UNIVERSITÀ DEGLI STUDI DI PAVIA
DOTTORATO DI RICERCA IN FISICA – XXIII CICLO

# Modeling and Analysis of Financial Time Series beyond Geometric Brownian Motion

*Danilo Delpini*


**Tesi per il conseguimento del titolo**

**ISBN 978-88-95767-40-6**

Università degli Studi di Pavia

Dipartimento di Fisica Nucleare e Teorica

Istituto Nazionale di Fisica Nucleare


DOTTORATO DI RICERCA IN FISICA – XXIII CICLO

# MODELING AND ANALYSIS OF FINANCIAL TIME SERIES BEYOND GEOMETRIC BROWNIAN MOTION

dissertation submitted by

***Danilo Delpini***

to obtain the degree of

**DOTTORE DI RICERCA IN FISICA**

Supervisor: Prof. Oreste Nicrosini (I. N. F. N. & Università di Pavia)

Referee: Prof. Fabrizio Lillo (Università di Palermo & Santa Fe Institute)

**Cover:** *Pompeii: Gauss vs fat-tails,*
2008 Copyright Espen Gaarder Haug
(http://web.mac.com/espenhaug/iWeb/Art/QuantArt.html)


ISBN 978-88-95767-40-6

Dottorato di Ricerca in Fisica
Università di Pavia, Italy
Printed in Pavia, November 2010

*To my family*

# Contents

# Introduction

This Ph.D. thesis aims at speaking about physics and finance; in its intent it would like to focus on modeling and analysis of financial time series rather than finance itself, even though financial applications are a quite natural way out of the former. As a physicist, I appreciate the term *Econophysics*, coined in the mid nineties by H. Stanley to address the growing number of papers by physicists venturing into the lands of financial markets and trying to find out the underlying price dynamics, if any. Physicists had already recognized that market dynamics could not be tackled with the tools of deterministic physics, while more recent attempts to single out analogies with quantum mechanics may sound at least risky. Nevertheless, statistical mechanics and the intrinsic probabilistic character of the quantum world set up a connection with social sciences, as stated by E. Majorana [1] in 1942. More recently, the study of complex systems strengthened the link, being clearer and clearer that collective behaviors, not easily justifiable in terms of individual rational decisions, have to be taken into account when addressing relevant financial phenomena like bubbles and crashes: many is not complex and complex is more. For sure, physicists are trained to think about predictions as probabilistic concepts as well, and many of them master the statistical tools adequate to analyze financial prices viewed as the outcome of many endogenous and exogenous factors subject to uncertainty. In particular, statistical mechanics provided concepts and frameworks which have found unexpected analogies in financial contexts. As an example, the Tsallis distribution, emerging in the scope of non extensive statistical mechanics, has proved to be one of the best candidate to model high frequency price changes and it can be justified as the result of a generalized Langevin dynamics for the price return in which the damping coefficient is stochastic either. Even more important, physicists often own the curiosity which is necessary to go beyond sample analysis, asking for the dynamics which produced that sample.

The other way round, concepts first appeared in social sciences have now permeated physics. In 1900, the random walk process had been used by L. Bachelier to describe the price evolution for the first time, a few years before the seminal papers about Brownian motion by A. Einstein [2] and M. Smolu-

chowski [3]. Not to mention power law distributions: they deeply characterize the scaling laws of many long range correlated systems, but they were introduced by the Italian social economist V. Pareto [4] when investigating the distribution of wealth of individuals. Besides motivations and historical chronicles, eminent physicists like R. N. Mantegna, J. P. Bouchaud, H. E. Stanley and many others, have given clear contributions to the analysis of financial data and phenomena, supporting a quantitatively backed questioning of established paradigms, one for all the Gaussianity of yields which was the building block of standard derivative pricing practice far too long.

The work illustrated in this thesis aims to put itself in the above perspective. The one of Econophysics as an interdisciplinary sphere in which tools coming from different areas are welcome whenever they help in improving our comprehension of financial or economic phenomena, or at least the quantitative description of the known empirical facts and the huge amount of "experimental" data coming from financial markets nowadays. Some of those tools may be familiar to physicists and mathematicians, such as the theory of stochastic processes and the formalism of stochastic calculus, some other are may be less, like Bayesian inference. In the first chapter, we will revise briefly the main evidences of non Gaussianity, known as "stylized fact", focusing on the slow decay of the empirical distributions, especially for high frequency data, and on the non constant erratic behavior of the amplitude of price fluctuations. Even more important, the return-volatility correlation and the volatility autocorrelation, which reflect an underlying dynamics and support the idea of the volatility as a fundamental, even though hidden, process. In Chapter 2 we will introduce a general class of diffusion processes with multiplicative noise which are recurrent in study of many physical systems. We will present an exact algorithmic solution for the moments of the associated probability density function; it will permit to characterize the process and its relaxation modes at all times, revealing interesting scaling properties, not least its ability to generate power law tailed stationary distributions. Some specific limits of this general process turn out to have been considered extensively in the literature to describe the dynamics of the stochastic process driving the volatility in the context of Stochastic Volatility Models, which will be introduced in Chapter 3. In particular we will present the linear approximation of the so called exponential Ornstein-Uhlenbeck model, deriving closed form expressions for the cumulants. Their analysis will confirm the non Gaussian nature of the linear model and will be useful for pricing purposes. Then we will present a different model, in which the volatility distributes asymptotically as an Inverse Gamma variable, in agreement with past empirical analysis. Here, the volatility dynamics is again a special limit of the general process studied in Chapter 2; we will carry out an exhaustive characterization of this model, drawing exact expressions for the return-volatility correlation and the volatility autocorrelation, and we will describe a mechanism by which the power law tails of the volatility induce power law tails for price returns. We will also estimate the

model parameters from daily financial data and we will test to what extent the proposed dynamics is able to capture the scaling properties of empirical return distributions. In Chapter 4 a possible application to financial series of Bayesian Product Partition Models (PPMs) will be described, in which returns are modeled as random variables drawn from possibly different distributions and clusters of statistically similar observations are elicited. We will describe two different approaches: in the first one cluster selection will be made based on the mean of the returns distribution, while in the second the partition structure will refer to the volatilities of the observations. We will also describe a suitable Markov Chain Monte Carlo technique to implement these approaches and we will design an efficient algorithm for outliers identification.

