collapse into a single one. Regularized Particle Filter (RPF) has been proposed in Musso, Oudjane, and Le Gland [2001] as a potential solution to the problem.

Another important point that is worth to remark, is that using $p(L_t|L_{t-1}^{(n)})$ as a proposal density, SIR algorithm is not taking into account any information whenever a new observation r_t is made available. This drawback may be solved deploying another version of the Particle Filter, namely Auxiliary

Particle Filter (APF). The approach has been implemented and analyzed in Berntsen [2015], but the slightly better results obtained for the model at hand, were not enough to justify its increased complexity compared to the SIR method.

SIR Particle Filter

1: At time $t = 0$, for all $i \in \{1, \cdots, N\}$:
2: Initialize $L_t^{(i)}$ and $\pi_t^{(i)}$
3: At time $t \geq 1$, for all $i \in \{1, \cdots, N\}$:
4: Simulate $L_t^{(i)} \sim p(L_t | L_{t-1}^{(i)})$
5: Compute $\pi_t^{(i)} \leftarrow p(r_t | L_t^{(i)}) / \sum_{i=1}^{N} p(r_t | L_t^{(i)})$
6: Draw $z(i) \sim \text{Multinomial}(N; \pi_t^{(1)}, \cdots, \pi_t^{(N)})$
7: Set $L_t^{(i)} \leftarrow L_t^{(z(i))}$
8: Return:
9: $p(L_t | r_{1:t}) \approx p^N(L_t | r_{1:t}) \leftarrow \{L_t^{(i)}, \frac{1}{N}\}$

The last part of this chapter is dedicated to the introduction of an *online* optimization method, that could be directly implemented within the Sequential Monte Carlo algorithm introduced in section 3.1.2.

3.2 Online Parameter Estimation

As noticed in Kantas et al. [2015], no perfect method exists with regards to dual estimation problems. The same method could perform differently from one model to another and every situation should be tackled in a specific manner. One important distinction that has been done in Kantas et al. [2015] was between *online* and *offline* estimation. Precisely, the former approach is the one in which the estimate is done when new data is available during the operation of the model. It is typically performed using a recursive algorithm and it leads to values that may vary over time. Differently, the values estimated with the latter methods are not time varying, since estimation is done only once all the significant data has been collected. Even though *offline* approach is considered better in terms of statistical efficiency, we preferred here an *online* implementation. Financial data are indeed published and updated continuously and a "real time" estimation procedure has been considered to be more appropriate for the model at hand, aiming to define a dynamic calibration method.

3.2.1 Stochastic Optimisation

The original Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm was introduced and fully analyzed in Spall [1987] and Spall [1992]. The method is part of the more general stochastic optimization framework widely employed in many areas of engineering, physical and social sciences with applications spanning from model fitting and statistical parameter estimation, to adaptive control and pattern classification. Although in our specific case the gradient of the objective function is known in closed-form, see Malevergne and Sornette [2014], we decided to deploy a gradient-free method in order to achieve better performance at the expense of computational precision. Concerning the implementation of the method within SIR particle filter instead, guidelines from Chan, Doucet, and Tadic [2003] have been used as a general reference.

Let's now consider the problem of minimizing (or maximizing) a differentiable loss function $F : \mathbb{R}^8 \to \mathbb{R}$, that is equivalent to find a root θ^* of the gradient

$$g(\theta) = \frac{\partial F(\theta)}{\partial \theta} = 0 \, .$$

whose approximation is based on a highly efficient and easily implemented "simultaneous perturbation". Let $\theta = (\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a)$ be our 8-dimensional vector of parameters we need to optimize. SPSA requires all the single component of θ to be varied randomly, simultaneously and independently to obtain two estimates of the cost function (likelihood in our case). Among other well known finite-difference methods, here only two measurements of the loss function are required regardless the dimension of θ. Starting from an initial guess θ_0, SPSA algorithm iteratively update the parameters following:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \gamma_t \cdot \hat{g}(\hat{\theta}_{t-1}) \, , \quad t \in \{1, \cdots, T\} \, , \tag{3.9}$$

where $\hat{g}(\cdot)$ and $\{\gamma_t\}$ are respectively the *gradient approximation* and a *gain sequence* that will be introduced in the next sections.

Gradient approximation

Let's define the "simultaneous perturbation" estimate for the gradient and consider a 8-dimensional vector of mutually independent mean-zero random variables, namely Δ_t. Regular conditions specified in Spall [1992] have to be satisfied in order to guarantee strong convergence and asymptotically normality of $\hat{\theta}_t$. The most important one is probably that for all $i \in \{1, \cdots, 8\}$ $\mathbb{E}(|\Delta_{t,i}^{-1}|)$, or some higher-order inverse moment, needs to be bounded, which precludes $\Delta_{t,i}$ to be normally or uniformly distributed. Even though no additional assumption has to be made on the distribution of

$\Delta_{t,i}$, symmetric Bernoulli has been suggested in Spall [1992] as a preferred practical choice. At time t the "simultaneous perturbed" gradient is then given by

$$\hat{g}(\hat{\theta}_t) = \begin{pmatrix} \frac{y_t^{(+)} - y_t^{(-)}}{2c_t\Delta_{t,1}} \\ \vdots \\ \frac{y_t^{(+)} - y_t^{(-)}}{2c_t\Delta_{t,8}} \end{pmatrix}, \tag{3.10}$$

where

$$\begin{aligned} y_t^{(+)} &= F(\hat{\theta}_t + c_t\Delta_t) \,, \\ y_t^{(-)} &= F(\hat{\theta}_t - c_t\Delta_t) \,, \end{aligned} \tag{3.11}$$

and $\{c_t\}$ is the other gain sequence (of positive scalars) together with the previously introduced $\{\gamma_t\}$.

Gain Sequences γ_t and c_t

In every stochastic optimization method, the choice of the algorithm coefficients is critical to its convergence and SPSA is no exception in this regard. Optimal gain sequence selections is theoretically derived from asymptotical performance, but as shown in Wang [2013], for practical applications things become different. The two sequences, γ_t and c_t, are defined as

$$k_t = \frac{k}{(K+t+1)^\rho} \quad \text{and} \quad c_t = \frac{c}{(t+1)^\gamma} \,,$$

where the feasible domain for ρ and γ is defined in Spall [1992] by $\rho \leq 1$, $2\rho - 2\gamma > 1$ and $3\gamma - \rho/2 \geq 0$ together with their optimal asymptotical choices given by $\rho = 1$ and $\gamma = 1/6$. Accordingly to Wang [2013] for finite sample cases, sequence k_t should not be too small, that means smaller values for K and ρ (suggested equal to 0.602 in Spall [1992]) and a larger one for k. Letting N be the maximum number of iterations, we are obviously forced to choose K to be proportional to N in order to achieve some reasonable performance within the fixed limited number of iterations, that is $K = \varrho N$. Additionally since the effect of K should disappear in later iterations without leading to a too small gain step, ϱ has to be chosen from $\{0.1, 0.01, 0.001 \cdots\}$ while a larger k must be preferred. Finally a rule for k can be easily determined considering the performance of the algorithm in early iterations. Another important consideration must be added to the above discussion. Precisely, due to the flatness of the likelihood with respect to some of the parameters, both k_t and c_t should be defined as vector and not just scalar sequences, in order to accommodate the different changes of

magnitude of each parameter from one step to the other. To conclude, c is generally set equal to the standard deviation of the measurement noise, but since direct measurement is available, a small positive value is then assigned to c.

The last part of the chapter is used to present the *pseudo-code* describing step-by-step our SPSA implementation. Following this guidelines it should be fairly easy to implement the optimisation routine in every desired filtering framework.

SPSA Algorithm within SIR Particle Filter

Sequential Importance Sampling
At time $t = 0$, for all $i \in \{1, \cdots, N\}$:
Initialize $L_t^{(i)}$ and $\pi_t^{(i)}$
At time $t \geq 1$, for all $i \in \{1, \cdots, N\}$:
Simulate $L_t^{(i)} \sim p(L_t | L_{t-1}^{(i)})$
Compute $\pi_t^{(i)} \leftarrow p(r_t | L_t^{(i)}) / \sum_{i=1}^{N} p(r_t | L_t^{(i)})$
Cost function evaluation
Generate a 8-dimensional simultaneous perturbation vector Δ_t
Compute $\theta_t^{(i)+} \leftarrow (\theta_t^{(i)} + c_t \Delta_t^{(i)})$ and $\theta_t^{(i)-} \leftarrow (\theta_t^{(i)} - c_t \Delta_t^{(i)})$
Evaluate cost function $F(\theta_t^{(i)+})$ and $F(\theta_t^{(i)-})$
Gradient approximation
$\hat{g}(\hat{\theta}_t) \leftarrow \left(\frac{F(\theta_t^{(i)+}) - F(\theta_t^{(i)-})}{2c_t \Delta_{t,1}}, \cdots, \frac{F(\theta_t^{(i)+}) - F(\theta_t^{(i)-})}{2c_t \Delta_{t,8}} \right)^T$
Parameter Update
$\theta_{t+1}^{(i)} \leftarrow \theta_t^{(i)} + c_t \cdot \hat{g}(\theta_t^{(i)})$
Resampling
Draw $z(i) \sim \text{Multinomial}(N; \pi_t^{(1)}, \cdots, \pi_t^{(N)})$
Set $L_t^{(i)} \leftarrow L_t^{(z(i))}$ and $\theta_{t+1}^{(i)} \leftarrow \theta_{t+1}^{z(i)}$

Chapter 4

Offline parameter estimation

A different estimation approach is outlined in this chapter, where all the information related to the observable process r_t are assumed to be available by the time of the analysis. Following the conclusions that has been made in Malevergne and Sornette [2014] and Berntsen [2015], that is direct maximization of the cost function is not feasible due to both its sloppiness with respect to some of the parameters and to the high computational costs, a profile and modified profile likelihood approach are introduced and implemented following the guidelines given in Filimonov et al. [2016]. `DLIB` C++ library has been used to perform the maximization, deploying an unconstrained optimizer and defining artificial boundaries in order to satisfy the different constraints of some parameters. `float_128` data type has also been considered trying to take into account the highest numerical precision in order to detect fluctuations of the log-likelihood within its flat sections. Direct maximization lead to some interesting result, but profile likelihood method wasn't able to conclude anything valid from a statistical point of view.

4.1 Maximum Likelihood

The (usually) easiest and probably mostly used way of estimating a set of unknown parameters is through *maximum likelihood*. The idea behind it is pretty simple and straightforward. Let $\{r_t\}_1^T$ be an observable stochastic process depending on a static parameter vector θ and let $p(r_t|\theta)$ be the density function describing a single obervation. Assuming that $\{r_t\}_{t=1}^T$ are independent and identically distributed, the *likelihood function* of the process is equal to the joint density function $p(r_1, \cdots, r_T|\theta)$, that given our iid assumption is equal to

$$\mathcal{L}(\theta\,|r_1, \cdots, r_T) = \; p(r_1, \cdots, r_T|\theta) = \prod_{t=1}^{T} p(r_t|r_1, \cdots, r_{t-1}; \theta)\;. \tag{4.1}$$

For simplicity a log-transform of the previous equation, called *log-likelihood*, is usually considered

$$\ln \mathcal{L}(\theta \mid \{r_1, \cdots, r_T\}) = \sum_{t=1}^{T} ln\ p(r_t | r_1, \cdots, r_{t-1}; \theta)\ . \tag{4.2}$$

Within our framework the dynamics of the stochastic process r_t are given by the jump-diffusion model described by the eq. 2.27. Conditionally on $\mathcal{F}_{t-1}$, J_t and I_t, our process is then distributed as:

$$r_t | \mathcal{F}_{t-1}, J_t, I_t\ \sim\ \mathcal{N}\ (\mu_t - \kappa \cdot J_t \cdot I_t,\ \sigma_t^2)\ . \tag{4.3}$$

Integrating out both J_t and I_t from the above distribution following the guidelines in Malevergne and Sornette [2014] the conditional distribution of the log-returns reads

$$r_t | \mathcal{F}_{t-1}\ \sim\ \begin{cases} \text{EMG}(\mu_t, \sigma_t, \kappa_t) & \text{with probability } \lambda_t\ , \\ \mathcal{N}(\mu_t, \sigma_t^2) & \text{with probability } 1 - \lambda_t\ , \end{cases} \tag{4.4}$$

where $\text{EMG}(\mu, \sigma, \kappa)$ defines the Exponentially Modified Gaussian distribution with density

$$f_{\text{EMG}}(x | \mu, \sigma, \kappa) = \frac{\sigma}{|\kappa|} \cdot \varphi(x | \mu, \sigma) \cdot R\left(\text{sgn}(\kappa) \cdot \left(\frac{x - \mu}{\sigma} + \frac{\sigma}{\kappa}\right)\right). \tag{4.5}$$

The function $R(\cdot)$ given by

$$R(x) = \frac{1 - \Phi(x)}{\varphi(x | 0, 1)}\ , \tag{4.6}$$

can be easily implemented following the guidelines in Marsaglia [2004], while $\Phi(\cdot)$ and $\varphi(\cdot)$ represent respectively the cumulative distribution and the density function of a normal random variable. Log-likelihood is then described as

$$\begin{aligned} \ln \mathcal{L}(\theta \mid \{r_1, \cdots, r_T\}) = \sum_{t=0}^{T} \Bigg[& \lambda_t \cdot f_{\text{EMG}}\left(r_t | \mu_t, \sigma_t, \kappa_t\right) + \\ & + (1 - \lambda_t) \cdot \varphi\left(r_t | \mu_t, \sigma_t^2\right) \Bigg]\ . \end{aligned} \tag{4.7}$$

Our goal is to find θ^* that maximizes the function $\ln \mathcal{L}(\cdot)$. That is we wish to find the θ_{ML} such that

$$\theta_{ML} = \arg\max_{\theta} \mathcal{L}(\theta | r_1 \cdots r_T) \, . \tag{4.8}$$

As shown in Malevergne and Sornette [2014] and Berntsen [2015] direct maximisation of the likelihood doesn't perform well due to the *sloppiness* of some parameters with respect to the objective function, in the sense expressed by Waterfall, Casey, Gutenkunst, Brown, Myers, Brouwer, Elser, and Sethna [2006]. In his work P. Berntsen proved also that a *brute force approach* deploying a cross-section analysis of the 8-dimensional parameter space, would be too computationally expensive and time consuming without even spanning efficiently all the possibile parameter combinations. To tackle the sloppiness of the cost function Hessian matrix is then the only way to take into account and properly deal with the different orders of magnitude of the curvature of the likelihood with respect to the parameter.
To overcome the difficulties in dealing with a mixture of distributions is crucial to consider both J_t and I_t as latent states. Expectation-Maximization (EM) algorithm is considered one common technique to solve similar problems. More precisely, given an incomplete data set, EM algorithm is a general method of finding the maximum-likelihood estimate of the parameters of the underlying distribution. Incompleteness of data stands here for

- missing data in the set due to limitations (e.g. noise) of the observation process;
- likelihood is analytically intractable and extra but *hidden/latent* parameters need to be added to the data set.

Our framework is exactly within the latter application. Let indeed assume that we are not only observing log-returns $\{r_t\}_{t=1}^T$, but also jump times and sizes (latent states) $\{J_t, I_t\}_{t=1}^T$. With this assumption and recalling the dynamics of the model described in 2.27 the complete likelihood conditional on $\mathcal{F}_{t-1}$ is given by:

$$(r_t, J_t, I_t) \;\sim\; \mathcal{N}(\mu_t - \kappa \cdot J_t \cdot I_t, \sigma_t^2) \cdot \mathcal{E}(1) \cdot \lambda_t^{I_t} \cdot (1-\lambda_t)^{1-I_t} \, , \tag{4.9}$$

that is

$$\begin{aligned} \ln \mathcal{L}(\theta \mid \{r_t, J_t, I_t\}_{t=0}^T) = \sum_{t=0}^{T} \Bigg[& -\frac{1}{2} \ln 2\pi\sigma_t^2 - \frac{1}{2\sigma_t^2}(r_t - \mu_t + \kappa \cdot J_t \cdot I_t)^2 \\ & - J_t + I_t \cdot \ln\lambda_t + (1 - I_t) \cdot \ln(1-\lambda_t) \Bigg]. \end{aligned} \tag{4.10}$$

The idea behind the EM algorithm developed by Dempster, Laird, and Rubin [1977] is pretty simple and it works with the following two steps:

1. Expectation Step: it averages over the latent states according to the model estimation of their probability distributions;
2. Maximisation Step: the model parameters are estimated by maximising the averaged likelihood.

Unfortunately the aforementioned method has been already exploited and proved to be inefficient in Malevergne and Sornette [2014] due to log-likelihood *sloppiness* we mentioned before. Even though no satisfying results has been obtained, their approach was enlightening and extremely helpful for our calibration purposes. The auxiliary function $Q^{\theta|\theta_0}$ introduced below was shown indeed to be pretty useful to give an analytical expression of both the gradient and the Hessian of the log-likelihood function given in 4.10.

$$\begin{aligned} Q_t^{\theta|\theta_0} &= \mathbb{E}_{t-1}^{\theta_0}\Bigg[-\frac{1}{2}\ln\sigma_t^2 - \frac{(r_t-\mu_t+\kappa\cdot J_t\cdot I_t)^2}{2\sigma_t^2} - J_t + \\ &\quad + I_t\cdot\ln\lambda_t + (1-I_t)\cdot\ln(1-\lambda_t)|S_t\Bigg] , \\ &= -\frac{1}{2}\ln\sigma_t^2 - \frac{(r_t-\mu_t)^2}{2\sigma_t^2} - \frac{\kappa(r_t-\mu_t)}{\sigma_t^2}\cdot\mathbb{E}_{t-1}^{\theta_0}[J_t\cdot I_t|S_t] + \\ &\quad - \frac{\kappa^2}{2\sigma_t^2}\cdot\mathbb{E}_{t-1}^{\theta_0}[J_t^2\cdot I_t|S_t] - \mathbb{E}_{t-1}^{\theta_0}[J_t|S_t] + \\ &\quad + \mathbb{E}_{t-1}^{\theta_0}[I_t|S_t]\cdot\ln\lambda_t + \mathbb{E}_{t-1}^{\theta_0}[1-I_t|S_t]\cdot\ln(1-\lambda_t) , \end{aligned} \tag{4.11}$$

where $\mathbb{E}_{t-1}^{\theta_0}[\cdot]$ stands for the expectation conditional on $\mathcal{F}_{t-1}$ under the parameter θ_0. Gradient and Hessian implementation has been done following the guidelines in Malevergne and Sornette [2014]. Given the sensibility of our problem, an extreme precision is crucial while performing the calculations and `float_128` data type has been used in our first C++ implementation. Unfortunately, `DLIB` and `EIGEN` optimization libraries were unable to process `float_128` data type we went back to the "standard" `double` precision.