In the last chapter we will address specific financial issues: derivative pricing and risk management. The calibration of the linearized version of the exponential Ornstein-Uhlenbeck model, presented in Chapter 3, will be carried out exploiting real option prices from the Milan Stock Exchange and an efficient and general calibration procedure will be implemented. Along the way we will see also an example of how the features of the return distribution are reflected by market option prices. Then PPMs will be used to compute risk exposure for single assets and the results will be compared with classical Maximum Likelihood estimates based on fat tailed Student-$t$ distributions. Finally we will consider the case of portfolios containing derivative instruments, where non Gaussian features emerge as a consequence of the nonlinearity of those contracts. A general methodology based on the generalized Fourier transform formalism will be proposed to efficiently compute risk estimates as well as their sensitivities for these portfolios. All the obtained results will be summarized at the end of the thesis, along with some possible extension and related research topics and perspectives.

The studies and the results which are the object of this work are documented by the following scientific publications:


[5] G. Bormetti, M. E. De Giuli, D. Delpini and C. Tarantola, *Bayesian Value-at-Risk with Product Partition Models*, Quantitative Finance (2010), doi: 10.1080/14697680903512786.

[6] D. Delpini, G. Bormetti, M. E. De Giuli and C. Tarantola, *Estimating Value-at-Risk with Product Partition Models*, In: S. Co. 2009 - Complex data modeling and computationally intensive statistical methods for estimation and prediction, Proceedings. 2009, Maggioli Editore.

[7] G. Bormetti, V. Cazzola, D. Delpini, G. Montagna and O. Nicrosini, *The low volatility fluctuations regime of the exponential Ornstein-Uhlenbeck model*, Journal of Physics: Conference Series, **221**, 012014 (2010).

[8] G. Bormetti, V. Cazzola and D. Delpini, *Option pricing under Ornstein-Uhlenbeck stochastic volatility: a linear model*, International Journal of Theoretical and Applied Finance, **13**, 1047 (2010).

[9] G. Bormetti and D. Delpini, *Exact moment scaling from multiplicative noise*, Physical Review E, **81**, 032102 (2010).

[10] G. Bormetti, V. Cazzola, D. Delpini and G. Livan, *Accounting for risk of non linear portfolios*, The European Physical Journal B, **76**, 157 (2010).

They have been presented also in the form of oral presentations in the following international conferences, meetings and seminars:

- *Option Pricing under Ornstein-Uhlenbeck Stochastic Volatility Models*, Bayesian Analysis and Econophysics for Financial Markets, mini workshop at the Department of Statistics - Athens University of Economics and Business, Athens, June 2nd-4th, 2009.
- *Estimating Value-at-Risk with Product Partition Models*, S.Co. 2009 - Complex Data Modeling and Computationally Intensive Statistical Methods for Estimation and Prediction, Milano, September 14th-16th, 2009.
- *Ornstein-Uhlenbeck Stochastic Volatility Models and Option Pricing*, Econophysics Colloquium 2009, Erice, October 25th-31st, 2009.
- *Exact Moment Scaling from Multiplicative Noise*, seminar at the Department of Political Economics and Quantitative Methods - University of Pavia, April 28th, 2010.

Chapter 1

# Stylized facts in financial markets

In this chapter, we will review the first attempts to model the price of a financial asset as a stochastic process, from the work of Bachelier, who proposed a Gaussian process for the price itself, to the much more realistic Geometric Brownian Motion (GBM), introduced by Osborne and Samuelson and predicting a Gaussian law for price logarithm. Then, we will summarize a body of empirical evidences about financial time series, surprisingly stable across different markets and known as "stylized facts". We will see that the empirical distributions of price variations are characterized by a probability for extreme events which is much more higher than predicted by a Gaussian law. Contrarily to what assumed by the GBM model, the amplitude of price fluctuations, the so called *volatility*, is not constant. We will see that it varies with time, and that its evolution closely resembles the paths of a stochastic process. We will also inspect the correlations which are present in financial time series, in particular between past returns and future volatility and between volatility values at different times. We will conclude that Geometric Brownian Motion is just an approximation of the price actual dynamics while more realistic descriptions are mandatory in order to capture the empirical evidences discussed here.

## 1.1 The Gaussian paradigm

The modeling of price evolution in financial markets is carried out typically in terms of stochastic processes, of which the most cited example is probably the Brownian motion. The theoretical description of this process was accomplished by Einstein when dealing with the determination of the Avogadro number [2], even though the formalization of the random walk, of which the Brownian motion is a scaling limit, was first attempted by Louis Bachelier in its pioneering Ph.D. thesis [11]. Having in mind the problem of the rational pricing of

derivative instruments in speculative markets, for the first time he proposed to describe prices of financial instruments as stochastic processes. Bachelier considered a Markovian evolution for the stock price, so that the conditional probability of the future price $S_{t+\tau}$ at time $t+\tau$ only depends on the present value $S_t$, and not on the knowledge of the whole past history:

$$p(S_{t+\tau} \mid S_{t'}, t' \leq t) = p(S_{t+\tau} \mid S_t)\,.$$

From this Markov property, Bachelier was able to recognize that the above transition probability satisfies what is now called the Chapman-Kolmogorov equation [1]

$$p(S_{t+\tau} \mid S_0) = \int p(S_{t+\tau} \mid S_t) p(S_t \mid S_0)\, dS_t\,. \tag{1.1}$$

Implicitly, Bachelier assumed also the homogeneity and stationarity conditions

$$p(S_{t+\tau} \mid S_t) = p(S_{t+\tau} - S_t; \tau)$$

and from Eq. (1.1) he deduced the Gaussian law for the distribution of absolute price changes, not recognizing that possible solutions also included the whole family of Levy-Khintchine infinitely divisible distributions (e.g. the stable Levy-Pareto and the Poisson distributions). From the point of view of Stochastic Differential Equations (SDEs), the above assumptions are incorporated assuming that the price change evolution is governed by the following equation

$$dS_t = \mu\, dt + \sigma\, dW_t\ , \quad S_{t=t_0} = S_0\ , \tag{1.2}$$

where the constant $\mu$ is the mean growth rate of $S_t$, $dW_t \doteq W_{t+\mathrm{d}t} - W_t$ represents the infinitesimal variation of a standard Wiener process modeling price fluctuations, and $\sigma$ is a constant modulating their amplitude. It follows directly from the properties of $dW$ that the solution $S_t - S_0$ of Eq. (1.2) for the time scale $\Delta t = t - t_0$ is a Gaussian random variable with mean $\mu\Delta t$ and variance $\sigma^2\Delta t$, and that equation describes what is called an *Arithmetic Brownian Motion* (ABM).

Despite its intuition, the Bachelier's solution is not enough motivated from an economic point of view, since it predicts negative prices with possibly strong probabilities. In 1959 the physicist M. Osborne [13], followed by P. Samuelson and R. Merton [14, 15] during the sixties, proposed a much more realistic description which is obtained by considering a Brownian motion for the price logarithm and replacing Eq. (1.2) with the following

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t\ , \quad S_{t=t_0} = S_0\ . \tag{1.3}$$

What is Normally distributed is now the logarithmic variation[2] $\log(S_t/S_0)$ (*log-return*), instead of the absolute variation, while the price itself distributes

[1] For an introduction to stochastic processes and Îto stochastic calculus, whose rudiments will be used in the following pages, see or instance [12].

[2] We choose the convention of indicating with "log" the logarithm to the base $e$.

according to a Log-Normal law:

$$S_t = S_0\, e^{\left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma W_t} \ .$$

In the context of a multiplicative model such as (1.3), the parameter $\sigma$ is called the *volatility* of the price, and in financial analysis it is a fundamental quantity whose empirical features are to be carefully taken into account when looking for a realistic model of financial data, as we will see in the following section. Since the exponential of the arithmetic mean of logarithms is the geometric mean of actual prices, it is said that Eq. (1.3) defines a *Geometric Brownian Motion*. The choice of modeling returns instead of absolute variations is common in theoretical finance, where it is assumed that price changes are proportional to prices themselves and the fundamental quantities to be modeled are the relative increments[3] $\eta_t$, with $S_t = S_0(1 + \eta_t)$. This approach is justified since stock prices are somewhat arbitrary while relative increments are much more stable across different markets and stocks; however, for high frequency data (variations over time scales of order of minutes of less) the assumption of a multiplicative model is much less justified (see chapter 6 in [16]). As it will be discussed in detail later on, the dynamics (1.3) and a bunch of other assumptions about the financial market, are the basis of the celebrated Black, Scholes and Merton theory of rational option pricing.

## 1.2 Power laws in financial markets

Today, prices of financial products, ranging from stocks, commodities, exchange rates and derivative instruments written on these underlying assets, are available in huge databases and price changes are recorded over time scales of minutes or less. In recent years, the statistical analysis of this data, especially those from the stock and foreign exchange markets, has been one of the main concerns for the Econophysics community and led to collect a number of well defined empirical evidences, surprisingly stable across different markets, known as "stylized facts".

Before discussing these evidences we briefly recall some basic definitions. For a given probability density function (PDF) $p(x;t)$ (generally time dependent), we define the *characteristic function* (CF) as its Fourier transform:

$$f(\omega;t) = \int_{-\infty}^{+\infty} p(x;t)\, e^{i\omega x} dx \ ; \tag{1.4}$$

the *cumulants* can be expressed in terms to $f(\omega;t)$ through the relationship

$$k_n(t) = (-i)^n \left.\frac{\partial^n \log f(\omega;t)}{\partial \omega^n}\right|_{\omega=0} , \tag{1.5}$$

[3]For time scales $\Delta t$ of the order of one trading day of less, relative variations are usually small and in this limit their modeling can be confused with that of the log-returns: $\log(S_t/S_0) \approx (S_t - S_0)/S_0 \doteq \eta_t$.

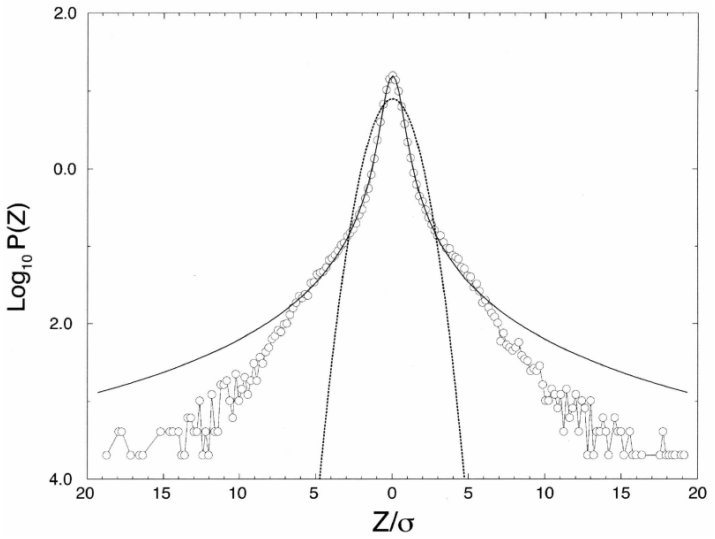



Figure 1.1: Probability density function of high frequency (one minute) changes in the Standard & Poor 500 index from January 1984 to December 1989, compared with the Gaussian distribution fitting the data (dotted line) and with a fit with a Levy stable distribution. From [17].

and they define the relevant features of the shape of the PDF. Besides the mean and the variance, the concepts of *skewness* $\zeta$ and *kurtosis* $\kappa$ are particularly relevant:

$$\zeta(t) = \frac{k_3(t)}{k_2(t)^{3/2}}\,, \quad \kappa(t) = \frac{k_4(t)}{k_2(t)^2}\,. \tag{1.6}$$

Non zero skewness indicates an asymmetric PDF, the right (left) tail being heavier than the other for $\zeta > 0$ ($\zeta < 0$); on the other hand, the kurtosis is an indicator of how "fat" the tails are with respect to those of a Normal distribution which has $\kappa = 0$ by definition.

Even before the introduction of the GBM, systematic deviations from Normality in the empirical distributions of price changes had been collected. More precisely, the observed distribution of the differences $\log(S_{t+\Delta t}) - \log(S_t)$ exhibits a higher and narrower peak and much larger probabilities for extreme events than the Gaussian distribution fitting the data (an example is shown in Fig. 1.1) qualifying itself as a *leptokurtic* distribution, the degree of leptokurtosis being much higher for high frequency data. The first attempt to take into account explicitly this leptokurtic nature dates back to Mandelbrot's studies about cotton prices [18], soon after supported also for stocks by Fama [19], who proposed for $\log(S_t)$ a process whose increments are distributed accordingly to a non Gaussian Levy stable law, whose PDF decays with power law tails of index $0 < \nu < 2$:

$$p(x) \underset{|x|\to\infty}{\sim} \frac{1}{x^{1+\nu}}\,,$$

with $X = \log(S_{t+\Delta t}/S_t)$.