The following section will be an introduction to a hierarchical sequential optimization implemented accordingly to the work done by Filimonov et al. [2016]. Additional guidelines has been given also in Filimonov and Sornette [2013] where a similar estimation problem was faced, proving that a sequential optimization of the objective function may lead to a way less complex and more stable fitting procedure. With this method the cost function

should be indeed characterized by good smooth properties with in general a single minimum whenever the model fits properly the empirical data. In the first part of section 4.2 *profile likelihood* procedure is introduced, while in the second one a *modified* version of it is proposed and adapted to model at hand.

4.2 Profile Likelihood

To summarize, our aim is to solve the following

$$\mathcal{L}(\theta \mid \{r_t\}_{t=1}^{T}) \quad \longrightarrow \quad \max_{\theta}, \tag{4.12}$$

where $\mathcal{L}(\theta \mid r_1, \cdots, r_T)$ stands for the complete likelihood of the model over the whole observed time series $\{r_t\}_1^T$. The general idea behind the method we are interest in, it is to detect those parameters over which the cost function is more or less sensitive, that is equivalent to find a way to distinguish between *rigid* and *sloppy* parameters. In order to understand their sensitiveness, eigenvalues and eigenvectors relative to the Hessian matrix of the likelihood function are computed and ranked. Parameters corresponding to the largest eigenvalues are considered *rigid*, while the smallest ones are considered *sloppy*. Once parameters have been ordered and the ranking has been recognized, n clusters are created following a hierarchical classification of the 8-dimensional parameter space. After *rigid* (top) and *sloppy* (bottom) clusters have been detected, the following n-step sequential optimization is performed:

- fix the values of the sloppy parameters, θ_{sloppy};
- optimize the cost function with respect to the rigid ones

$$\mathcal{L}_p(\theta_{\text{sloppy}}) \;=\; \max_{\theta_{\text{rigid}}} \mathcal{L}(\theta_{\text{rigid}} \mid \theta_{\text{sloppy}}) \,. \tag{4.13}$$

Often the function $\mathcal{L}_p(\theta_{\text{sloppy}})$, known as *profile likelihood*, is wrongly considered a genuine likelihood, since it treats the sloppy parameters as if they were known. With this assumption the inference on θ_{sloppy} based on $\mathcal{L}_p(\theta_{\text{sloppy}})$ may indeed be inaccurate, leading to unstable estimates relatively to small perturbations in the observed data.

4.3 Modified Profile Likelihood

To correct the drawback highlighted in the last part of the previous section, a pretty useful and reliable approach called *modified profile likelihood* has been

introduced by Barndorff-Nielsen [1983]. This method is trying to correct the aforementioned deficiencies of the *profile likelihood*, multiplying the latter one with an additional factor, $M(\theta_{\text{sloppy}})$, that is

$$\begin{aligned}\mathcal{L}_m(\theta_{\text{sloppy}}) &= M(\theta_{\text{sloppy}}) \cdot \mathcal{L}_p(\theta_{\text{sloppy}}) \,, \\ &= |I(\hat{\theta}_{\text{rigid}|\text{sloppy}})|^{-\frac{1}{2}} \cdot \left| \frac{\partial \hat{\theta}_{\text{rigid}}}{\partial \hat{\theta}_{\text{rigid}|\text{sloppy}}} \right| \cdot \mathcal{L}_p(\theta_{\text{sloppy}}) \,,\end{aligned} \tag{4.14}$$

where $|\cdot|$ is the absolute value of a matrix determinant and $I(\theta_{\text{rigid}})$ is the observed Fisher information matrix on the rigid parameters, assuming the sloppy ones are known:

$$I(\hat{\theta}_{\text{rigid}|\text{sloppy}}) = -\frac{\partial \ln \mathcal{L}(\theta)}{\partial \theta_{\text{rigid}} \partial \theta'_{\text{rigid}}}\bigg|_{\theta_{\text{rigid}}=\hat{\theta}_{\text{rigid}|\text{sloppy}}} \,. \tag{4.15}$$

Intuitively, (4.15) is a penalty term correcting the overestimation that may arise in $\mathcal{L}_p(\cdot)$ and with its inclusion we are taking into account the curvature of the likelihood, crucial in our calibration framework since its *sloppiness* has been shown to be the main issue. To conclude, with the Jacobian term $J(\theta_{\text{sloppy}}) = |\partial \hat{\theta}_{\text{rigid}} / \partial \hat{\theta}_{\text{rigid}|\text{sloppy}}|$ *modified profile likelihood* acquires the fundamental property to be invariant with respect to any reparametrization of θ_{sloppy}. Despite some limitation due to the difficulties that may arise in evaluating the Jacobian term in many realistic models $\mathcal{L}_m(\theta_{\text{sloppy}})$ can be considered a genuine likelihood function. As pointed out by Filimonov et al. [2016], in comparison with other Bayesian approaches and integrated likelihood functions it also has the advantage to do not require any prior density specification for the *sloppy* parameters. Finally, since likelihood $\mathcal{L}(\theta)$ is meaningful up to some constants, its following normalized version called *relative likelihood*, is usually considered:

$$R(\theta) = \frac{\mathcal{L}(\theta)}{\max_\theta \mathcal{L}(\theta)} \in [0, 1] \,. \tag{4.16}$$

The same idea holds true for both *profile* and *modified profile likelihood*. An additional remark has to be done also with regards to the uncertainty in estimated parameters and to such a purpose our inference is based on the *likelihood ratio* $R(\theta)$ defined in eq. (4.16). MLE approach indeed provides us not only the estimated *points* of interest, but also the probabilistic *ranges* within which they could live given the observed data. We will rely here on the same procedure described in Filimonov et al. [2016]. Precisely, we will deploy the intuitive *likelihood interval (LI)* approach, allowing us to avoid any kind of regularity assumption on the model. To define such an interval

some arbitrary *cutoff* has to be chosen and only values of $R(\theta)$ above this fixed threshold, will be considered statistically valid. Hence

$$LI(\theta^*) = \left\{ \theta^* : R_m(\theta^*) = \frac{L_m(\theta^*)}{L_m(\hat{\theta}^*)} > 0.05 \right\}, \tag{4.17}$$

describes the *likelihood interval* at the 5% *cutoff* point while θ^* defines the sloppy parameter we are trying to profile. The problem with this approach is that due to the sloppiness of the likelihood with respect to the parameter of interest, no meaningful likelihood ratio may be defined. Indeed for every choice of the sloppy parameter in the feasible range the likelihood of the model won't be particularly affected, leading to a likelihood interval as wide as the original parameter domain.

Now that we have introduced some of the tools needed to estimate the parameters of a given model, in the following chapter we will assume that a successful calibration method has been performed already. Two of the most important risk measures will be introduced together with some of their most important properties and some statistical backtesting techniques to assess their reliability.

Chapter 5

Risk Measures

This chapter is an introduction on how to measure and estimate, from a statistical point of view, the risks associated with a given portfolio. The main topic at the beginning will be the loss operator used to express the variation of the portfolio value in terms of risk factor changes. More precisely the first section introduces some basic probabilistic and statistical tools that are then used to define risk measures, like Value at Risk (VaR) and Expected Shortfall (ES). In section 5.2 some common estimation techniques are presented, while some pitfalls and backtesting methods are then introduced in the final sections. McNeil et al. [2015] is used as a main reference for most of the chapter.

5.1 Risk Factors and Loss Operator

Modeling the statistical properties of random (future) losses is considered one of the main task of every risk manager, aiming to predict with a given confidence, changes in the value of the portfolio owned during a specific time span.

To begin with, let's define a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_1^T, \mathbb{P})$ where all the possible events live and a random variable $L : \Omega \to \mathbb{R}$ describing a loss. Let t be the current time, V_t denote the value of a portfolio of assets and/or liabilities known at time t and Δt be the considered time horizon. The following assumptions must hold:

(i) no changes in the portfolio composition in Δt;

(ii) no intermediate payments during Δt.

Different time windows could be considered, e.g. 1-day and 10-days are usually mostly used for market related risks while for credit related ones longer intervals like 1-month and 1-year are analyzed.

Even though risk managers are interested in the so called profit-and-loss (P&L) distribution of the portfolio, given by its change in value over the *short time period* Δt, their main concern is with regard to the left tail of the distribution, that is the probability of large losses. Following the exact same approach, from now on we will drop the 'P' from the previous notation and we will focus only on the losses. Thus our basic idea will be to estimate the distribution of

$$L_{t+1} = -\Delta V_{t+1} = -(V_{t+1} - V_t) \,, \tag{5.1}$$

where V_{t+1} is a non-$\mathcal{F}_t$-measurable random variable. For a given t and a d-dimensional vector of risk factor $\boldsymbol{Z}_t = (Z_{t,1}, \cdots, Z_{t,d})$, represented for example by log-prices of stocks, FX rates or inflation, the function $g : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ defines the following mapping

$$V_t = g(\tau_t, \boldsymbol{Z}_t) \quad \text{where} \quad \tau_t = t(\Delta t) \,. \tag{5.2}$$

Let $\boldsymbol{X}_{t-N}, \cdots, \boldsymbol{X}_t$ be a time series of historical risk factor changes where each term is given by $\boldsymbol{X}_{t+1} = \boldsymbol{Z}_{t+1} - \boldsymbol{Z}_t$. Plugging the last expression into (5.1) we get

$$\begin{aligned} L_{t+1} &= -\left(V_{t+1} - V_t\right) \,, \\ &= -\left(g(\tau_{t+1}, \boldsymbol{Z}_{t+1}) - g(\tau_t, \boldsymbol{Z}_t)\right) \,, \\ &= -\left(g(\tau + \Delta t, \boldsymbol{Z}_t + \boldsymbol{X}_{t+1}) - g(\tau_t, \boldsymbol{Z}_t)\right) \,. \end{aligned} \tag{5.3}$$

Recalling that $\boldsymbol{Z}_t$ is $\mathcal{F}_t$-measurable, eq. (5.3) is basically telling us that the loss L_{t+1} is solely driven by the risk factor changes $\boldsymbol{X}_{t+1}$. We can finally define the loss operator at time t as

$$L_{t+1} = l_{[t]}(\boldsymbol{x}) = -g(\tau_t + \Delta t, \boldsymbol{z}_t + \boldsymbol{x}) - g(\tau_t, \boldsymbol{z}_t) \,, \tag{5.4}$$

where $\boldsymbol{z}_t$ is simply a realization of $\boldsymbol{Z}_t$. Depending on the purpose of our estimation, on the available data and especially on the assumptions we are making, two different distributions can be estimated:

- *Conditional loss distribution*: given all the information up to time t, we can look at the conditional distribution

  $$F_{L_{t+1}|\mathcal{F}_t} = \mathbb{P}(l_{[t]}(\boldsymbol{X}_{t+1}) \leq l \,|\mathcal{F}_t) \,.$$

  Usual case in quantitative finance, where we want to include in our risk measurement the most recent information about financial markets.

- *Unconditional distribution*: the stochastic process $(\boldsymbol{X}_s)_{s\leq t}$ is assumed to be stationary and we are interested in the distribution of the loss operator $l_{[t]}(\cdot)$ under $F_{\boldsymbol{X}}$, where $\boldsymbol{X}$ is a d-dimensional random vector with df $F_{\boldsymbol{X}}$. This approach may be appropriate when the time interval considered is longer or we would like to perform stress testing during periods characterized by low volatility.

It is trivial to show that if $(\boldsymbol{X}_i)_{i\in\mathbb{N}_{>0}}$ generate an iid stochastic process then $F_{L_{t+1}|\mathcal{F}_t} = F_{\boldsymbol{X}}$. The usage of conditional and unconditional distribution leads to many differences in the risk management process that are well highlighted and described in McNeil and Frey [2000].

Most advanced concepts of linear and quadratic operator might have been introduced in the last part of this section, but they are completely outside the scope of this master thesis and, if the reader is willing to deepen is knowledge in this direction, the argument is fully analyzed in McNeil et al. [2015]. The next section will be dealing with two of the most important risk measures that are built upon the loss operator concept we have just introduced. Some standard estimation methods are also presented, together with some of their most important (and dangerous) theoretical and practical drawbacks.

5.2 Standard Methods for Risk Measurement

As previously stated, the goal of this section is to estimate the loss distribution

$$L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1}) \,, \tag{5.5}$$

or an approximation thereof. The estimation problem is composed by three main parts:

(a) Statistical problem: estimation of the distribution of $\boldsymbol{X}_{t+1}$;

(b) Numerical problem: derivation of the distribution of L_{t+1};

(c) Evaluation problem: compute a risk measure from $F_{L_{t+1}}$.

Three main approaches could be considered in order to face the task and they are described in the following subsections.

5.2.1 Variance-Covariance method

The Variance-Covariance approach belongs to the class of the analytical methods that implicitly require a closed-form solution to the problem at

hand, without deploying any computationally expensive simulation. In order to guarantee an analytical solution, basic (and critical) assumptions to this approach are:

+ Multivariate normality of the conditional distribution of risk-factor changes, that is $\boldsymbol{X}_{t+1}|\mathcal{F}_t \sim \mathcal{N}_d(\boldsymbol{\mu}_{t+1}, \Sigma_{t+1})$.
+ Sufficiently accurate linear loss operator, $l^{\Delta}_{[t]}(\cdot)$.

Simplicity of the method comes along with some straightforward drawbacks encompassed within the assumptions listed above:

- Multivariate normality assumptions may lead to a severe underestimation of the tail of the loss distribution together with the associated risk measure.
- Linearization of the loss operator may be inaccurate and/or approximation may be complex and computationally expensive to compute.

To take into account some of the extreme risk we can adapt the method using multivariate Student t or multivariate hyperbolic risk-factor changes, but in this case linearization will become even more delicate and crucial.

5.2.2 Historical Simulation

Most of the financial institutions estimate the risk related to their trading book relying on the historical simulation method. Its basic idea is to estimate the distribution of the loss operator under the empirical distribution of historical data

$$\hat{F}_{L_{t+1},n}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\tilde{L}_{t-i+1} \leq x\}} \ , \quad x \in \mathbb{R} \ ,$$

where the process $\tilde{L}_k = L(\boldsymbol{X}_k) = -(g(t+\Delta t, \boldsymbol{Z}_t + \boldsymbol{X}_k) - g(t, \boldsymbol{Z}_t))$ shows what would happen if the past n risk-factor changes were to randomly recur.

+ Easy to implement.
+ No explicit parametric model has to be defined for $\boldsymbol{X}$ and no estimation needed.
- It may be difficult to collect a sufficient amount of data for all the risk factors.
- Tail risk may be underestimated due to the possible lack of *extreme events* in the past time frame considered.

As shown in McNeil et al. [2015], to align our estimation to the tail risk we are actually exposed to, we may deploy techniques from *extreme value theory* in order to estimate the risk of low-probable events, based on the historical losses $L_{t-n+1}, \cdots, L_t$.

5.2.3 Monte Carlo

Monte Carlo method is the most expensive method from a computational point of view, because inference about the loss L is done by simulating new risk factor data. In order to do that, $\boldsymbol{X}_{t+1}$ is assumed to follow some explicit parametric model, without any prior assumption concerning the analytical tractability of the loss operator. To perform a Monte Carlo estimation the simple steps to follow are:

i. Choose and calibrate a specific model accordingly to some given historical time series of risk-factor data, say $\boldsymbol{X}_{t-n+1}, \cdots, \boldsymbol{X}_t$.

ii. Simulate $\boldsymbol{X}_{t+1}^{(1)}, \cdots, \boldsymbol{X}_{t+1}^{(N)}$ for the next time period where each $(\boldsymbol{X}_{t+1}^{(n)})_{n=1}^{N}$ is independent from each other .

iii. Apply the loss operator to the simulated data

$$L_k = l_{[t]}(\boldsymbol{X}_{t+1}^{(k)}) \,, \qquad k \in \{1, \cdots, N\} \,.$$

iv. Infer on the loss distribution F_L and estimate the risk measure of interest. Simple empirical quantile could be considered, but, as in the historical simulation approach, tail risk may be estimated fitting some *extreme value distribution* to the simulated data.

Basically any kind of distribution could be considered for $\boldsymbol{X}_{t+1}$ making *generality* one of the main pro of MC method. On the other side, depending on the size and the complexity of the portfolio the evaluation of the considered loss operator may be difficult and computationally expensive. Lastly but most importantly, this approach does not tackle at all the problem of finding the distribution of $\boldsymbol{X}_{t+1}$.

5.2.4 Measuring risk

Within *Basel* or *Solvency* framework, a risk measure for a financial position with random loss L is basically a real number expressing the amount of capital required to make a position with loss L acceptable to a regulator, internal or external. Determine the risk of an insurance contract, limit the amount of risk of a business unit and define the amount of capital to hold against unpredictable (and extreme) future losses are only some of the reasons why the use of risk measure has become some popular in financial industries. Different approaches might be used in order to achieve this riskiness valuation and they are grouped in

a. *Notional-amount approach:*

$$\text{portfolio risk} = \sum_i (\text{notional value of security}_i) \times (\text{riskiness factor}_i) \,.$$

Simplicity of the method is partially offset by the absence of both netting and differentiation between long and short position.

b. *Scenario-based measure:* given certain possible future risk-factor changes, said $\mathcal{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ with the corresponding weights $(w_i)_1^n$, the risk of a portfolio is,

$$\psi_{\mathcal{X},\boldsymbol{w}} = \max_{1\leq i\leq n}\{w_i \cdot L(\boldsymbol{x}_i)\} \ .$$

c. *Loss-distribution-based measure:* some characteristics of the underlying (estimated) loss distribution are exploited in order to assess the exposure of a given portfolio, e.g. variance, Value at Risk, Expected Shortfall.