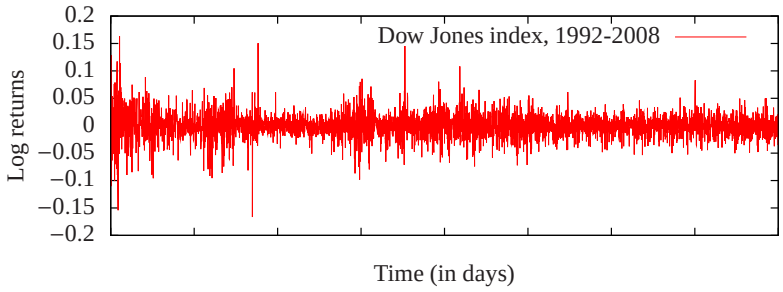



Figure 1.2: Daily log-returns for the Dow-Jones index during the period from 1992 to 2008.

With this choice, as confirmed by the example reported in Fig. 1.1, these distributions capture the shape of the central region of the data, even though they tend to overweight the tails. Since $\nu < 2$ a major consequence of Mandelbrot's assumptions is that returns would have infinite variance, posing serious interpretation problems when dealing with real data (from finite size samples). A way out of this drawback is represented by the Truncated Lévy Flights [20], introduced by R. N. Mantegna, which are processes characterized by a non Gaussian Lévy stable distribution in the central region, but whose tails are truncated to zero or feature a smooth exponential cutoff [21]. Nevertheless, the power-law decay of return distributions was confirmed by more recent studies [22], showing for stocks of individual companies that the distributions decay in agreement with a tail index $2.5 < \nu < 4$ (well outside the Lévy stable regime). In particular, this kind of decay can be well reproduced by means of Student's $t$-distributions [16, 23, 24, 17, 25], which feature power-law tails while having finite variance whenever the number of degree of freedom is greater than 2.

## 1.3 The random nature of the volatility

As discussed above, empirical return distributions decay much more slowly than expected for the GBM model (1.3); the observed fluctuations in the second moment of the distribution or, equivalently, the non constant behavior of the volatility $\sigma$, represent the other prominent evidence which can not be accounted for by such a simple model. This effect can be seen clearly by looking at the typical behavior of a financial series, as in Fig. 1.2, which shows the non constant nature of the amplitude of price fluctuations (referred to as *heteroskedasticity*). More precisely, we can see that the volatility tends to concentrate and then to "explode" as time goes by, and the periodic alternation of regimes of high and low volatility is known as *persistence*.

Before addressing this issue in more detail, let us discuss the more common *proxies* by which we can access the volatility from return data. Indeed, unlike prices, volatilities are not directly observed on the market, and suitable defini-

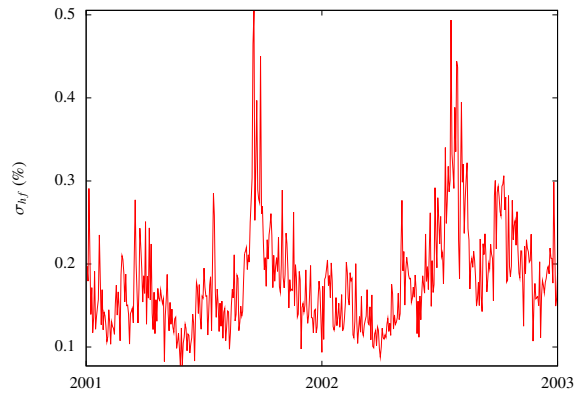



Figure 1.3: Time evolution of the high-frequency volatility $\sigma_{hf}$ for the Spanish company Telefonica from 2001 to 2003.

tions in term of the former are needed. Inspired by the GBM model, for which the variance of log-returns at the scale $\Delta t$ is simply $\sigma^2 \Delta t$, the most natural definition of the volatility is that of standard deviation of the sample. Given the price time series $\mathcal{S} = \{S_0, S_1, S_2, \dots\}$, if we indicate as $L_k = \log(S_{t_{k+1}}/S_{t_k})$ the $k$-th return of the sample $\mathcal{R} = \{L_0, L_1, L_2, L_3, \dots, L_{N-1}\}$, the *historical volatility* $\sigma_{\text{hist}}$ is defined by

$$\sigma^2_{\text{hist}} = \frac{1}{(N-1)\Delta t} \sum_{k=0}^{N-1} (L_k - \hat{\mu}_L)^2 \ , \tag{1.7}$$

where $\hat{\mu}_L = \sum_{k=0}^{N-1} L_k/N$ stays for the log-return sample mean. A slightly different definition is the one given by Bouchaud in [16] (see also [26, 27]) of *high frequency volatility* as the local average of absolute returns $R_k = (S_{t_{k+1}} - S_{t_k})/S_{t_k}$ over small time scales ($t_{k+1} - t_k = 5$ minutes)

$$\sigma_{\text{hf}} = \frac{1}{N_d} \sum_{1}^{N_d} |R_k(5')| \ , \tag{1.8}$$

where $N_d$ is the number of 5 minutes intervals within a trading day.

In a more general way than Eq. (1.3), zero mean returns $dX_t$ could be modeled with a stochastic equation of the form

$$dX_t = \frac{dS_t}{S_t} - \left\langle \frac{dS_t}{S_t} \right\rangle = \sigma_t \, dW_t \, , \tag{1.9}$$

accounting for a non constant volatility. This way, returns and volatilities are related by the average relation $\langle \sigma_t \rangle = \sqrt{\langle dX_t^2 \rangle / dt}$. In empirical analysis, when the time resolution $\Delta t$ is small enough, we can approximate $dt \approx \Delta t$ and $dX_t$ with $\Delta X_t$ and, even though not justified from the analytical point of view, we could disregard the average value, and adopt as a definition of *instantaneous*

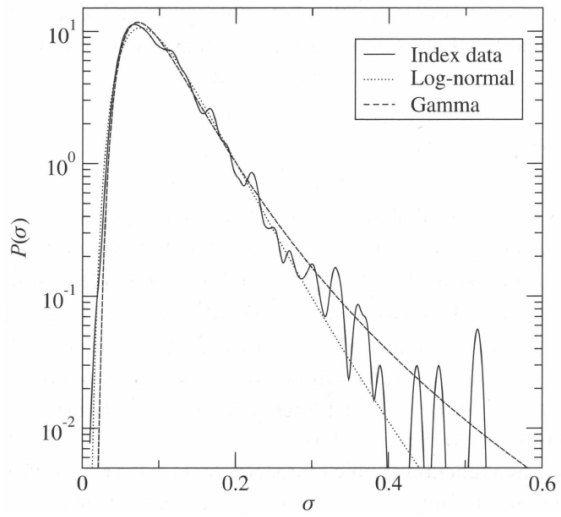



Figure 1.4: Distribution of $\sigma_{hf}$ for the S&P500 index during the period 1990-2001, fitted with a Log-Normal or an Inverse Gamma distribution. From [16].

*volatility* [28] the relation

$$\sigma_{\text{inst},t} = \sqrt{\frac{\Delta X_t^2}{\Delta t}}\,. \tag{1.10}$$

No matter what the chosen proxy is, the volatility is not constant in time and its evolution closely resembles the path of a random variable, as illustrated in the Fig. 1.3. If we take for serious the idea of modeling the volatility as a random quantity, we could ask ourselves which its distribution is. Independent analysis [16, 29] clearly show that a good fit of the volatility for small time scales can be obtained with a Log-Normal or an Inverse Gamma distribution. For the case of the Standard & Poor 500 (S&P500) index, the two fits of the high frequency volatility (1.8) are compared in Fig. 1.4; even though the results are comparably good, the Inverse Gamma law better reproduces extreme realizations. It is interesting to notice that the emerging of an inverse gamma distribution can be motivated. Indeed, it is reasonable to postulate that the past activity, as measured for instance by the absolute value of close to close price differences, influences the future volatility and the Inverse Gamma distribution turns out to be the equilibrium distribution of the volatility for the simplest case of a model incorporating this feedback effect (see Chapter 20 in [16]).

## 1.4 Correlations in financial data

Besides the distributions of returns or the volatility, several studies have been devoted to quantitatively asses the nature of time correlations in financial data. In most empirical studies, returns time series are analyzed as if they were made

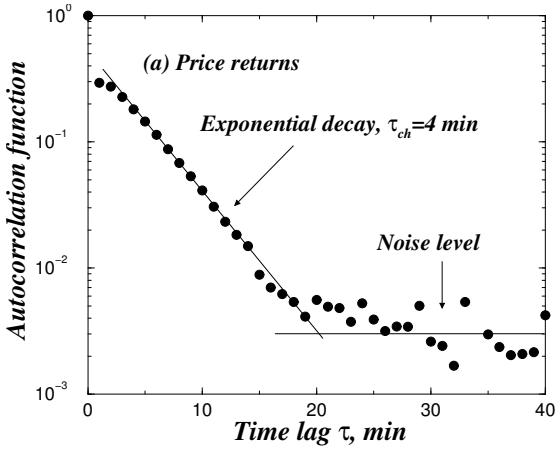



Figure 1.5: Log-linear plot of the autocorrelation function for the S&P500 index log-returns sampled at the scale $\Delta t = 1$ min. From [30].

of independent random variables. The pairwise correlation between returns can be investigated by looking at the return autocorrelation function

$$\mathcal{C}_{\Delta t}(\tau) = \frac{\langle L_t L_{t+\tau}\rangle - \langle L_t^2\rangle}{\langle L_t^2\rangle - \langle L_t\rangle^2} \ ,$$

where the superscript $\Delta t$ stands for the time lag the return is referred to, that is $L_t = \log(S_{t+\Delta t}/S_t)$, and it will be frequently omitted if not necessary. As a matter of fact, $\mathcal{C}_{\Delta t}(t)$ decays exponentially and very rapidly, vanishing with time constants of the order of some minutes for liquid markets [31, 32, 30] as shown in Fig .1.5. The *return-volatility correlation* and the *volatility autocorrelation* are more interesting. Taking into account that the square of the return $\mathrm{d}X_t^2$ is proportional to $\sigma_t^2$, these quantities can be empirically estimated through the following correlation functions [28, 33, 34]:

$$\mathcal{L}_{\Delta t}(\tau) = \frac{\left\langle dX_t dX_{t+\tau}^2\right\rangle}{\left\langle dX_t^2\right\rangle^2} \tag{1.11}$$

$$\mathcal{A}_{\Delta t}(\tau) = \frac{\left\langle dX_t^2 dX_{t+\tau}^2\right\rangle - \left\langle dX_t^2\right\rangle \left\langle dX_{t+\tau}^2\right\rangle}{\mathrm{Var}\left[dX_t^2\right]} \ . \tag{1.12}$$

As for the return-volatility correlation, it is found to be compatible with zero for $\tau < 0$, negative and exponentially decaying [16] for positive values of the correlation time $\tau$:

$$\mathcal{L}_{empirical}(\tau) = -H(\tau)\ A\ \mathrm{e}^{-\frac{\tau}{\tau^{\mathcal{L}}}} \ ,$$

where $H$ stands for the Heaviside step function. However, the amplitude of the correlation $A$ and the decay time $\tau^{\mathcal{L}}$ are found to be rather different for single stocks and indexes. In Fig. 1.6 it is reported the exponential fit of the S&P 500 data from 1970 to 2010, corresponding to $A = 37.5$ and $\tau^{\mathcal{L}} = 15.1$ days

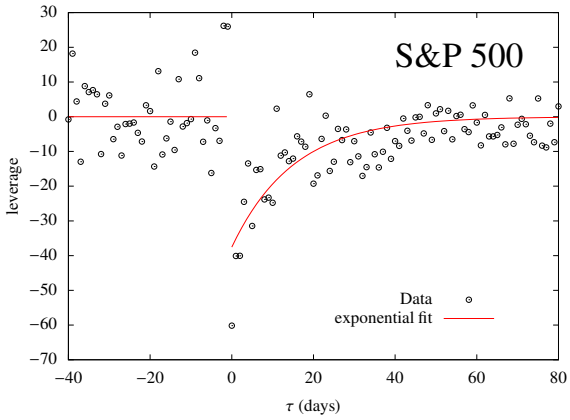



Figure 1.6: Return-volatility correlation estimated from the daily returns of the S&P500 index during the period 1970-2010.