From now on, our attention will be on the last approach, while for further details and references on the other methods we invite the reader to refer to McNeil et al. [2015]. Before giving the formal definition of Value at Risk we need to introduce the following concept.

Definition 5.1 (Generalized Inverse) *For any increasing function $T : \mathbb{R} \to \mathbb{R}$, with $T(-\infty) = \lim_{x\downarrow -\infty} T(x)$ and $T(\infty) = \lim_{x\uparrow\infty} T(x)$, the generalized inverse $T^{\leftarrow} : \mathbb{R} \to \bar{\mathbb{R}}$ of T is defined by*

$$T^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : T(x) \geq y\}, \quad y \in \mathbb{R}, \tag{5.6}$$

with the convention that $\inf\emptyset = \infty$. If T is a distribution function, $T^{\leftarrow} : [0,1] \to \bar{\mathbb{R}}$ is the quantile function of T while if T is continuous and $\uparrow$, then $T^{\leftarrow} \equiv T^{-1}$.

Hence, we can now finally give the following

Definition 5.2 (Value-at-Risk) *The Value-at-Risk of a portfolio at a given confidence level $\alpha \in (0,1)$, is given by the smallest number x such that the probability of the loss L exceeds x is no larger than $(1-\alpha)$. In mathematical notation that means*

$$VaR_\alpha(L) = F_L^{\leftarrow}(\alpha) = inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\}.$$

VaR is basically the α-quantile of the distribution function F_L and its simplicity is one of the main reasons for its widely usage in both financial and non-financial institutions. The concept has been introduced by JP Morgan in the early 80s by its former CEO Dennis Weatherstone. The framework became known worlwide only few years later, precisely in 1994, when it has been published by the bank, through the famous CreditMetrics Technical Documentation, while recent academic developments in this direction are given for example in Mina, Xiao, et al. [2001]. Being a frequency-based measure, unfortunately Value-at-Risk is not assessing and/or providing any information on the size of the loss that may occur with a probability $\leq 1-\alpha$.

To overcome this issue, other measures, introduced soon, must be deployed. Before even starting with the estimation process, two main parameters has to be chosen: Δt and α. Depending on the risk that is being measured and accordingly to the specific regulatory framework, we can set those two values as:

1. Basel II
   a) Market Risk: $\alpha = 0.99$ and $\Delta t = 10d$.
   b) Credit Risk: $\alpha = 0.999$ and $\Delta t = 1yr$.
2. Solvency II
   a) $\alpha = 0.995$ and $\Delta t = 1yr$.

Before moving to the next section, it is worthwhile to introduce another extremely important risk measure able to overcome some of the theoretical deficiencies that come together with the Value-at-Risk definition. Those drawbacks will be shortly presented in section 5.3.

Definition 5.3 (Expected Shortfall) *Let L be a loss with distribution function F_L and $\mathbb{E}(|L|) < \infty$. The Expected Shortfall of a portfolio with loss L at confidence level $\alpha \in (0,1)$ is the average VaR over all $u \geq \alpha$. Formally*

$$\mathbb{ES}_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 F_L^{\leftarrow}(u)du = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(L)du. \tag{5.7}$$

In comparison to Value-at-Risk, Expected Shortfall looks even further into the extremes of the loss distribution investigating also in the $(1-\alpha)$ probability events "hidden" in the tails.

We can now finally move on to the next sections of the chapter, where some of the major properties of the risk measures are presented together with some backtesting techniques useful (and crucial) to evaluate the quality of our estimates.

5.3 Coherence and Convexity

The idea of this third section, it is to introduce and familiarize with some important properties that a risk measure has to satisfy in order to be considered a good measure. The following axioms should be considered as practical tools that may be used to assess, compare and rank all different risks faced by a financial and non-financial institution.
The first four postulates, were initially introduced by Delbaen [2002], extended on infinite probability spaces some years later by Delbaen [2002] and finally defined in a more general (*convex*) framework by Föllmer and Schied

[2002]. Let $\mathcal{L}^0(\Omega, \mathcal{F}, \mathbb{P})$ be the set of all random variables that are almost surely finite on $(\Omega, \mathcal{F})$ and let $\mathcal{M} \subseteq \mathcal{L}^0(\Omega, \mathcal{F}, \mathbb{P})$ be a linear space representing portfolio losses over a fixed time horizon Δ. The set $\mathcal{M}$ is often assumed to be a *convex cone*, that is

Definition 5.4 (Convex Cone) *A given set $C \subseteq \mathcal{M}$ is said to be a convex cone if $(1-\gamma)x + \gamma y \in C$ and $\lambda x \in C$ for all $x, y \in C$, $\gamma \in (0,1)$ and $\lambda > 0$.*

In this framework, a *risk measure* is defined as a real-valued map defined on the (convex) cone $\mathcal{M}$ as

$$\varrho : \mathcal{M} \to \mathbb{R} ,$$

where for all $L \in \mathcal{M}$, $\varrho(L)$ represents from an economical point of view, the amount of capital needed to make a position with loss L, acceptable by internal/external risk controllers. Additionally this measure can be either positive or negative depending if the risk-free capital must be increased or can be eventually reduced.

Before introducing the axioms, let's now consider at time zero a simple portfolio composed by two basic instruments: a risk-free asset of amount x and a risky one with the terminal value L. Having no risk associated with it, at time 1, x is mapped to $x \cdot (1 + r_f)$ where r_f represent the risk-free rate. Thus,

Axiom 1 (Translation invariance)

$$\forall L \in \mathcal{M},\ x \in \mathbb{R}, \qquad \varrho(L + x \cdot (1 + r_f)) = \varrho(L) - x \, . \tag{5.8}$$

Reasonably, capital requirements are decreased investing a real value x in the risk-free asset.

Axiom 2 (Sub-additivity)

$$\forall (L_1, L_2) \in \mathcal{M} \times \mathcal{M}, \qquad \varrho(L_1 + L_2) \leq \varrho(L_1) + \varrho(L_2) \, . \tag{5.9}$$

VaR is in general non sub-additive and using a non-additive measure regulatory capital requirements could be easily lowered by any institution splitting up in subsidiaries. The main idea is that in order to minimize the overall risk associated with a specific portfolio, a portfolio manager is encouraged to apply diversification.

Axiom 3 (Positive Homogeneity)

$$\forall L \in \mathcal{M},\ \lambda \geq 0, \qquad \varrho(\lambda \cdot L) = \lambda \cdot \varrho(L) \, . \tag{5.10}$$

The main criticism that has been done to this axiom, it is that for large positive λ liquidity risk comes into play and concentration of risk should be penalized. Considerations which lead us to consider $\varrho(\lambda L) > \lambda \varrho(L)$, with the immediate consequence of contradicting the previous axiom.

Axiom 4 (Monotonicity)

$$\forall (L_1, L_2) \in \mathcal{M} \times \mathcal{M} \text{ such that } L_1 \leq L_2 \text{ (a.s)}\,,$$
$$\varrho(L_1) \leq \varrho(L_2)\,. \tag{5.11}$$

That means, positions that almost surely lead to higher losses require more capital.

Now, we can finally give the following

Definition 5.5 (Coherent risk measure) *A risk measure ϱ on the convex cone $\mathcal{M}$ is coherent if it satisfies axioms 1-4.*

As already mentioned previously in the section, some of the requirements that need to be satisfied by a risk measure in order to be defined coherent, may be relaxed leading to a more general group of measures. Precisely conditions expressed in axioms 2 and 3 may be softened and substituted by the weaker property of convexity, defined as

Definition 5.6 (Convex risk measure) *A risk measure $\varrho : \mathcal{M} \to \mathbb{R}$ is said to be convex if it is monotone, translation invariant and for all $(L_1, L_2) \in \mathcal{M} \times \mathcal{M}$ and $\lambda \in [0, 1]$ satisfies*

$$\varrho(\lambda \cdot L_1 + (1 - \lambda) \cdot L_2) \leq \lambda \cdot \varrho(L_1) + (1 - \lambda) \cdot \varrho(L_2)\,. \tag{5.12}$$

Being the two properties one the generalization of the others, coherence imply convexity, but in general the converse does not hold true. Additionally, as previously stated VaR is not a coherent risk measure (at least in the general case, e.g. with heavy-tailed distributions) whereas Expected Shortfall is. Detailed proofs and counter-examples, can be found in McNeil et al. [2015].

5.4 Consistent Measures of Risks

Different properties may be analyzed and defined for a given risk measure and before moving to the final section dealing with backtesting, here we would like to introduce a different approach that might be a convenient complement to coherency and convexity.
As introduced and discussed in Malevergne and Sornette [2002] and Malevergne and Sornette [2006] let's recall the same simplified economic framework previously defined, with the difference that the newly defined risk measure $\bar{\varrho}$ is forced by the following axiom to take values only in $\mathbb{R}_{\geq 0}$

Axiom 5 (Positivity)

$$\forall L \in \mathcal{M} \qquad \bar{\varrho}(L) \geq 0. \tag{5.13}$$

In addition we would like $\bar{\varrho}$ to be *translational invariant*, that is, to guarantee that the fluctuation do not change if any deterministic amount is added to the random payout of the risky position. Formally

Axiom 6 (Translational Invariance)

$$\forall L \in \mathcal{M},\ \forall x \in \mathbb{R} \qquad \bar{\varrho}(L + x)\bar{\varrho}(L). \tag{5.14}$$

In order to overcome the liquidity risk that, as previously said, is not properly tackled by Axiom 3, our risk measure is forced to satisfy the following

Axiom 7 (Positive Homogeneity)

$$\forall L \in \mathcal{M},\ \forall \lambda \in \mathbb{R}_+,\ \exists \xi \geq 1 \qquad \bar{\varrho}(\lambda \cdot L) = \lambda^{\xi} \cdot \bar{\varrho}(L). \tag{5.15}$$

where the liquid market case could be easily retrieved forcing $\xi = 1$. Various practical examples of risk measures obeying to axioms 5-7 are analyzed in Malevergne and Sornette [2006], in which great emphasis has been put on centered moments (variance in particular) and cumulants. Let's now move on to the last section of the chapter where a step-by-step backtest guide will be presented.

5.5 Backtesting

Whenever a risk measure is estimated by a model, its quality needs to be evaluated with appropriate backtesting methods. In order to do that we basically need to define a statistical procedure aiming to assess the validity of both our model and estimation. The idea is then to compare the predicted losses against the actual ones realized at the end of the considered time window. Precisely, we say that an exception has occurred and VaR has been underestimated, if by that time, our portfolio has experienced a loss greater than the estimated VaR. In this direction two types of test may be performed to analyze those kind of events:

- *Unconditional Coverage:* it determines whether the frequency in the exceptions during Δt is in line with the confidence interval that has been considered. It does not take into account the time of the exception.
- *Conditional Coverage:* it captures if the different exceptions occurred independently from each other. It indicates if the model is able to detect changes in market dynamics like volatility or correlation.

Given the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$ let VaR_α^t be the Value-at-Risk at a specific confidence level α. Let's now consider $I_{t+1} = \mathbb{1}_{\{L_{t+1} > \text{VaR}_\alpha^t\}}$ counting 1 if the violation has happened and 0 if it hasn't. Assuming continuous

loss distribution and exploiting the definition of quantile, we have that I_{t+1} is a Bernoulli random variable with probability $(1-\alpha)$, whose sum is

$$M = \sum_{t=1}^{m} I_t \ \sim \text{Bin}(m, 1-\alpha).$$

If we then assume the timings of the exceptions to be $1 \leq T_1 < \cdots < T_M \leq m$ with $T_0 = 0$, then $S_k = T_k - T_{k-1}$ with $k = (1, \cdots, M)$ will be Geometrically distributed with probability mass function given by

$$\mathbb{P}(S_k = j) = \alpha^{k-1}(1-\alpha), \qquad j \in \mathbb{N}.$$

Both concepts can easily be tested on empirical data, computing for example Likelihood Ratios defined in Christoffersen [1998] and Kupiec [1995] for spacings of violations and for their Binomial behaviour, respectively. We performed both 95% and 99% confidence estimates leading us to reject the null hypothesis whenever Likelihood Ratio was greater then $\chi^2_1(5\%) = 3.841$ and $\chi^2_1(1\%) = 6.635$. In order to consider a mixture of the previous methods the two ratios can be also combined in a single one. The sum has to be then compared to a chi-square distribution with two degrees of freedom. Critical values become $\chi^2_2(5\%) = 5.991$ and $\chi^2_2(1\%) = 9.210$.

Chapter 6

Simulations and Results

This chapter is divided in three main parts and it has been structured with the idea that each one of them gives a detailed summary of the results for a specific purpose and method that have been used. Section 6.1 is dedicated to model simulation and it will present how a synthetic time series has been able to capture the most important stylized facts observed in real financial data. First results are then introduced in section 6.2 where the code developed in Berntsen [2015] is shown to be slightly improved in performance deploying some Object-Oriented features, while SPSA online estimation presented in chapter 3, is used to estimate the unknown parameters. Even though online calibration has produced interesting results and estimation was quite satisfactory from a statistical point of view, especially for some parameters, convergence could not be guaranteed for all of them and maximum and profile likelihood are then adopted as an alternative calibration methodology.

6.1 Model simulations

In order to investigate the properties of the model developed by Malevergne and Sornette [2014] a first analysis has been done by considering 40-year synthetic time series produced with a set of parameters that has been chosen in order to simulate what is observed empirically in financial returns. In the considered framework, discrete observations are made available on a daily basis while each year consists of 250 trading days. Further remarks should be made in order to clarify and understand the values that have been chosen to run the simulation. The first four parameters related to the GARCH(1,1) dynamics define realized volatility $\bar{\sigma}$ and return $\bar{r}$ respectively equal to 25% and 7% on an annual basis. Paramters α and β are instead connected to the dependency of the process on past squared returns and volatility, with the values fixed at 0.05 and 0.94. Concerning the jump size the value κ

has been set to $1/25 = 0.04$ expecting to observe average jumps of around 4%. Parameter a has been selected in order to consider a memory of the mispricing equal to 250 days, that is $a = 1 - 1/250 = 0.996$. Finally, $\bar{X}$ and η are fixed to values equal to -5 and 3, respectively. It is worthwhile to remark, that given eq. (2.26), volatility of the mispricing index X_t, is then expected to be 30-35 times bigger than the volatility of the returns r_t, whilst its mean to be equal to -5.

With respect to $\bar{X}$ and η suggestions of Malevergne and Sornette [2014] have been followed to select their most significant values. Finally the memory of the model is then uniquely described by the single parameter a that with the value 0.996, is of the order of 250 trading days. To summarize

$\bar{r}$	$\bar{\sigma}$	α	β	κ	$\bar{X}$	η	a
0.07	0.25	0.05	0.94	0.04	-5.0	3.0	0.996

The above parameter combination produced a time series in line with well known stylized facts related to empirical log-returns. To begin with, the upper plot in figure 6.2 shows the ability of the model to capture volatility clustering while the bottom one illustrates log-returns and squared log-returns together with their sample autocorrelation functions (ACFs). The confidence intervals have been estimated using asymptotic theory exploiting, in particular the well known result, that for long iid time series of length n, estimated autocorrelation, say $\hat{\rho}$, is normally distributed with mean zero and variance inversely proportional to n. Thus the 95% confidence interval is simply given by $\hat{\rho} \pm 1.96/\sqrt{n}$. Lastly, as an additional check we also measured skewness and kurtosis of the simulated log-returns, respectively defined as the third and fourth standardized moments, formally:

$$skew(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right], \qquad kurt(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]. \tag{6.1}$$

Negative skewness (-0.3171) and kurtosis bigger than 3 (7.2141), implied for our synthetic time series, smaller mean than the median and leptokurtic distribution characterized by more concentration around the mean and slower decay in the tails, in comparison to a normal distribution. All characteristics observable in empirical financial data.

Dependency between log-prices, their explosive growth and crash probability are easily observable in figure 6.1. Reading together the top and the bottom plot we can spot how the super-exponential growth in the log-price is interconnected with the mispricing index and on a second instance with the crash-hazard rate. Basically 4 different regimes may be observed in the time

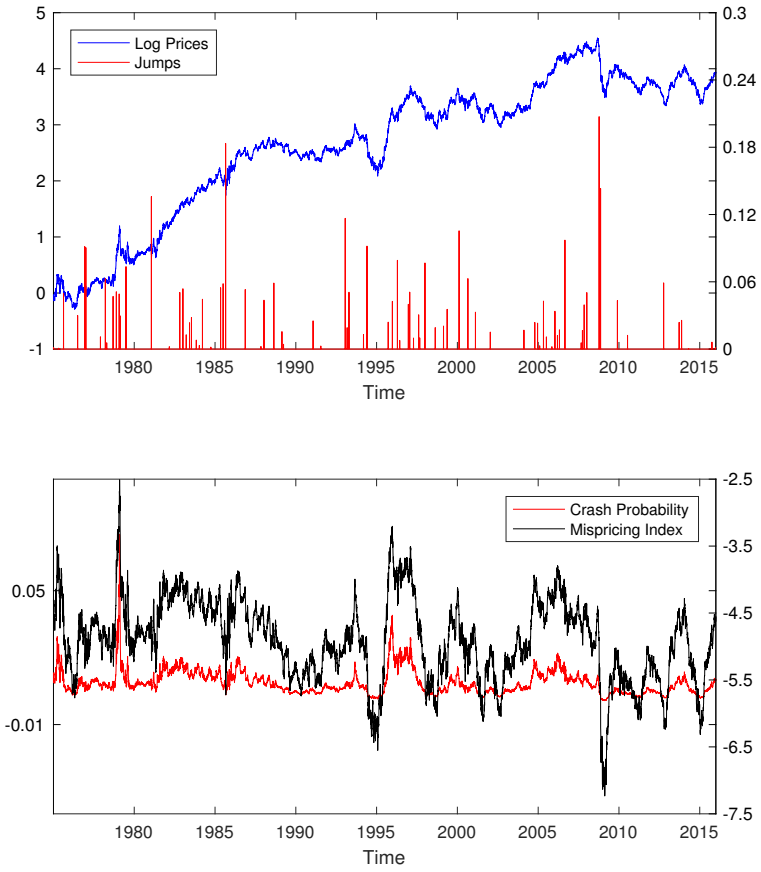

Figure 6.1: Top: simulated log-prices (blue line, left scale) and jumps (red line, right scale). Bottom: mispricing index (black line, right scale) and crash probability (red line, left scale) corresponding to the simulated log-price dynamic showed in the top plot.

frame considered. A pretty flat behaviour is shown up to around 1978-79 followed by an unusual boost in the log-price with an obvious consequent spike in the crash probability justified by the increased risk the investors have to bear in order to keep the position open. Just before 1980 as soon as multiple crashes brought the log-price closer to its fundamental value another change of regime may be easily spotted. An upward trend is indeed observable in the log-price up to 1988-1989 where a flat period drove both the mispricing index and the crash-hazard rate to low and reasonable levels. Frequency of jumps spiked again especially between 1995 and 2000 where crash probability rose to almost record levels, but also right before the his-

torical maximum hit by log-prices in 2008-2009. Finally, some jump activity is observed in the last analyzed decade, but overall the trend remains flat.

As pointed out in chapter 2 the revolutionary and enlightening elements of Malevergne and Sornette [2014] model are a non-local and self-referencing mispricing index X_t and a crash-hazard rate λ_t, joined to the latter through a non-linear and S-shaped standard logistic function. X_t has been defined as a process following a normal distribution with mean given by the parameter $\bar{X}$ and standard deviation proportional to the volatility of the simulated log-returns. The qq-plot in the bottom right corner of figure 6.3 is basically verifying this behaviour. Quantiles of the mispricing index are indeed plotted against the theoretical quantiles values from a normal distribution. A linear shaped line proves the analogy of the distributions. On its left side the scatter plot is showing instead the high correlation between the crash-hazard rate at time t and $t+1$ displaying higher variation in the top-right corner. Those points relate to events belonging to the time frame when the crash probability reached its local peak in February 1979 before a subsequent drop in the months ahead. Finally in the top plot in figure 6.3 gaussian kernel smoothing function is used to estimate the density of the crash-hazard rate clearly matching its theoretical counterpart defined by the logistic-normal distribution introduced in eq (2.25).

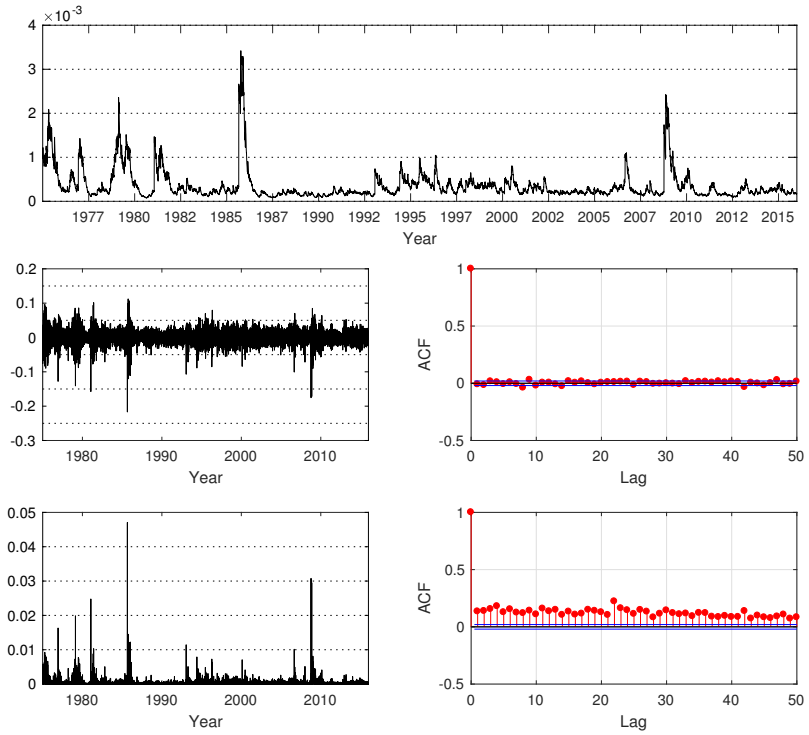

Figure 6.2: Top: trajectory of the volatility of the simulated jump-diffusion model. Bottom plots: log-returns and squared log-returns with the corresponding autocorrelation functions up to 50 legs.

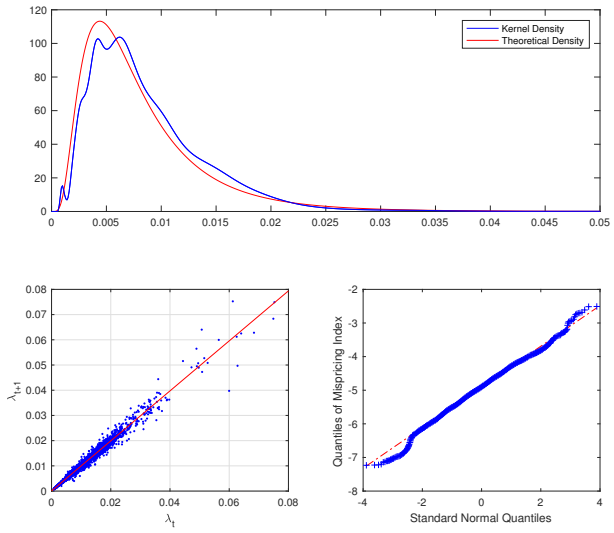

Figure 6.3: Top: normal kernel smoothing function (blue) and theoretical density (red) of the crash-hazard rate λ_t of the simulated time series. Bottom left: scatter plot of crash probability at time t vs $t+1$ with corresponding least-squared fitted line (red). Bottom right: quantiles of the mispricing index X_t (blue crosses) versus the theoretical quantiles from a normal distribution (red dashed line).

6.2 Particle Filter and Online Estimation

We decided to divide this section in two main parts both extremely tied with the work that has been done in Berntsen [2015] which will be used as a proxy for some of our results and as a starting point for many considerations. The simulated time series that has been analysed here is still the same one displayed in figure 6.1 produced with the set of known parameters given in section 6.1.

6.2.1 Code Optimization

Taking over the work that has been done in Berntsen [2015], my first task has been to increase portability and performance of the code. Three versions already implemented in Berntsen [2015], namely Sequential Importance Resampling (SIR), Auxiliary Particle Filter (APF) and SIR with State Augmentation (SIRSA), were grouped into one single code. A `TimeSeries class` was created with such a purpose and, depending on the input, SIR, APF or SIRSA were performed with a specific number of particles and time series length. As shown by the black dashed line in the top plot of figure 6.4, CPU running time was on average, decreased by 18.2% compared to the procedure developed in Berntsen [2015]. The code has been run on the same synthetic 10000-step time series of log-prices introduced above and showed in Figure 6.1.

Exceeding our expectations, the new algorithm has been also able to better detect the timings of jump arrivals. This result can be seen in the second plot from the top of Figure 6.4. Blue bars, corresponding to the new algorithm, in most of the cases exhibit greater height than the red ones, representing the older version of the code. In that specific plot a bar going all the way up to 1 represents the ability of the algorithm to fully capture the actual occurrence of a jump.

Overall all the 83 jumps have been identified by the filter, but a low confidence has been assigned to many of them. Here, confidence is expressed in probabilistic terms given by the heights of the bars in the second plot of Figure 6.4. Bottom plots of the same Figure, analyze how well the algorithm behaved with respect to false and true positive, with respect to changes in this confidence level. We can indeed consider our filter as a binary classifier giving us the output *jump* or *no jump* at a given time t. If the jump is detected, but no jump actually occurred, we have a *false positive*. By logic, a *true positive* is when the jump is both captured and observed. As shown by the bottom left plot of Figure 6.4, many false positive results are produced by the new code when threshold is set to a level below 0.5. With respect to true positive instead, the new code is always preferred to the old one.

6.2.2 SPSA Results

Here the assumption of known parameters is relaxed and the estimation results obtained with the online procedure described in chapter 3 are presented. In the first place it is worthwhile to specify once more the 8-dimensional space where the parameters live.

$$(\bar{r}, \kappa, \bar{X}, \eta) \in \mathbb{R}^4, \tag{6.2}$$

$$\bar{\sigma} \in \mathbb{R}_+, \tag{6.3}$$

$$(\alpha, \beta) \in \{(x, y) \in [0,1)^2, x + y < 1\}, \tag{6.4}$$

$$a \in [0,1]. \tag{6.5}$$

The constraints specified above are crucial for the calibration and they must be satisfied throughout the whole estimation process in order to guarantee the existence of the dynamics defined by the model. In order to do that, Spall [1998] simply suggested that given a parameter $\theta \in [\theta_{min}, \theta_{max}]$, the two lines below may be added to SPSA algorithm right after the update step

$$\theta = \min(\theta, \theta_{max}) \tag{6.6}$$

$$\theta = \max(\theta, \theta_{min}). \tag{6.7}$$

In our case though this approach was proven to be inefficient due to the fact that all the particles tend to fall into this imposed boundary condition and remain there until the end of the estimation procedure. We then decided to proceed differently. In our second implementation, every time a particle was perturbed or a parameter was updated, if any of the condition was not met, the corresponding perturbation or parameter update was recomputed until all the constraints were satisfied or a maximum number of iterations was reached. This methodology together with gain sequences specifically defined to accommodate the different changes of magnitude of each parameter from one step to the other, lead us to the results of Figure 6.5.

Constraints within SPSA algorithm

```
1: while All constraints not satisfied do
2:     while All constraints not satisfied do
3:         Generate a 8-dimensional simultaneous perturbation vector
4:     Parameter Update
```

As in every optimization procedure, the starting values are crucial in order to perform a successful and meaningful calibration and achieve any reasonable outcome. Precisely, starting points for κ, η and a have been drawn

independently for each particle, from a uniform distribution centered in the corresponding true value. The same technique has been adopted with the couple (α, β), subject to $\alpha + \beta < 1$. $\bar{r}$ and $\bar{\sigma}$ instead have been drawn respectively from a Normal and a Chi-Squared distribution. $\bar{r}$ describes the unconditional expected return of the risky asset and drawing its starting values from a normal distribution with mean and standard deviation equal to the sample ones, was a straightforward choice. On the other hand concerning $\bar{\sigma}$, that is describing the long-term volatility of the GARCH(1,1) process, we know that in the case of a Gaussian sample (log-returns in the present case) its variance follows a χ^2_k distribution, where the degrees of freedom (equal to $T-1$ in our case) have to be selected in order to control the spread of the distribution. Using also the following relationship between chi-squared and Gamma distributions

$$\chi^2_n \sim \Gamma\left(\frac{n}{2}, 2\right) \quad \text{and} \quad a \cdot \chi^2_n \sim \Gamma\left(\frac{n}{2}, 2a\right) \quad \forall a \in \mathbb{R}_{>0} \tag{6.8}$$

we have drawn our starting values from $\Gamma(\bar{a}, \bar{b})$ where $\bar{a} = \frac{T-1}{2}$, $\bar{b} = \sigma_s^2 \cdot \frac{2}{T-1}$ where T and σ_s^2 define the length of the time series and the sample variance, respectively.

Although Spall [1998] and Mina et al. [2001] showed convergence of the method in similar highly non-linear problems, in our specific case no convergence is guaranteed. Even reducing the uncertainty moving from 8 unknown parameters to 7 (e.g. assuming parameter a to be given), no improvement in the calibration is shown. Against our expectation Figure 6.5 shows indeed that even after 10000 time steps, no reduction is observed in the volatility of the estimated parameters. No improvement was obtained increasing or decreasing the number of particles.

In Berntsen [2015] the parameter a was shown to be the most difficult to calibrate, while our results proved something different. In our framework indeed, parameter a, together with $\bar{r}, \bar{\sigma}, \alpha, \beta$ was either fluctuating around the corresponding true value or moving towards it. Different behaviour instead was observed for κ, but especially for $\bar{X}$ and η, whose calibrated values were far away from the their true counterparts. During the estimation process values of $\bar{X}$ and η spanned from -5.5 to -4.5 and from -20 to +17, respectively. These results showed once more the sloppy nature of our problem as anticipated in both Malevergne and Sornette [2014] and Berntsen [2015]. The interesting result was that the problematic parameters have been identified to be $\bar{X}$ and η. In the next section, eigenvalues and eigenvectors of the Hessian of the cost function are analyzed in order to further investigate our findings.

To conclude, our dual estimation procedure has proved inefficient for the model at hand. The reason stems from the fact that particle filter and SPSA

algorithm we implemented failed to disentangle the dependencies between hyper-parameters and hidden states, leading to highly unstable results.

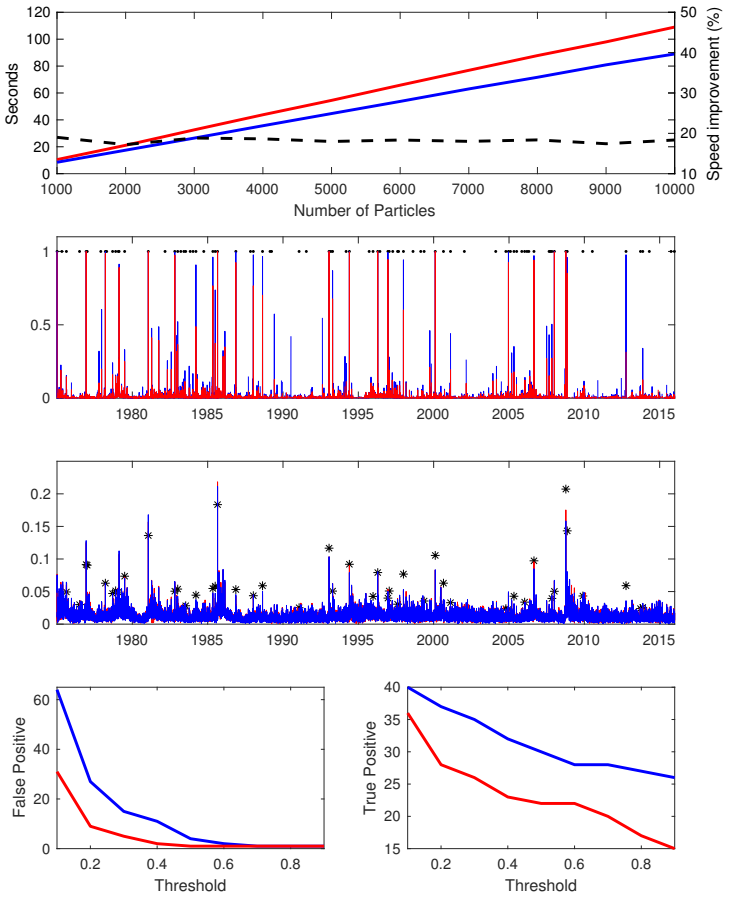

Figure 6.4: Red and blue lines/bars represent the old and the new code respectively. Top: performance improvement with Object Oriented features analysing 10000-step synthetic time series with different number of particles and run on a MacBook Pro 2,26 GHz Intel Core 2 Duo with OS X El Capitan. Dashed line refers to the right y-axis and describe the percentage speed improvement. Central plots: jump timings (upper plot) and size (lower plot) filtered using SIR particle filter with number of particles and time steps both equal to 10000. Black dots and stars respectively represent timings and sizes of observed jumps. Blue and red bars indicate how well the filer was able to capture them. In the second plot from the top, bars represent the confidence in the estimates. Full timing detection is represented by bars with height equal to 1. Bottom plot: number of false and true positives with respect to changes in the threshold of confidence.

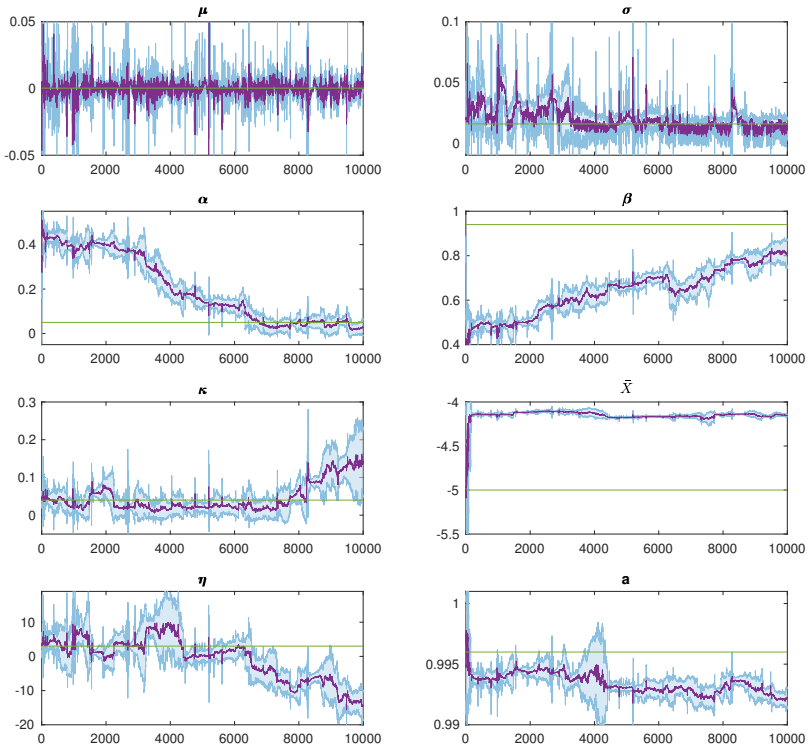

Figure 6.5: Parameters estimated online using SPSA algorithm within SIR particle filter. Purple lines correspond to the mean of the parameter values assumed by all the particles at each time t while shaded areas represent confidence intervals at 95%. 1000 particles and 10000 time steps are considered. True value represented by the green lines.

6.3 Maximum and Profile Likelihood

In this section, a modified and tailored version of the standard maximum likelihood method is presented together with its results. Given the needs to better understand the importance of the different parameters with respect to the cost function, profile likelihood has been also implemented. Finally due to the statistically insignificant results we obtained while profiling the likelihood, its modified version introduced in chapter 4 has not been implemented.