for a time resolution of $\Delta t = 1$ day. The same kind of analysis carried out by averaging over 437 U.S. stock [16] shows that the amplitude is much smaller for single stocks, $A \approx 1.9$, and the correlation time much longer, $\tau^{\mathcal{L}} \approx 69$ days. The correlation (1.11) is often referred to as "leverage" correlation, since it can be associated with the *leverage effect*, the financial effect by which price drops increase the subsequent volatility. Indeed, as discussed before, correlations exist between past returns and future volatilities, but the economic underlying mechanism is still debated. For the case of single stocks, the small amplitude of the correlation may indicate a simple retarding effect [16]: indeed, while price variations are independent on the price, the volatility in the representation (1.9) is proportional to the absolute value of the return

$$\sigma_t^2 \propto \mathrm{d}X_t^2 = \frac{\mathrm{d}S_t^2}{S_t^2} \ ,$$

so that past drops in the price are able to increase the future volatility since the price enters the denominator. On the other hand, for indexes, which are a combination of the main representative stocks in a sector of the market, the strength of the correlation suggests the presence of a panic effect in which down moves of the market as a whole increase the uncertainty for the future and they trigger a rise of the activity, thus making the market sector more volatile.

Even though pairwise correlation between price changes decays very rapidly, it does not imply that price changes are independent random variables. Indeed, various studies showed that non linear functions of those variables, for instance their absolute values or their squares, have much longer memory. The volatility autocorrelation (1.12) measures how much the volatility at time $t+\tau$ differs from the value it had at time $t$. Unlike the leverage function, the empirical volatility-volatility correlation can not be described in terms of a single characteristic time. Fig. 1.7 shows that the two regions of $\mathcal{A}(\tau)$ for small and

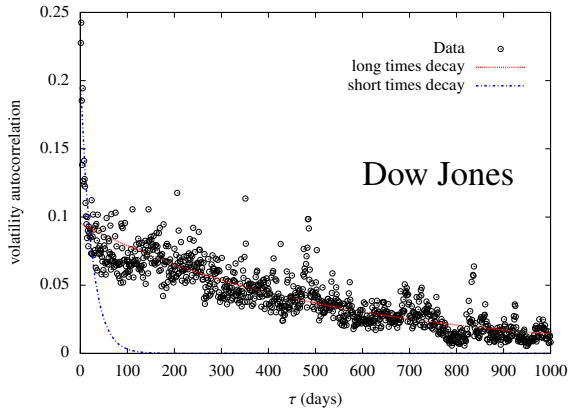



Figure 1.7: Daily volatility autocorrelation for the Dow Jones index from 1928 to 2003, along with two exponential fits, considering the short and the long $\tau$ region respectively. They show that $\mathcal{A}(\tau)$ exhibits at least two characteristic times and that it is a long range correlation.

large $\tau$ are nicely reproduced by an exponential fit if considered separately. This shows that $\mathcal{A}(\tau)$ has at least two characteristic times, one of the order of ten days (comparable with the decay time of the leverage correlation) while the other is much longer, approximately 500 days for the Dow-Jones index. The multiple time scale nature of the volatility autocorrelation is quite accepted, but the functional form of the decay is still under investigation. Past studies [35, 36, 37, 31, 27, 38] found evidence that the decay is of power law type, but, as remarked in [16], it is not trivial to discriminate between a real power law decay or, for instance, a superposition of exponentials with different characteristic times and adjustable amplitudes plus an additive constant. The presence of long term memory in the square of price changes, not showing in the price changes themselves, is a strong indication of the fact they are not the only relevant process. Another stochastic process may drive price variations and it is what we call the volatility.

Chapter 2

# Multiplicative noise diffusion process

Many different physical phenomena exhibit a complex behavior characterized by long range correlations, memory effects and the emergence of non Gaussian, most often power law, distributions. These systems are usually contrasted with classical statistical systems of free massive particles, showing the Gaussian distribution $e^{-p^2/2mT}$ in the momentum $p$, which are considered to be paradigmatic of equilibrium, uncorrelated or short range correlated systems. A microscopic description of the Maxwell Boltzmann statistics is provided by the Langevin equation [12, 39]

$$\dot{x} + \gamma x = \xi \,,$$

where $x$ represents the statistical variable of interest (the momentum for a system of free particles) and, as introduced originally [40], $\gamma$ is a constant damping coefficient while $\xi$ is the additive stochastic noise term. Recently it has been realized [41] that the Gaussian paradigm is not the most general description of the equilibrium state and that power law tails in the equilibrium distribution stem naturally when assuming also the damping coefficient $\gamma$ to be stochastic. From a macroscopic point of view, this superposition of an additive and a multiplicative noise leads to the following Fokker-Planck (FP) equation [41] with a quadratic diffusion coefficient

$$\frac{\partial p}{\partial t}(x,t) = -\frac{\partial}{\partial x}\left[(ax+b)p(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[(cx^2+dx+e)p(x,t)\right] \,, \qquad (2.1)$$

where $a, b, c, d, e$ are real constants. Until recently an analytical characterization of the process (2.1) was available only at the stationary state [41], while at finite times it could be accomplished only as a formal expansion on the eigenfunctions of the FP operator [42]. On the other hand, in [9] a finite time exact analysis has been carried out in terms of the moments of the associated PDF. A quite interesting point is that the dynamics (2.1), generalized for possibly time

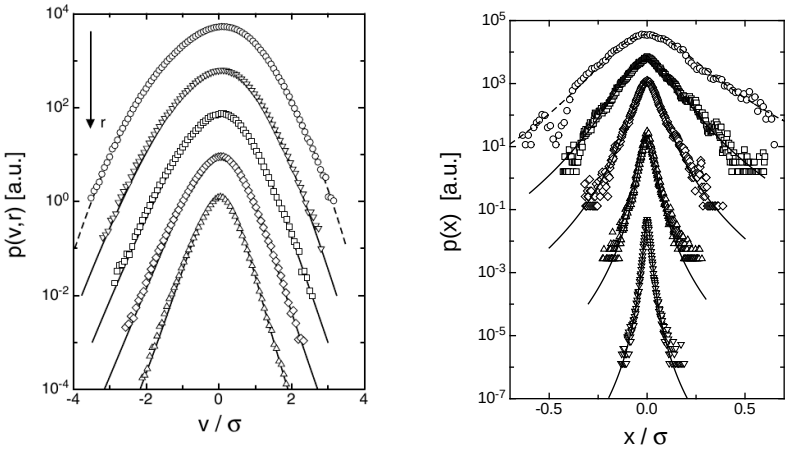



Figure 2.1: Left panel: PDF of longitudinal velocity variations in a turbulent fluid for different length scales, from [43]. Right panel: PDF of price changes in the U.S dollar - German mark exchange rate for different time delays, from [44]. In both the examples, real data are well reproduced by the PDF corresponding to specific instances of the dynamics (2.2).