6.3.1 Maximum Likelihood

Here we show the results of the likelihood maximization with respect to our 8-dimensional parameter vector $\theta = (\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a)$. As is often the case with optimization problems, we decided to exploit the relationship $\arg\max_\theta f(\theta) = \arg\min_\theta -f(\theta)$, since most of the optimization libraries provide only minimization routines. For our purpose `DLIB C++` Library is employed. Namely, the unconstrained minimization function `find_min` is adopted.

Due to the complexity of the problem we divide the optimization task in two parts. In the first place GARCH(1,1) estimation is assessed, meaning that $(\bar{r}, \bar{\sigma}, \alpha, \beta)$ are calibrated using the Limited-memory BFGS quasi-Newton search strategy. Wurtz, Chalabi, and Luksan [2006] showed the efficiency of the algorithm in a similar framework. The remaining parameters, $(\kappa, \bar{X}, \eta, a)$ are instead tackled in a second instance where `newton_search` strategy is used. This method gives the possibility to provide analytical gradient and Hessian, both crucial for algorithm convergence. Those information are indeed necessary in order to guarantee high computational precision and avoid being trapped in local maxima. Sadly, `DLIB` does not support any `float_128` or `long double` data type and we are forced to use basic `double` precision.

To sequentially optimize the problem, but also to speed up the computation function-objects are also implemented following `DLIB` documentation. Boundary conditions of the model are manually forced to be satisfied in this context simply assigning an unrealistic high value to the log-likelihood whenever one of the constraints is violated. We take a similar approach with regards to the gradient. Its closed form solution may not be defined outside the boundaries and then its numerical approximation is computed by `dlib::derivative` function.

Lastly, as a final step of the optimization, a gradient based line-search is performed. Different directions are given here by the eigenvectors of the Hessian. The idea of this last step is basically trying to minimize the norm

of the gradient of the cost function, unless it is not sufficiently small already. Sufficiency in this case is determined by the following quantity

$$\varepsilon = \nabla\mathcal{L}(\theta^*)^T \cdot \Lambda_i \tag{6.9}$$

where $\nabla\mathcal{L}(\theta^*)$ is the gradient of the log-likelihood evaluated in our optimal value, while Λ_i represents the eigenvector corresponding to the i^{th}-eigenvalue. In our specific case the threshold for ε is set to 1e-4.

Results

Here the aforementioned procedure is applied on the 40-year synthetic time series presented in section 6.1 and simulated with the following set of parameters: $\bar{r} = 0.07$, $\bar{\sigma} = 0.25$, $\alpha = 0.05$, $\beta = 0.94$, $\kappa = 0.04$, $\bar{X}$-5.0, $\eta = 3.0$ and $a = 0.996$. The estimated values together with their corresponding robust standard errors are displayed below. Calibration procedure took 430 sec on a MacBook Pro 2,26 GHz Intel Core 2 Duo with OS X El Capitan. With the minimum eigenvalue equal to 5.35444e-05 the Hessian of the log-likelihood is shown to be positive definite. This result confirms that the optimal parameter vector is an actual miminum and not a saddle point in which the procedure has been trapped. Overall the method performed well and lead to estimated values close to their true counterpart with small robust standard errors. Additionally, in contrast with Berntsen [2015] conclusions, parameter a is resulted to be calibrated correctly. On the contrary though, optimal $\bar{X}$ and η resulted in having unexpectedly high standard errors, forcing us to further investigations in this direction.

$$-\mathcal{L}(\hat{\theta}) = \text{-2.5950}$$

$$\|\nabla\mathcal{L}(\hat{\theta})\| = \text{5.2638e-04}$$

$\bar{r}$ = 0.0880 (0.0002)	κ = 0.0627 (0.0119)
$\bar{\sigma}$ = 0.2386 (0.0006)	$\bar{X}$ = -5.1276 (0.2776)
α = 0.0480 (0.0036)	η = 2.6120 (1.3642)
β = 0.9412 (0.0040)	a = 0.9936 (0.0045)

6.3.2 Profile Likelihood

In this second part of the section we aim to better understand the interconnections between the different parameters. Especially we would like to understand how important each one of them is with respect to the calibration process. We already know that the log-likelihood as a function of some

of the parameters is very flat and it is with respect to those "sloppy" parameters that profile likelihood is performed. The basic idea here is to exploit some information contained within the Hessian matrix of the log-likelihood. In order to do that, eigenvalues and eigenvectors are computed at the end of the full estimation procedure described before. Thus

$$\begin{array}{llll} \lambda_1 = 3484.92 & \lambda_2 = 296.64 & \lambda_3 = 75.9838 & \lambda_4 = 7.5176 \\ \lambda_5 = 3.6071 & \lambda_6 = 1.3313 & \lambda_7 = 0.0014 & \lambda_8 = 5.4\text{e-}05 \end{array}$$

$$\Lambda_1 = \begin{bmatrix} 0.9999 \\ -0.0107 \\ -0.0014 \\ -0.0006 \\ 0.0020 \\ -1.2\text{e-}05 \\ -6.4\text{e-}06 \\ -0.0003 \end{bmatrix} \quad \Lambda_2 = \begin{bmatrix} -0.0103 \\ -0.9873 \\ 0.1427 \\ 0.0685 \\ 0.0008 \\ -0.0004 \\ -6.6\text{e-}06 \\ 0.0026 \end{bmatrix} \quad \Lambda_3 = \begin{bmatrix} -0.0031 \\ -0.1526 \\ -0.7431 \\ -0.6515 \\ -0.0012 \\ -0.0005 \\ -3.0\text{e-}05 \\ -0.0098 \end{bmatrix} \quad \Lambda_4 = \begin{bmatrix} 0.0001 \\ 0.0004 \\ -0.0185 \\ 0.0059 \\ 0.0650 \\ 0.0032 \\ 0.0021 \\ 0.9977 \end{bmatrix}$$

$$\Lambda_5 = \begin{bmatrix} 0.0009 \\ 0.0421 \\ 0.6535 \\ -0.7555 \\ -0.0055 \\ 0.0003 \\ 0.0001 \\ 0.0169 \end{bmatrix} \quad \Lambda_6 = \begin{bmatrix} 0.0021 \\ -0.0008 \\ -0.0038 \\ 0.0053 \\ 0.9974 \\ -0.0308 \\ -0.0017 \\ 0.0650 \end{bmatrix} \quad \Lambda_7 = \begin{bmatrix} 0.0001 \\ -0.0005 \\ -0.0006 \\ 0.0001 \\ -0.0310 \\ 0.9978 \\ 0.0587 \\ -0.0013 \end{bmatrix} \quad \Lambda_8 = \begin{bmatrix} 5.2\text{e-}06 \\ 1.1\text{e-}05 \\ 4.8\text{e-}06 \\ 1.6\text{e-}05 \\ -2.2\text{e-}06 \\ -0.0587 \\ 0.9983 \\ -0.0019 \end{bmatrix}$$

The first thing to notice is the difference in the order of magnitudes between λ_1 and λ_8. Additionally, analyzing the corresponding eigenvectors, namely Λ_1 and Λ_8 we notice that they are dominated by the first ($\bar{r}$) and the 7^{th} parameter (η), respectively. This is telling us that the likelihood of the model is extremely sensitive to changes in the value assigned to $\bar{r}$, but on the contrary, it will not be affected by any variation made to the parameter η. Same reasoning holds true with respect to the parameter associated with λ_7, i.e. $\bar{X}$. By fixing a threshold ε to the value of λ_i we can now distinguish between *sloppy* and *rigid* parameters. Figure 6.6 shows the results obtained profiling the cost function following the guidelines given in Chapter 4.2.

Figure 6.6 clearly displays an improvement in the estimate of the calibrated η before (black dotted line) and after (green dotted line) the profile maximization. Opposite behaviour is instead observed for the parameter $\bar{X}$. The problem with this approach can be seen analysing the relative likelihoods,

$R_p(\theta|\eta)$ displayed in the second and in the fourth plot. Indeed no cutoff point can be set in neither of the cases and no probabilistic range can be defined in the parameter space. All the value we spanned lead to a value of the relative likelihood way above the 95% threshold is usually set. To conclude, uncertainty of the point estimates remains too high and against our expectations profile likelihood method is proven to be inefficient. A first attempt was made to implement its modified version described in chapter 3, but due to time constraints it was not successfully completed. It is believed that taking into account additional information contained in the Hessian of the log-likelihood, the method may be helpful for the our specific case. Further investigation in this direction may be exploited by future researchers.

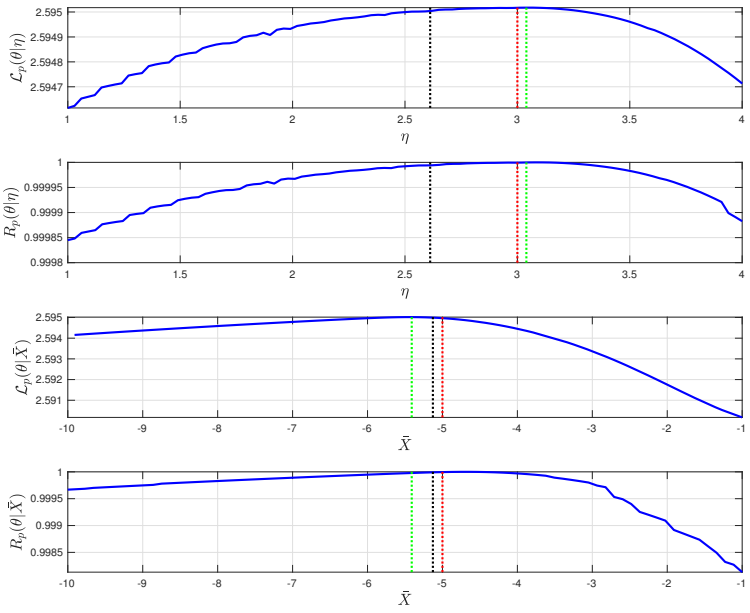

Figure 6.6: Profile ($\mathcal{L}_p$) and relative profile (R_p) likelihood estimates with respect to η and $\bar{X}$. Red dotted line: true values of the parameters. Black dotted line: estimated values of the parameters after the first full 8-parameter calibration, but before the profile optimization. Green dotted line: values of the parameter maximizing the profile likelihood.

6.3.3 Monte Carlo Estimation

Monte Carlo estimation has been performed in order to better understand statistical consistency of the calibration process introduced before. Knowing that a critical part of the estimation is due to the interconnection and rela-

tionship between the different parameters any kind of convergence has to be proven many combination. Theoretically an 8 dimensional grid should have been produced to exploit as many combinations as possible, but computational time and power are two constraints that must be always taken into account. Given our parameter vector $\theta = (\bar{r}, \bar{\sigma}, \alpha, \beta, \kappa, \bar{X}, \eta, a)$, we decided to freeze the first four component related to the GARCH(1,1) whose estimation has been shown to be less complicated while the other four were selected within the following 4-dimensional grid

$$\kappa = \{0.01, 0.04, 0.1\} \tag{6.10}$$

$$\bar{X} = \{-1, -4, -7\} \tag{6.11}$$

$$\eta = \{1, 4, 7\} \tag{6.12}$$

$$a = \{0.98, 0.996, 0.998\} \tag{6.13}$$

81 different combinations were tested. Each one of them has been run 1000 times and new log-price dynamics has been simulated every time. To check our results we compared our estimator against the theoretical one, known to be distributed as

$$\frac{\hat{\theta} - \theta^*}{\sigma^*} \sim \mathcal{N}(0, 1) \tag{6.14}$$

where θ^* and σ^* define respectively the true value and the true standard deviation of the specific parameter. In order to derive σ^* the inverse of information matrix evaluated in the true set of parameters has been computed. In practice though, only samples of finite size may be analyzed, leaving us for certain parameter combinations with a sample estimate of the Hessian not close enough to its theoretical counterpart. To overcome this issue, for the critical combinations we averaged the Hessian matrices obtained for each simulation in order to increase the number of replications and ensure closeness between the population Hessian and the one estimated via the sample.

As expected, out of the 1000 simulations that have been run for each combination, only a percentage completed the estimation. As shown in Figure 6.7 five main "clusters of convergence" has been detected. Analysing those 5 different buckets we came to the conclusion that convergence of the method was mainly driven by two parameters, namely κ and $\bar{X}$. Every drastic drop or surge in the acceptance rate coincides indeed with a change in one of those two value. No direct impact was related instead to the change in the other parameters. Table 6.1 highlights the interesting parameters showing that failure rate is rising or falling whenever specific combinations are set. In particular $(\kappa, \bar{X}) = (0.04, -1.0)$, $(\kappa, \bar{X}) = (0.1, -1.0)$ and $(\kappa, \bar{X}) = (0.1, -4.0)$ are proved to be quite unfortunate.

Figures C.1, C.2, C.3, C.4 and C.5 analyze the convergence of the ones that exceeded the 91% threshold, defined by the red dashed line in Figure 6.7. As previously stated we made use of the asymptotic theory in order to conclude that the standardized estimators are normally distributed with mean zero and standard deviation 1. From our simulation we can conclude that the estimators of $\bar{r}$, $\bar{\sigma}$, α and β follow the theory. Same holds true for κ and $\bar{X}$, but some exceptions are shown, for example in Figure C.1. Concerning η and a instead it is clear that in every combination their estimation is far from being normally distributed. Finally we can conclude that the (synthetic) sample size with T = 10000 observations is not big enough to guarantee convergence of the asymptotic theory for all the parameters.

Finally, with respect to the least successful combinations, no evidence has been produced in order to exclude them from the calibration process. We expect that slight modification to our estimation procedure, like different logic for the starting values and a diverse management of the boundary conditions will help in this direction. Unfortunately, modification has not been successfully implemented in the time frame of this master thesis.

Comb	1-9	10-18	19-27	28-36	37-45	46-54	55-63	64-72	73-81
κ	0.01	0.01	0.01	0.04	0.04	0.04	0.1	0.1	0.1
$\bar{X}$	-1.0	-4.0	-7.0	-1.0	-4.0	-7.0	-1.0	-4.0	-7.0

Table 6.1: Combinations used in the Monte Carlo estimation with the specific values of κ and $\bar{X}$.

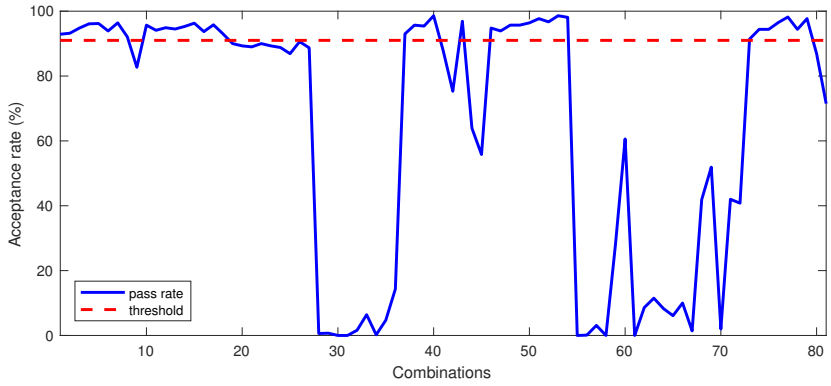

Figure 6.7: Acceptance ratio for the 81 parameter combinations that have been selected for the Monte Carlo estimation. A threshold of 91% pass rate has been set and only combinations that exceeded this level were further analyzed.

Chapter 7

Application to financial data

Two different methodologies has been presented in the previous chapter with maximum likelihood being the most promising one leading to estimated values close to the true ones with small standard errors except the sloppy ones, namely $\bar{X}$ and η. The idea of this chapter is to move away from the known simulated environment and to go into the some real world application. In this direction the chapter is divided in 2 main parts. The first one where maximum likelihood is applied to S&P 500 data and the estimated values are used to re-engineer and detect (ex-post) past jumps to measure the practical efficiency of the model. Using the same calibrated values section 7.2 is comparing instead the general Malevergne and Sornette [2014] framework against the simpler GARCH(1,1) with historical innovations. Both models are deployed to estimate 1d, 10d and 1y VaR. A general discussion on the difficulties encountered and some criticism to the model will finally close the chapter.

7.1 S&P 500 analysis

In order to keep consistency with the analysis done in Malevergne and Sornette [2014] and eventually compare the results we obtained, the same time series has been considered here. Precisely our analysis has been performed on Standard & Poor's 500 data going from 03 January 1950 to 09 June 2014. Maximum likelihood procedure described in the previous chapter has been run on the aforementioned time series. The outcome of our tailored calibration procedure is in line with the one showed in Malevergne and Sornette [2014] where Expectation Maximization (EM) algorithm has been employed. Positive-definite Hessian guarantees also that the estimated parameters, say $\hat{\theta}$ represent an actual maximum and that the procedure did not get stuck at some saddle point. Robust standard errors computed as the square rooted diagonal values of the Information matrix evaluated in the $\hat{\theta}$. As expected

from our previous analysis proving that log-likelihood is sloppy with respect to $\bar{X}$ and η, standard errors of both those parameters appeared to be hundreds times higher than the other ones. Below we show the estimated parameters with the corresponding robust standard errors. Log-likelihood evaluated in optimal point and smallest Hessian eigenvalue are also shown.

$$-\mathcal{L}(\hat{\theta}) = -3.4233$$
$$\|\nabla\mathcal{L}(\hat{\theta})\| = 7.3261e-04$$
$$\min \lambda_i = 3.7320e-05$$

$\bar{r}$ = 0.1125 (0.0001)	κ = 0.0096 (0.0011)
$\bar{\sigma}$ = 0.0976 (0.0004)	$\bar{X}$ = -3.0670 (0.2600)
α = 0.0652 (0.0038)	η = 2.9791 (1.2849)
β = 0.9235 (0.0042)	a = 0.9985 (0.0005)

Calibrated parameters have been then used to estimate ex-post jump activity of the Index within the considered time window. That means, the following expectation has been computed

$$\mathbb{E}(I_t \cdot J_t | \mathcal{F}_t) \tag{7.1}$$

where I_t and J_t describe respectively arrivals and intensities of the jumps while $\mathcal{F}_t$ represents all the information available at time t. A closed-form solution to eq. (7.1) is provided in Malevergne and Sornette [2014] and has been implemented in a C++ framework. Given the estimated jumps and the estimated crash-hazard rate, empirical γ-quantile has been calculated. Defining this quantity, we basically wanted to set a threshold $\bar{j}$, able to separate the "normal activity jumps" from the most "extreme" ones. In Figure 7.1 γ was set to 1 minus the average of the crash-hazard rate over the whole time horizon (in our case equal to 0.9632). In Figure 7.2 instead 99^{th}-percentile was chosen in order to consider only the 100 most extreme detected jumps. As in Figure 7.1, black line describes the following theoretical log-price dynamic of the index

$$\bar{s}_t = \begin{cases} \bar{s}_{t-1} \cdot \exp(\mu_t) & \text{if } J_t < \bar{j} \\ s_t & \text{if } J_t > \bar{j} \end{cases} \tag{7.2}$$

where recalling eq. (2.14) $\mu_t = \bar{r} + \kappa \cdot \lambda_t$ and $s_t = \log S_t$ describes the observed log-price at time t. In words this is exactly like saying that if at

time t the estimated jump would occur and its size would be bigger than the fixed threshold $\bar{j}$, then the theoretical log-price, namely $\bar{s}(t)$, will face a loss proportional to the value of λ_t, forcing it to match its observed counterpart given by $s(t)$. On the other side if the intensity of the estimated jump would be lower then $\bar{j}$ or no jump activity would be observed at all, $\bar{s}(t)$ will increase at a rate equal to the expected return conditional on no crash.