dependent coefficients, emerges in the study of a large and hetereogeneus set of physical phenomena, from turbulent velocity flows [45], power law spectra in $e^+e^-$, $p\bar{p}$ and heavy ion collisions [46], anomalous diffusion processes [47], non stationary Markov processes [48] with Hurst exponent $H \neq 1/2$, to heartbeat interval fluctuations [49]. Surprisingly, it can also describe exchange rate variations in foreign exchange markets [50, 44] and stock price variations in option markets [51, 52].

In this chapter, an extended and detailed version of the results published in [9] is provided. We will introduce and analyze the microscopic dynamics associated to Eq. (2.1) and we will generalize it for a time dependence both in the drift and the diffusion coefficients. We will carry out an analytical characterization of the considered process from the point of view of the moments of the associated PDF, detailing an original algorithmic solution for their exact computation at all orders and for all times. We will see how this analytical information allows to gain insight into the relaxation modes of the process, possibly but not necessarily toward a stationary regime. More precisely, the exact, time dependent expressions of the moments will allow to identify sets of parameters for which all of them do converge to a stationary level, and different sets for which higher order moments diverge asymptotically. In the latter case, when a stationary distribution exists, we will guess that its PDF has power law tails and the tail exponent predicted by our approach will match the results of different analysis based on the solution of the stationary FP equation. Most of all, the analytical information provided by our approach will turn out to be crucial every time the process does not admit a stationary limit, for which case no conclusion can be drawn, in general, from the analysis of the FP equation.

In particular, a simple connection between the time dependence in the microscopic equation and the time scaling of the moments will emerge, allowing to account for nearly arbitrary scaling laws.

We will devote large space to this topic since, as explained above, the considered process has universality character in the study of complex systems. However, in the scope of this thesis the main connection between the material presented here and the financial modeling and analysis will be manifest only in the next chapter dedicated to Stochastic Volatility Models. There we will realize that the stochastic process which drives the price volatility is modeled most often with a dynamics which is a specific, mean reverting instance of the more general process analyzed in this chapter.

## 2.1 The microscopic model

At the microscopic level, the process introduced above is summarized in the following SDE[1]

$$dX_t = \frac{a\,X_t + b}{g(t)}\,dt + \sqrt{\frac{c\,X_t^2 + d\,X_t + e(t)}{g(t)}}\,dW_t\,, \tag{2.2}$$

with initial time condition $X_{t_0} = X_0$, $t_0 \in D \subseteq [0, t_{\rm lim})$ with $t_{\rm lim}$ possibly $+\infty$; $W_t$ is the standard Brownian motion, $a$, $b$, $c$, $d$ are real constants, $1/g(t)$ and $e(t)$ are non negative (deterministic) smooth functions of the time over $D$. For the diffusion coefficient to be meaningful, it has to satisfy $d^2 - 4ce(t) \leq 0$ with $c \geq 0$. Due to the correspondence between SDEs and FP equation (see [39]), the general dynamics (2.2) exactly corresponds to the macroscopic equation (2.1) when $g(t)$ is a constant (time homogeneous process), and generalizes it introducing a time inhomogeneity otherwise, possibly through a time dependence in the coefficient $e(t)$. Application of the Itô Lemma to the function $f(X_t) = X_t^n$ leads to the following integral relation

$$\begin{aligned} X_t^n =& X_0^n + \int_{t_0}^{t} \frac{X_s^{n-2}}{g(s)} \left[F_n X_s^2 + A_n X_s + B_n(s)\right]\,ds \\ &+ n \int_{t_0}^{t} \frac{X_s^{n-1}}{\sqrt{g(s)}} \sqrt{cX_s^2 + dX_s + e(s)}\,dW_s\,, \end{aligned}$$

whose expectation readily provides the ordinary differential equation satisfied by the $n$-th order moment $\mu_n(t) = \langle X_t^n \rangle$ for $n \geq 1$

$$g(t)\frac{\mathrm{d}}{dt}\mu_n(t) = F_n\mu_n(t) + A_n\mu_{n-1}(t) + B_n(t)\mu_{n-2}(t)\,, \tag{2.3}$$

with boundary condition $\mu_n(t_0) = \langle X_0^n \rangle$. The coefficients read $F_n = na + \frac{1}{2}n(n-1)c$, $A_n = nb + \frac{1}{2}n(n-1)d$, $B_n(t) = \frac{1}{2}n(n-1)e(t)$, and we assume

[1] Here and then we assume the Îto prescription for the interpretation of SDEs.

$\mu_0(t) = 1$. When the starting PDF $p(x, t_0)$ has finite moments at all orders, the smoothness of $1/g(t)$ and $e(t)$ ensures the existence of a unique solution $\mu_n(t)$ over an arbitrary interval $D' = [t_i, t_f] \subseteq D$ with $t_0 \in D'$. After introducing a "rescaled" time $\tau(t) = \int_{t_0}^{t} 1/g(s)ds$, which is a monotonously increasing function of $t$ due to the hypothesis on $g(t)$, the solution of Eq. (2.3) reads

$$\mu_n(t) = \mathrm{e}^{F_n \tau(t)} \left[ \langle X_0^n \rangle + \int_0^{\tau(t)} e^{-F_n \tau_1} A_n \tilde{\mu}_{n-1}(\tau_1)\, d\tau_1 + \int_0^{\tau(t)} e^{-F_n \tau_1} \tilde{B}_n(\tau_1) \tilde{\mu}_{n-2}(\tau_1)\, d\tau_1 \right], \tag{2.4}$$

where $\tilde{\mu}_n(\tau) = \mu_n(t(\tau))$ and $\tilde{B}_n(\tau) = B_n(t(\tau))$. Now we will show that the previous expression lends itself to an expansion over $\left\langle X_0^{n-j} \right\rangle$, for $j = 0, \ldots, n$ by iteratively substituting the moments entering the *r.h.s.* with their closed-form solutions. For sake of simplicity, we will detail the procedure for the case $g(t) = 1$ and constant $e(t) = e \geq 0$.