Finally cumulative crash probability shown in the bottom plot of Figure 7.2 has been computed following the same logic. Thus

$$1 - \bar{q}_t = \begin{cases} 1 - \bar{q}_{t-1} \cdot (1 - \lambda_t) & \text{if } J_t < \bar{j} \\ \lambda_t & \text{if } J_t > \bar{j} \end{cases} \tag{7.3}$$

Accordingly to our ex-post estimation four main jump activities has been detected. All of them may be attributed correctly to unstable phases in financial markets. The crash occurred in 1987 has been detected with high accuracy. A huge spike in jump activity is indeed captured in the same period. On the other hand, the dot-com bubble burst around the 2000-01 and the 2008 financial crisis have not been properly detected. Both features may be observed in Figure 7.1 where $\bar{s}_t$ (black line) is behaving differently than its empirical counterpart s_t (blue line). Those remarks may be connected to the fact that the behaviour of the index in the years following the peak reached in 2000-01 is characterized by a fluctuating, but stable regime.

Bottom plot in Figure 7.2 shows (in blue) the cumulative crash probability resulting from the implementation of eq. (7.3). Being 1-$\bar{q}_t$ defined as a function of λ_t, we decided to include in the same plot the crash-hazard rate of the calibrated model in order to give a better comprehension of their tied relationship. In the plot it is clear how the crash-hazard rate is connected to the cumulative crash probability and on a second instance to the jump size. In our framework indeed whenever some jump activity is detected, λ_t is not automatically set to zero, but its value decreases proportionally to the intensity of the crash. Even though from a theoretical point of view eq. (7.3) should help the comprehension of the risk-return dynamics, in reality on this specific application no useful information may be extracted. Too many small activities are indeed detected by the model. Bottom plot in Figure 7.1 resulted then to be too noisy making it difficult to provide any kind of interpretation. The reason may be attributed to the estimated parameter κ. Its value equal to 0.0096 indeed, suggests an average jump intensity of around 1%. Such small value is telling us that in this specific case the model is not able to differentiate between the "normal" daily stock price dynamic and the actual jump (or crash) activity. In bottom plot in Figures 7.1 and 7.2 jumps are considered actual jumps only if their intensities are above

the average. Even though our choice seems logical from a theoretical point of view, there is no fixed rule with this regards. Since this threshold is basically arbitrary, an higher one is then considered in Figure 7.2. Precisely only top 1% of the jumps detected by the model are considered there. As expected, in the top plot of the figure the process defined by eq. (7.2) and described by the black line, moves away from its empirical counterpart, since smaller jumps are now neglected. On the other side the bottom plot is now less noisy and more comprehensible then before. Different phases can be now identified throughout the time window analysed. Higher activity obviously corresponds to periods with higher λ_t where many actual jumps are detected. On the contrary, during phases in which the crash-hazard rate is relatively small the plot shows how this probability is being accumulated. The arrival time of a jump with sufficiently high intensity might be seen indeed as a tipping point in which the risk related to the occurrence of a jump is brought back to its original and theoretical, value given by λ_t.

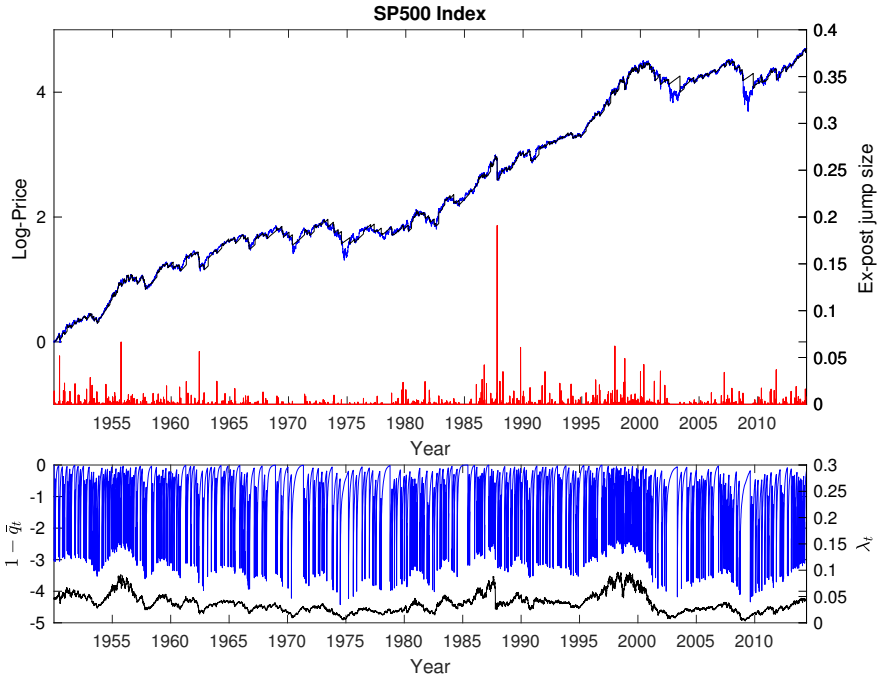

Figure 7.1: Top: observed S&P500 (blue line, left scale) and modified S&P500 (black line, left scale) prices represented in semi-logarithmic scale. Modified dynamic is computed using eq. (7.2). Red bars refer to right scale and describe ex-post jump activity computed using equation 7.1. Bottom: jump probability (black line, right scale) and cumulative jump probability (blue line, left scale) of the observed time series. Cumulative probability computed using eq. (7.3). In both plots the threshold for the jump intensity is set to the $(1 - \sum_t \lambda_t / T)^{\text{th}}$-percentile of the detected activities

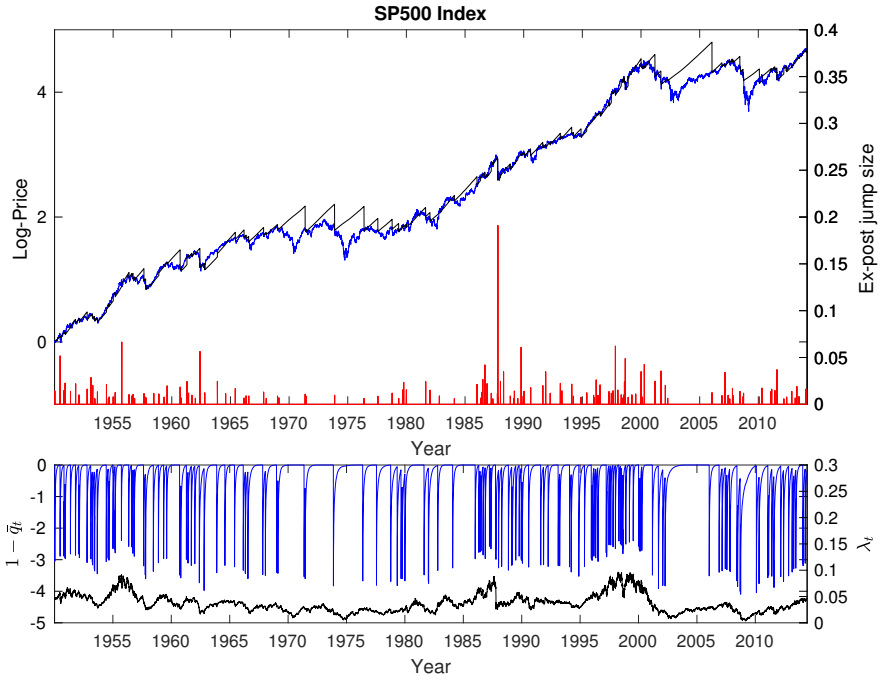

Figure 7.2: Top: observed S&P500 (blue line, left scale) and modified S&P500 (black line, left scale) prices represented in semi-logarithmic scale. Modified dynamic is computed using eq. (7.2). Red bars refer to right scale and describe ex-post jump activity computed using equation 7.1. Bottom: jump probability (black line, right scale) and cumulative jump probability (blue line, left scale) of the observed time series. Cumulative probability computed using eq. (7.3). In both plots the threshold for the jump intensity is set to the 99^{th}-percentile of the detected activities.

7.2 Risk Estimation

In this last section risk measure estimation is assessed and results from the model at hand are compared to the one obtained with a GARCH(1,1) with historical innovations. Even though both methods are generally able to capture changes in market dynamics, in our framework VaR violations are proved to occur in clusters. This behaviour is mainly due to the fact that a static time window is used here to perform the analysis. We further expect that a slightly more complex approach deploying a rolling window estimate would lead to a more dynamic procedure able to solve this specific issue.

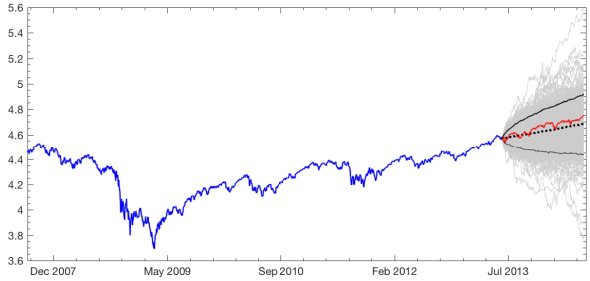

Figure 7.3: Calibrated model is used to simulate 250 daily returns. 1000 Monte Carlo repetitions are performed. Light grey lines: simulated paths attached to empirical S&P500 time series starting from June 2013. Black lines show the 90% confidence interval of the Monte Carlo simulations while its mean is specified by the black dotted one. Red line: observed S&P500 daily returns.

7.2.1 VaR

Here Malevergne and Sornette [2014] model is tested in the context of risk measure estimation in comparison against the well-known GARCH(1,1). Both the models are calibrated on a time window that is now shrunk down to 63 years. 250 empirical observations are removed compared to the calibration done in the previous chapter and they are used as out-of-sample data to backtest our findings. Even though only approximately 1.5% of the sample data are discarded from the previous procedure, minor discrepancies are found in the estimated values compared to the ones obtained in the earlier section. Specifically, this is the case with respect to $\bar{X}$ and η. This outcome was anyway expected due to the fact that the log-likelihood is shown to be flat as a function of those parameters and the can be widely changed without affecting its value.

Figure 7.3 shows how the simulated paths are able to capture what was ac-

tually observed in the out-sample year represented by the red line. In the same figure black continuous and dotted lines illustrate respectively 90% confidence interval and mean of the Monte Carlo trajectories. $VaR_{95\%}$ and $VaR_{99\%}$ are computed for both 10 days and 1 year holding periods. Full parametric approach is used to simulate new price trajectories from the the calibrated models. Although 1000, 10000 100000 Monte Carlo repetitions were performed, in the following figures only the first case is shown since no remarkable differences are found between them. In the table below estimated parameters are listed together with their robust standard errors.

Full Model	
$-\mathcal{L}(\hat{\theta}) = -3.42107$	
$\bar{r}$ = 0.1117 (0.0001)	$\bar{\sigma}$ = 0.0988 (0.0004)
α = 0.0653 (0.0038)	β = 0.9237 (0.0042)
κ = 0.0099 (0.0011)	$\bar{X}$ = -3.1366 (0.2644)
η = 3.0741 (1.2933)	a = 0.9986 (0.0005)
GARCH(1,1) Model	
$-\mathcal{L}(\hat{\theta}) = -3.4174$	
$\bar{r}$ = 0.1201 (0.0001)	$\bar{\sigma}$ = 0.1764 (1.6e-7)
α = 0.0820 (0.0015)	β = 0.9115 (0.0022)

Full Model

Estimated $VaR_{95\%}$ are displayed in Figure 7.4 by blue dashed lines, while the 99% cases are represented by the black ones. In the same Figure top 2 plots are related to the 10-day VaR estimation. Holding period of 1-year is assessed instead in the last two graphs. In every situation we considered here VaR estimates are exceeded by the observed time series. 95% cases shows also some high correlation in the timing of the exceedances. Only one violation is spotted instead when the higher threshold of 99% is selected. As mentioned in section 5.5 clustered exceedances should not be observed whenever a good estimator is in place. Indeed the underlying model should be able to capture changes in the risk factor dynamics. In the specific case of 10-days and 1-year $VaR_{99\%}$ both Kupiec's Proportion of Failures Test (POF) and Christoffersen's (IND) Tests did not reject the null hypothesis providing evidence that the model performs well on average. On the other side when $VaR^{1yr}_{95\%}$ estimation was tested for independence, alternative hypothesis was rejected in favor of the null. With respect to the number of observed viola-

tions instead, no statistical evidence has been provided by the coverage test in order not to reject the alternative hypothesis.

Garch(1,1)

Here Value-at-Risk is estimated simulating GARCH(1,1) dynamics. In order to generate more realistic scenarios historical innovations are used to generate the new trajectories. Results are grouped in Table 7.1 and Figure 7.5. GARCH(1,1) has proved to be more conservative with less violations occurred compared to the previous method. In particular $\text{VaR}^{1yr}_{99\%}$ is never exceeded. With respect to $\text{VaR}^{1\text{yr}}_{95\%}$ and $\text{VaR}^{1\text{yr}}_{99\%}$ instead, unconditional cover tests are proved inefficient in assessing the goodness of the model. Clustered violation displayed in the bottom plot of Figure 7.5 suggests that the model is not able to capture changes in the risk factor dynamics, but since for the analysis a static window is used, such results was expected. Finally, all the three tests performed on $\text{VaR}^{10d}_{95\%}$ and $\text{VaR}^{10d}_{99\%}$ provide statistical evidence to accept the null in favor of the alternative one.

Full Model	$\text{VaR}^{10d}_{95\%}$	$\text{VaR}^{10d}_{99\%}$	$\text{VaR}^{1yr}_{95\%}$	$\text{VaR}^{1yr}_{99\%}$
Exceptions	2	1	6	1
Coverage Test	2.7956	2.8896	***4.3686***	1.1764
Independence Test	1.0205	0.2507	0.2963	0.0081
Combined Test	3.8161	3.1403	4.6649	1.1845
GARCH(1,1)	$\text{VaR}^{10d}_{95\%}$	$\text{VaR}^{10d}_{99\%}$	$\text{VaR}^{1yr}_{95\%}$	$\text{VaR}^{1yr}_{99\%}$
Exceptions	1	1	5	0
Coverage Test	0.4130	2.8896	***6.0715***	***5.0252***
Independence Test	0.2507	0.2507	0.2049	NaN
Combined Test	0.6637	3.1403	***6.2764***	NaN

Table 7.1: Backtesting results for VaR estimation in the considered cases. Bold values highlight situations when critical value of the statistic was exceeded and null hypothesis rejected. When no violations happened Independence Test was not applicable and NaN is shown.

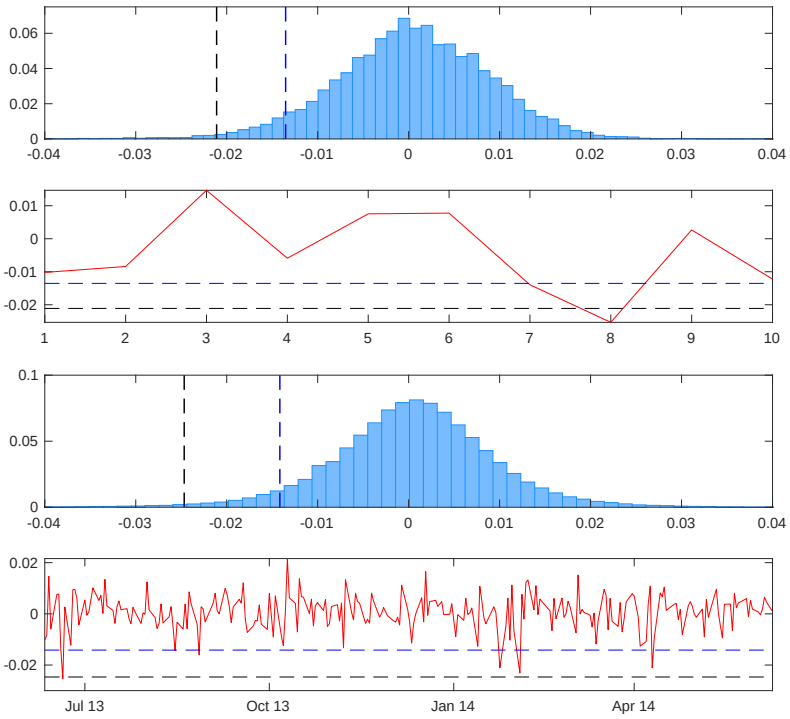

Figure 7.4: Calibrated full model is used to simulate 10 and 250 daily returns. 1000 Monte Carlo repetitions are performed. Black and blue dashed lines represent VaR at 99% and 95% confidence, respectively. 10-day $\text{VaR}_{95\%}$ is estimated around -1.35%, while $\text{VaR}_{99\%}$ is -2.11%. As expected instead higher estimates are shown for the 1-year holding period, namely $\text{VaR}_{95\%} = -1.403\%$ $\text{VaR}_{99\%} = -2.458\%$. First and third plot from the top display normalized histogram of simulated returns for respectively 10 and 250 days. Finally red lines in the second and fourth graph show S&P500 realized return.

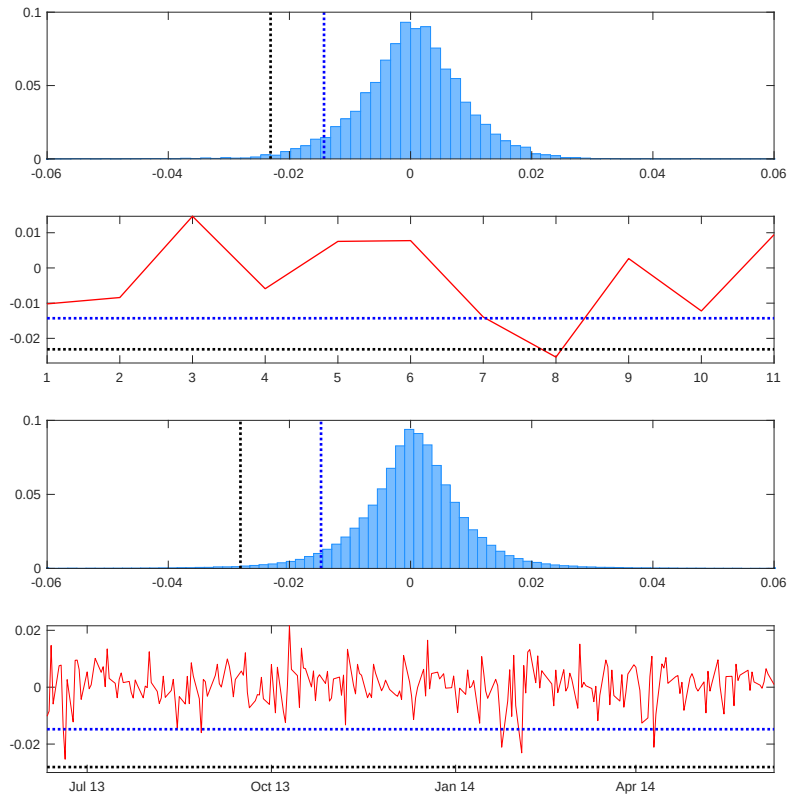

Figure 7.5: Calibrated GARCH(1,1) model is used to simulate 10 and 250 daily returns. 1000 Monte Carlo repetitions are performed. Black and blue dotted lines represent $\text{VaR}_{99\%}$ and $\text{VaR}_{95\%}$, respectively. 10-day $\text{VaR}_{95\%}$ is estimated around -1.39%, while $\text{VaR}_{99\%}$ is -2.19%. As expected instead higher estimates are shown for the 1-year holding period, namely $\text{VaR}_{95\%} = -1.47\%$ $\text{VaR}_{99\%} = -2.81\%$. First and third plot from the top display normalized histogram of simulated returns for respectively 10 and 250 days. Finally red lines in the second and fourth graph show S&P500 realized return.

Chapter 8

Conclusion

The two-fold goal of this master thesis was to calibrate the jump-diffusion model developed by Malevergne and Sornette [2014] and to extend the work that has been done in Berntsen [2015]. Strengths and weaknesses of the model were investigated and highlighted throughout the work. We describe how synthetically generated data was able to capture dynamics widely observed in financial data. In addition we demonstrate the difficulty in distinguishing between regular and abnormal activity.

With regards to the calibration two different procedures were adapted to the framework. In the first approach the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm developed by Spall [1987] was embedded within the particle filter implemented in Berntsen [2015]. Convergence of the method was not guaranteed due to the complicated dependencies between the hyper-parameters and the latent variables extracted by the filter. In addition flatness of the likelihood with respect to some of the parameters made their calibration even more cumbersome.

In a second part of the thesis, offline maximization of the likelihood was performed. The analytic gradient and Hessian derived in Malevergne and Sornette [2014] were also implemented in order to guarantee the maximum possible precision. The offline approach proved to be more stable than the online counterpart, but still for some of the parameters standard errors are shown to be too large.

In the next step, a spectral analysis of the Hessian was performed in order to disentangle the importance of the different parameters. Profile likelihood showed promise in this direction. Unfortunately no statistical significance was provided to the estimates, mainly due to the sloppy nature of our problem, but also to the highly ill-conditioned Hessian of the cost function.

A Monte Carlo calibration approach was finally implemented to analyze 81 different parameter combinations. Success rate of the estimation was proved

to be influenced primarily by two parameters, namely the ones describing the average jump size and mispricing index. Variation of the other ones neither improved nor worsened the quality of the estimation.

In the last part of the work the calibration procedure was finally applied to 64 years of empirical financial data. The model estimates matched the results obtained in Malevergne and Sornette [2014]. Calibrated parameters were then used to infer ex-post jump probabilities. The model detected and recovered some of the past extremes. In particular it was able to fully capture the 1987 crash. On the contrary many other negative jumps were miserably missed, namely the ones at the end of dot-com bubble in the early 2000 and the 2008 subprime crisis. Lastly, Value-at-Risk estimation is assessed to show applicability of the model in a general risk management environment.

Many difficulties have been encountered in the calibration of the model at hand. The framework aimed to capture extreme behaviours, but in reality has proved insufficient in distinguishing between regular and abnormal regimes. Additionally, in this work no attention is paid to the composition of the analysed index over time. We expect that further investigations in this direction, may shed lights on the reason why the model has proved inefficient in certain occasions.

We further notice that in the economy of the model certain parameters are extremely important, whilst others are not. This difference can be shown in the high condition number of the Hessian associated to the log-likelihood of the model and it is crucial in the calibration procedure. In future work matrix shrinkage approaches, like the one presented in Ledoit and Wolf [2003], might also prove helpful in this respect.

Another line of research is to implement the adaptive variation of the SPSA algorithm described in Spall [2000], which is taking into account first and second derivative of the cost function. Our choice to deploy a gradient-free method, was based on its suitability for better performance. Unfortunately, due to the complexity of the model, closed-form gradient and Hessian are essential in order to perform a meaningful estimation.

Appendix A

GARCH(1,1)

Here is the procedure has been used to derive and implement the log-likelihood of the GARCH(1,1) process together with its analytical gradient and Hessian. Their implementation has been crucial to give a more accurate comparison of the results obtained by the DLIB optimisation library. Given the following general settings of a GARCH(1,1) process

$$\varepsilon_t = v_t \sigma_t \qquad v_t \sim \mathcal{N}(0,1) \tag{A.1}$$

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

ε_t is normally distributed with mean zero and conditional variance equal to σ_t^2, while its density conditionally on the past information is

$$p(\varepsilon_t | \varepsilon_{t-1}, \cdots, \varepsilon_0) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{\varepsilon_t^2}{2\sigma_t^2}} \tag{A.3}$$

Let now $\theta = (\omega, \alpha, \beta)'$ be the vector containing the unknown parameters. The log-likelihood function, without constants, reads

$$l(\theta) = -\sum_{t=2}^{T} \left(\ln \sigma_t^2 + \frac{\varepsilon_t^2}{\sigma_t} \right) \tag{A.4}$$

Deriving with respect to the parameter vector θ and rearranging the term, the gradient of the log-likelihood (A.4) is given by

$$\frac{\partial l(\theta)}{\partial \theta} = \sum_{t=2}^{T} \left(\frac{\varepsilon_t^2}{(\sigma_t^2)^2} - \frac{1}{\sigma_t^2} \right) \frac{\partial \sigma_t^2}{\partial \theta} \tag{A.5}$$

where

$$\frac{\partial \sigma_t^2}{\partial \theta} = (1, \varepsilon_{t-1}^2, \sigma_{t-1}^2)' + \beta \frac{\partial \sigma_{t-1}^2}{\partial \theta} \tag{A.6}$$

Re-parametrization

In order to increase the stability of our estimates and to have a gradient with components having similar order of magnitude, we re-parametrize the dynamics A.2 using the long-time variance $\bar{\sigma}^2$ in the following way

$$\sigma_t^2 = \underbrace{\bar{\sigma}^2(1-\alpha-\beta)}_{=\omega} + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{A.7}$$

The above modification leads to the same analytical gradient shown in (A.5), beside the last term $\partial\sigma_t^2/\partial\theta$ that is now equal to

$$\frac{\partial \sigma_t^2}{\partial \theta} = \begin{pmatrix} 1-\alpha-\beta \\ -\bar{\sigma}^2 + \varepsilon_{t-1}^2 \\ -\bar{\sigma}^2 + \sigma_{t-1}^2 \end{pmatrix} + \beta \frac{\partial \sigma_{t-1}^2}{\partial \theta}. \tag{A.8}$$

Non-central dynamics
Suppose now that the observed process of log-returns is given by

$$r_t = \bar{r} + \varepsilon_t \tag{A.9}$$

where ε_t is described by the aforementioned GARCH(1,1) model, that is

$$r_t \sim \mathcal{N}(\bar{r}, \sigma_t^2) \tag{A.10}$$

and accordingly the variance σ_t^2 is

$$\sigma_t^2 = \bar{\sigma}^2(1-\alpha-\beta) + \alpha(r_{t-1}-\bar{r})^2 + \beta\sigma_{t-1}^2 \tag{A.11}$$

Essentially the idea is now to consider the mean of the observed time-series, say $\bar{r}$, as an additional parameter that need to be estimated together with the long-time variance, $\bar{\sigma}^2$ and the couple (α, β), whose sum is describing the persistence. Log-likelihood is of course the same as before, while a

small difference needs to be considered in the gradient construction, due to the 4^{th}-parameter involved. Precisely, given the vector $\theta_{\text{new}} = (\bar{r}, \bar{\sigma}^2, \alpha, \beta)$ the 4 components of the gradient at time t are given by

Version 1

$$\frac{\partial l_t(\theta_{\text{new}})}{\partial \bar{r}} = \left(\frac{\varepsilon_t^2}{(\sigma_t^2)^2} - \frac{1}{\sigma_t^2} \right) \frac{\partial \sigma_t^2}{\partial \bar{r}} + \frac{2\varepsilon_t}{\sigma_t^2} \tag{A.12}$$

$$\frac{\partial l_t(\theta_{\text{new}})}{\partial \bar{\sigma}^2} = \left(\frac{\varepsilon_t^2}{(\sigma_t^2)^2} - \frac{1}{\sigma_t^2} \right) \frac{\partial \sigma_t^2}{\partial \bar{\sigma}^2} \tag{A.13}$$

$$\frac{\partial l_t(\theta_{\text{new}})}{\partial \alpha} = \left(\frac{\varepsilon_t^2}{(\sigma_t^2)^2} - \frac{1}{\sigma_t^2} \right) \frac{\partial \sigma_t^2}{\partial \alpha} \tag{A.14}$$

$$\frac{\partial l_t(\theta_{\text{new}})}{\partial \beta} = \left(\frac{\varepsilon_t^2}{(\sigma_t^2)^2} - \frac{1}{\sigma_t^2} \right) \frac{\partial \sigma_t^2}{\partial \beta} \tag{A.15}$$

Version 2

$$\frac{\partial l_t(\theta_{\text{new}})}{\partial \theta_{\text{new}}} = \left(\frac{\varepsilon_t^2}{(\sigma_t^2)^2} - \frac{1}{\sigma_t^2} \right) \frac{\partial \sigma_t^2}{\partial \theta_{\text{new}}} + \left(\frac{2\varepsilon_t}{\sigma_t^2}, 0, 0, 0 \right)^T \tag{A.16}$$

Version 3

$$\begin{pmatrix} \partial l_t(\theta_{\text{new}})/\partial \bar{r} \\ \partial l_t(\theta_{\text{new}})/\partial \bar{\sigma}^2 \\ \partial l_t(\theta_{\text{new}})/\partial \alpha \\ \partial l_t(\theta_{\text{new}})/\partial \beta \end{pmatrix} = \left(\frac{\varepsilon_t^2}{(\sigma_t^2)^2} - \frac{1}{\sigma_t^2} \right) \frac{\partial \sigma_t^2}{\partial \theta_{\text{new}}} + \begin{pmatrix} 2\varepsilon_t/\sigma_t^2 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{A.17}$$

where, similarly as before, the gradient of the process σ_t^2 is

$$\frac{\partial \sigma_t^2}{\partial \theta_{\text{new}}} = \begin{pmatrix} -2\alpha\varepsilon_{t-1} \\ 1 - \alpha - \beta \\ -\bar{\sigma}^2 + \varepsilon_{t-1}^2 \\ -\bar{\sigma}^2 + \sigma_{t-1}^2 \end{pmatrix} + \beta \frac{\partial \sigma_{t-1}^2}{\partial \theta_{\text{new}}}. \tag{A.18}$$

Hessian

$$\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'} = \sum_{t=2}^{T}\left[\left(1-\frac{2\varepsilon_t^2}{\sigma_t^2}\right)\frac{1}{(\sigma_t^2)^2}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}+\left(\frac{\varepsilon_t^2}{\sigma_t^2}-1\right)\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'}\right] \tag{A.19}$$

Appendix B

SPSA Pseudo-code

A more detailed pseudo-code for the SPSA algorithm implementation is presented here.

1. Initialization. Given some prior distribution for the starting values of the parameters, for each particle $\{i\}_0^{N-1}$, simulate

$$\theta_0^{(i)} = (\bar{r}^{(i)}, \bar{\sigma}^{(i)}, \alpha^{(i)}, \beta^{(i)}, \kappa^{(i)}, \bar{X}^{(i)}, \eta^{(i)}, a^{(i)}) \tag{B.1}$$

2. Initialize SPSA parameters c, A, α, r, a

3. Compute $(\hat{X}_t^{(i)}, \hat{V}_t^{(i)})$ for each parameter particle $\theta_{t-1}^{(i)}$ and simulate new jump times and sizes

$$\hat{I}_t^{(i)} \sim Bern(L(\hat{X}_t^{(i)}))$$

$$\hat{J}_t^{(i)} \sim Exp(1) \text{ if } \hat{I}_t^{(i)} = 1$$

4. Compute SPSA gain sequences $c_t = \frac{c}{(t+1)^r}$ and $\gamma_t = \frac{a}{(A+t+1)^\alpha}$.

5. Generate a p-dimensional ($p = 8$ in our case) simultaneous perturbation vector Δ_t following a Bernoulli ± 1 distribution with probability $\frac{1}{2}$ for each ± 1 independently. I built it like

$$\Delta_t = [\underbrace{\Delta_{t,0}, \cdots, \Delta_{t,N-1}}_{\text{1st parameter}}, \cdots, \underbrace{\Delta_{t,N(p-1)}, \cdots, \Delta_{t,pN-1}}_{\text{p-th parameter}}]$$

so the N p-dimensional perturbation vectors are given by

$$\Delta_t^{(0)} = (\Delta_{t,0}, \Delta_{t,N}, \cdots, \Delta_{t,N(p-1)})$$

$$\vdots$$

$$\Delta_t^{(N-1)} = (\Delta_{t,N-1}, \Delta_{t,2N-1}, \cdots, \Delta_{t,pN-1})$$

6. Compute the perturbed parameter particles

$$\theta_{t-1}^{(i)+} = (\theta_{t-1}^{(i)} + c_t\Delta_t^{(i)})$$

$$\theta_{t-1}^{(i)-} = (\theta_{t-1}^{(i)} - c_t\Delta_t^{(i)})$$

7. For $i =$ 0:N-1 compute

$$\hat{X}_{\pm}^{(i)} = (1-a^{\pm})\bar{X}^{\pm} + a^{\pm}\hat{X}_{t-1} + \eta^{\pm}(r_{t-1} - \mu^{\pm})$$

$$\hat{V}_{\pm}^{(i)} = \hat{\sigma}^{\pm}(1-\alpha^{\pm}-\beta^{\pm}) + \alpha^{\pm}(r_{t-1}-\mu^{\pm})^2 + \beta^{\pm}\hat{V}_{t-1}$$

and sample

$$\tilde{I}_{\pm}^{(i)} \sim Bern(L(\hat{X}_{\pm}^{(i)}))$$

$$\tilde{J}_{\pm}^{(i)} \sim Exp(1) \text{ if } \tilde{I}_{\pm}^{(i)} = 1$$

8. Evaluate the cost function (likelihood) for both perturbations

$$F(\theta_{t-1}^{(i)+}) = p(r_t|\tilde{I}_{+}^{(i)}, \tilde{J}_{+}^{(i)}, \theta_{t-1}^{(i)+})$$

$$F(\theta_{t-1}^{(i)-}) = p(r_t|\tilde{I}_{-}^{(i)}, \tilde{J}_{-}^{(i)}, \theta_{t-1}^{(i)-})$$

9. Gradient approximation. For each parameter particle the corrisponding gradient is

$$\hat{\nabla}F(\theta_{t-1}^{(i)}) = (\hat{\nabla}F_1(\theta_{t-1}^{(i)}), \cdots, \hat{\nabla}F_p(\theta_{t-1}^{(i)}))$$

where the components of the gradient approximation are

$$\hat{\nabla}F_m(\theta_{t-1}^{(i)}) = \frac{\hat{F}(\theta_{t-1}^{(i)+}) - \hat{F}(\theta_{t-1}^{(i)-})}{2c_t\Delta_{t,m}^{(i)}}$$

and $\Delta_{t,m}^{(i)}$ denote the m^{th} component of $\Delta_t^{(i)}$

10. Parameter update.

$$\theta_t^{(i)} = \theta_{t-1}^{(i)} + \gamma_t\hat{\nabla}F(\theta_{t-1}^{(i)})$$

11. Compute the weights

$$\begin{aligned}\omega_t^{(i)} &\propto p(r_t|\hat{I}_t^{(i)}, \hat{J}_t^{(i)}, \theta_t^{(i)}) \\ &= \varphi(r_t|\mu_t + \kappa_t L(X_t) - \kappa_t\hat{J}_t^{(i)}\hat{I}_t^{(i)}, V_t)\end{aligned} \tag{B.2}$$

where $\varphi(x;\mu,\sigma)$ is the Gaussian density.