## 2.2 Moments: exact solution

Starting form the implicit solution (2.4), we can replace $\mu_{n-1}$ and $\mu_{n-2}$ with their expressions involving $\mu_{n-2}$, $\mu_{n-3}$ and $\mu_{n-4}$ obtaining

$$\begin{aligned}
\mu_n(t) = {} & \langle X_0^n \rangle\, e^{F_n t} \\
& + e^{F_n t} \left\{ A_n \left\langle X_0^{n-1} \right\rangle \int_0^t e^{-(F_n - F_{n-1})t_1}\, dt_1 + B_n \left\langle X_0^{n-2} \right\rangle \int_0^t e^{-(F_n - F_{n-2})t_1}\, dt_2 \right. \\
& + A_n A_{n-1} \int_0^t e^{-(F_n - F_{n-1})t_1} \int_0^{t_1} e^{-F_{n-1} t_2} \mu_{n-2}(t_2)\, dt_1\, dt_2 \\
& + B_n B_{n-2} \int_0^t e^{-(F_n - F_{n-2})t_1} \int_0^{t_1} e^{-F_{n-2} t_2} \mu_{n-4}(t_2)\, dt_1\, dt_2 \\
& + A_n B_{n-1} \int_0^t e^{-(F_n - F_{n-1})t_1} \int_0^{t_1} e^{-F_{n-1} t_2} \mu_{n-3}(t_2)\, dt_1\, dt_2 \\
& \left. + B_n A_{n-2} \int_0^t e^{-(F_n - F_{n-2})t_1} \int_0^{t_1} e^{-F_{n-2} t_2} \mu_{n-3}(t_2)\, dt_1\, dt_2 \right\}.
\end{aligned}$$

This procedure can be iterated, keeping on to replace the moments entering the *r.h.s.* with their expressions involving lower order moments, until you arrive to the explicit expressions of $\mu_0 = 1$. In this way, by direct inspection it can be recognized that the expression of $\mu_n(t)$ reduces to an expansion on the moments at $t_0$

$$\mu_n(t) = e^{F_n t} \left\{ \sum_{j=0}^{n} \left\langle X_0^{n-j} \right\rangle \left[ \sum_{N_A + 2N_B = j} \sum_{\Pi_{N_A N_B}} \Delta_n(\pi_{N_A N_B}; t) \right] \right\}. \tag{2.5}$$

In a naive way, the contribution $\Delta_n(N_A, N_B, \pi_{N_A N_B}; t)$ can be associated to an oriented string of two types of "knots", let us name them type-$A$ and type-$B$ knots. Every knot has a subscript $k$ and the move from one knot to the next reduces progressively the value of $k$, keeping to the following rule:

$$A_k \begin{array}{l} \nearrow A_{k-1} \\ \searrow B_{k-1} \end{array} \qquad B_k \begin{array}{l} \nearrow A_{k-2} \\ \searrow B_{k-2} \end{array} \tag{2.6}$$

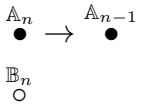
The number of type-$A$ knots, $N_A$, and of type-$B$ knots, $N_B$, must satisfy the constraint $N_A + 2N_B = j$ and the order of the first knot must always be equal to $n$. Every knot $A$ or $B$ gives a contribution equal to

$$\begin{aligned} \mathbb{A}_k &= A_k \int_0^t e^{-(F_k - F_{k-1})s}\, ds \\ \mathbb{B}_k &= B_k \int_0^t e^{-(F_k - F_{k-2})s}\, ds\,, \end{aligned} \tag{2.7}$$

and the integrals corresponding to "higher order" $(k)$ knots are nested into the ones corresponding to lower values of $k$. For instance, the pattern of two knots $\mathbb{B}_k \mathbb{A}_{k-2}$ brings a contribution[2] equal to

$$\mathbb{B}_k \mathbb{A}_{k-2} = \int_0^t e^{-(F_k - F_{k-2})t_1} \int_0^{t_1} e^{-(F_{k-2} - F_{k-3})t_2}\, dt_1\, dt_2$$

Given $n$, the coefficient of $\left\langle X_0^{n-j} \right\rangle$ is obtained considering all the patterns satisfying the above rules, that is considering the set $\Pi_{N_A N_B}$ of permutations without repetitions of $N_A$ type $A$ elements and $N_B$ type $B$ elements, picking up the strings associated to every element $\pi_{N_A N_B} \in \Pi_{N_A N_B}$ and then adding up the corresponding contributions; for example, when $j = 2$ the following patterns have to be taken into account

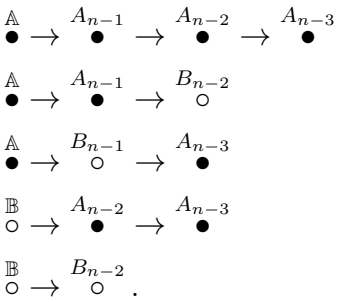



corresponding to a numerical contribution equal to

$$A_n A_{n-1} \int_0^t e^{-(F_n - F_{n-1})t_1} \int_0^{t_1} e^{-(F_{n-1} - F_{n-2})t_2}\, dt_1\, dt_2 + B_n \int_0^t e^{-(F_n - F_{n-2})t_1}\, dt_1\,,$$

[2]With a slight abuse of notation, we identify the abstract object "knot" with its numerical contribution to the computation of the $\Delta$ in the expansion (2.5).

while for $j = 4$ the diagrams to be considered are the following five:



The special case $j = 0$ corresponds to the null pattern, without any knot, and we assume the corresponding contribution to be equal to 1.

We now turn to the case of constant coefficients in the model (2.2) with possibly non constant $g(t)$. The solution of Eq. (2.3) reads

$$\langle X^n \rangle_t = e^{F_n \int_{t_0}^t \frac{1}{g(t')} dt'} \left\{ \langle X_0^n \rangle + \int_{t_0}^t \frac{e^{-F_n \int_{t_0}^{t_1} \frac{1}{g(t')} dt'}}{g(t_1)} \left[ A_n \mu_{n-1