12. Multinomial re-sampling

$$z(i) \sim Mult(N; \omega_t^{(0)}, \cdots, \omega_t^{(N-1)})$$

and set

$$\hat{I}_t^{(i)} = \hat{I}_t^{z(i)} \qquad \hat{J}_t^{(i)} = \hat{J}_t^{z(i)} \qquad \theta_t^{(i)} = \theta_t^{z(i)}$$

Re-assign importance weights $\omega_t^{(i)} = \frac{1}{N}$

13. Output. The estimate of the parameters is

$$\theta_t = \sum_{i=0}^{N-1} \omega_t^{(i)} \theta_t^{(i)}$$

Appendix C

Monte Carlo Estimation Results

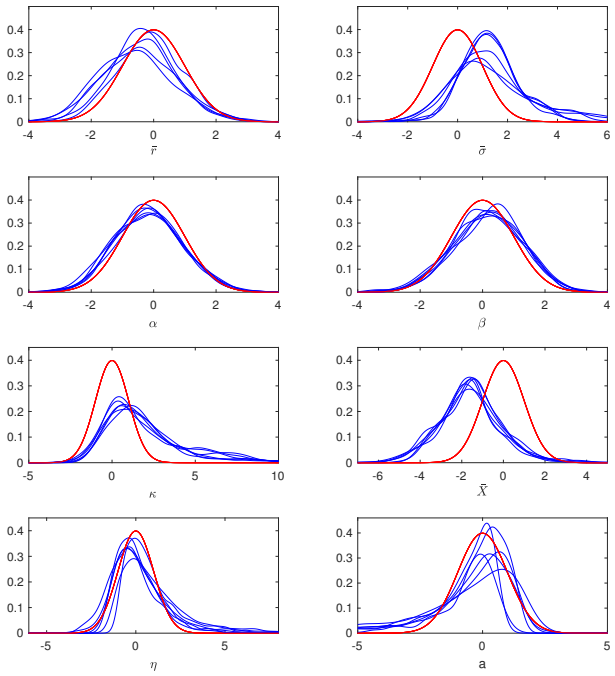

Figure C.1: Gaussian kernel smoothing functions vs theoretical distributions of the estimated parameters with true values equal to $\bar{r} = 0.07$, $\bar{\sigma} = 0.25$, $\alpha = 0.05$, $\beta = 0.94$, $\kappa = 0.01$, $\bar{X} = -1$, $\eta = \{1, 4, 7\}$, $a = \{0.98, 0.996, 0.998\}$

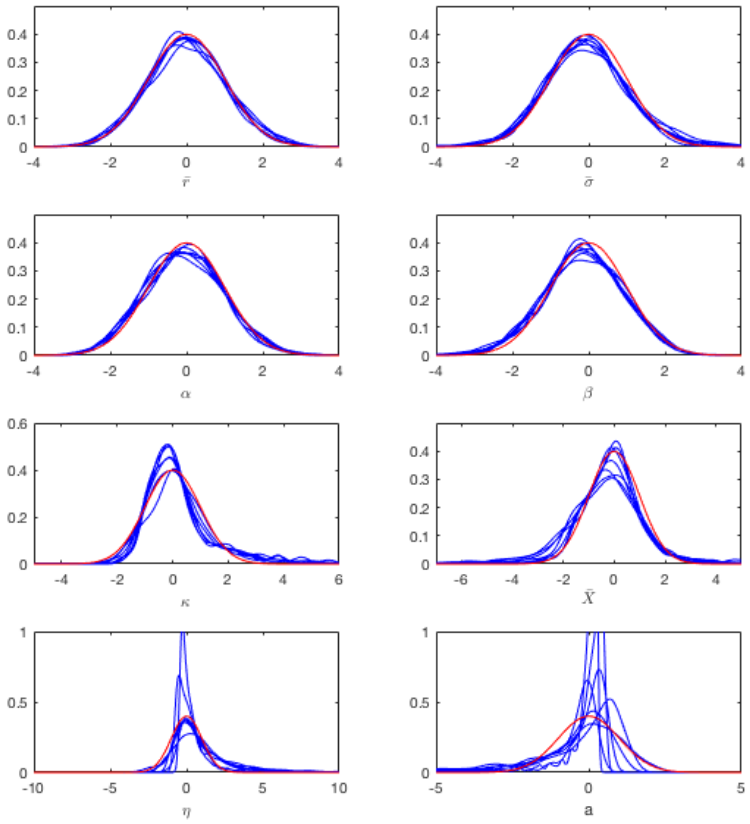

Figure C.2: Gaussian kernel smoothing functions vs theoretical distributions of the estimated parameters with true values equal to $\bar{r} = 0.07$, $\bar{\sigma} = 0.25$, $\alpha = 0.05$, $\beta = 0.94$, $\kappa = 0.01$, $\bar{X} = -4$, $\eta = \{1, 4, 7\}$, $a = \{0.98, 0.996, 0.998\}$

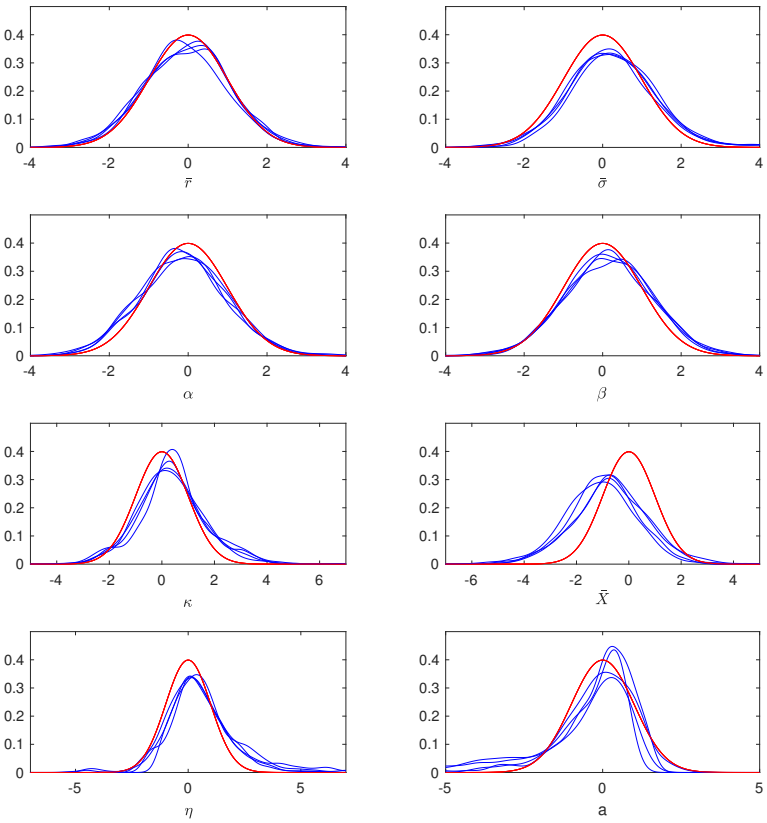

Figure C.3: Gaussian kernel smoothing functions vs theoretical distributions of the estimated parameters with true values equal to $\bar{r} = 0.07$, $\bar{\sigma} = 0.25$, $\alpha = 0.05$, $\beta = 0.94$, $\kappa = 0.04$, $\bar{X} = -4$, $\eta = \{1, 4, 7\}$, $a = \{0.98, 0.996, 0.998\}$

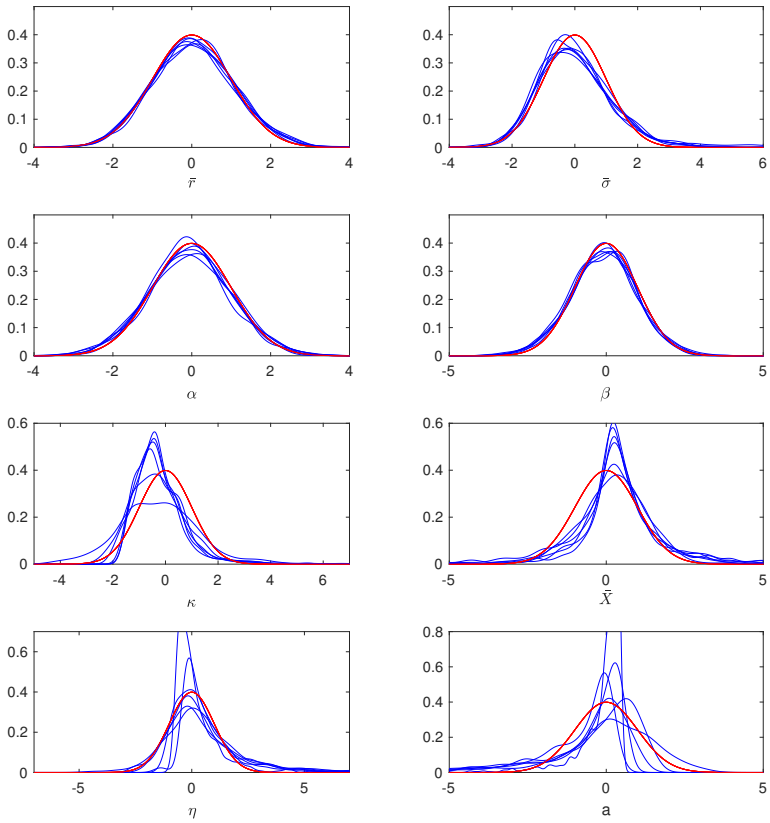

Figure C.4: Gaussian kernel smoothing functions vs theoretical distributions of the estimated parameters with true values equal to $\bar{r} = 0.07$, $\bar{\sigma} = 0.25$, $\alpha = 0.05$, $\beta = 0.94$, $\kappa = 0.04$, $\bar{X} = -7$, $\eta = \{1, 4, 7\}$, $a = \{0.98, 0.996, 0.998\}$

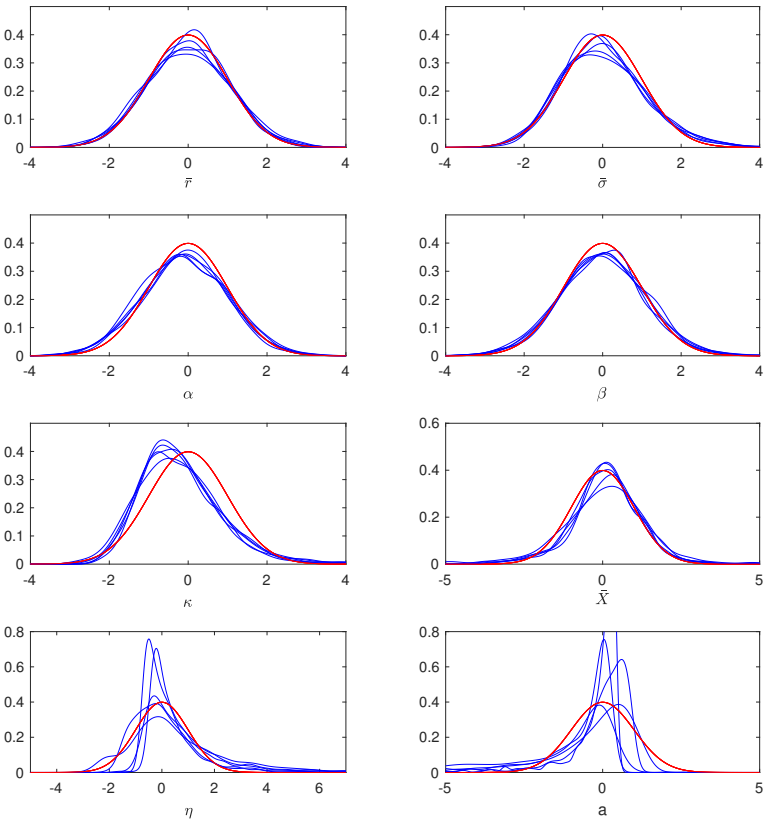

Figure C.5: Gaussian kernel smoothing functions vs theoretical distributions of the estimated parameters with true values equal to $\bar{r} = 0.07$, $\bar{\sigma} = 0.25$, $\alpha = 0.05$, $\beta = 0.94$, $\kappa = 0.1$, $\bar{X} = -7$, $\eta = \{1, 4, 7\}$, $a = \{0.98, 0.996, 0.998\}$

Bibliography

J. Aitchison and S.M. Shen. Logistic-normal distributions: Some properties and uses. *Biometrika*, 67(2):261–272, 1980.

D. Alspach and H. Sorenson. Nonlinear bayesian estimation using gaussian sum approximations. *IEEE transactions on automatic control*, 17(4):439–448, 1972.

O. Barndorff-Nielsen. On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, 70(2):343–365, 1983.

P. Berntsen. Particle filter adapted to jump-diffusion model of bubbles and crashes with non-local crash-hazard rate estimation. Master's thesis, ETH, 2015.

F. Black. Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, pages 177–181, 1976.

O.J. Blanchard and M.W. Watson. Bubbles, rational expectations and financial markets, 1982.

O. Cappé, E. Moulines, and T. Rydén. Inference in hidden markov models. In *Proceedings of EUSFLAT Conference*, pages 14–16, 2009.

B.L. Chan, A. Doucet, and V.B. Tadic. Optimisation of particle filters using simultaneous perturbation stochastic approximation. In *Acoustics, Speech, and Signal Processing*, volume 6, pages VI–681. IEEE, 2003.

P. Christoffersen, K. Jacobs, and K. Mimouni. Models for s&p 500 dynamics: evidence from realized volatility, daily returns, and option prices. *CREATES Research Paper*, (2007-37), 2007.

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

P. Del Moral. Feynman-kac formulae. In *Feynman-Kac Formulae*, pages 47–93. Springer, 2004.

F. Delbaen. Coherent risk measures on general probability spaces. In *Advances in finance and stochastics*, pages 1–37. Springer, 2002.

A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

B. Eraker, M. Johannes, and N. Polson. The impact of jumps in volatility and returns. *The Journal of Finance*, 58(3):1269–1300, 2003.

V. Filimonov and D. Sornette. A stable and robust calibration scheme of the log-periodic power law model. *Physica A: Statistical Mechanics and its Applications*, 392(17):3698–3707, 2013.

V. Filimonov, G. Demos, and D. Sornette. Modified profile likelihood inference and interval forecast of the burst of financial bubbles. *Swiss Finance Institute Research Paper*, (16-11), 2016.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447, 2002.

A. Johansen, O. Ledoit, and D. Sornette. Crashes as critical points. *International Journal of Theoretical and Applied Finance*, 3(02):219–255, 2000.

N. Kantas, A. Doucet, S.S. Singh, J. Maciejowski, N. Chopin, et al. On particle methods for parameter estimation in state-space models. *Statistical science*, 30(3):328–351, 2015.

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

O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621, 2003.

Y. Malevergne and D. Sornette. Multi-moments method for portfolio management: Generalized capital asset pricing model in homogeneous and heterogeneous markets. *Available at SSRN 319544*, 2002.

Y. Malevergne and D. Sornette. *Extreme financial risks: From dependence to risk management*. Springer Science & Business Media, 2006.

Y. Malevergne and D. Sornette. Jump-diffusion model of bubbles and crashes with non-local behavioral self-referencing. 2014.

G. Marsaglia. Evaluating the normal distribution. *Journal of Statistical Software*, 11(4):1–7, 2004.

A.J. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3):271–300, 2000.

A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: Concepts, techniques and tools*. Princeton university press, 2015.

R.C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144, 1976.

J. Mina, J.Y. Xiao, et al. Return to riskmetrics: the evolution of a standard. *RiskMetrics Group*, 2001.

C. Musso, N. Oudjane, and F. Le Gland. Improving regularized particle filters. In *Sequential Monte Carlo methods in practice*, Statistics for Engineering and Information Science, pages 247–271. Springer, 2001.

J.C. Spall. A stochastic approximation technique for generating maximum likelihood parameter estimates. In *American Control Conference, 1987*, pages 1161–1167. IEEE, 1987.

J.C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE transactions on automatic control*, 37(3):332–341, 1992.

J.C. Spall. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transactions on aerospace and electronic systems*, 34(3):817–823, 1998.

J.C. Spall. Adaptive stochastic approximation by the simultaneous perturbation method. *IEEE transactions on automatic control*, 45(10):1839–1853, 2000.

Q. Wang. Formal analysis for practical gain sequence selection in recursive stochastic approximation algorithms. In *2013 47th Annual Conference on Information Sciences and Systems (CISS)*, 2013.

J.J. Waterfall, F.P. Casey, R.N. Gutenkunst, K.S. Brown, C.R. Myers, P.W. Brouwer, V. Elser, and J.P. Sethna. Sloppy-model universality class and the vandermonde matrix. *Physical review letters*, 97(15):150601, 2006.

P. Weil. Confidence and the real value of money in an overlapping generations economy. *The Quarterly Journal of Economics*, pages 1–22, 1987.

D. Wurtz, Y. Chalabi, and L. Luksan. Parameter estimation of arma models with garch/aparch errors an r and splus software implementation. *Journal of Statistical Software, forthcoming*, 2006.

Università della Svizzera italiana

Facoltà di scienze economiche

DICHIARAZIONE

Dichiaro sul mio onore che quanto scritto nel presente lavoro è stato da me redatto e che, citazioni escluse, nessuna parte è stata copiata da pubblicazioni scientifiche, Internet o da lavori di ricerca già presentati in ambito accademico da me o da altri studenti.

Nel caso di parti tratte da pubblicazioni scientifiche, da Internet o da altri documenti, ne ho espressamente e direttamente indicato la fonte alla fine della citazione o a pié di pagina.

Dichiaro inoltre di aver preso atto delle sanzioni previste in caso di plagio dal vigente regolamento degli studi.

Nome e cognome: Mattias Franchetto

Numero di matricola: 13-995-246

Lugano il 5/12/2016 Firma

DECLARATION OF ACADEMIC HONESTY

I, the undersigned, hereby declare that I am the author of the present paper and that, except where specifically acknowledged, no parts have been copied from other authors or sources, or from papers previously submitted for assessment by myself or other students.

Any paragraph or portion of text that I have excerpted from a scientific publication, the Internet, or other sources of information has been duly placed in quotation marks and explicitly cited in a clear footnote reference.

Additionally, I declare that I have read and understood the Faculty's provisions with regard to student plagiarism, and am aware of the penalties outlined under the appropriate articles of the current (December 2007) *Student Regulations.*

Student's Name and Surname: Mattias Franchetto

Student ID card no.: 13-995-246

Lugano the 5/12/2016 Signature: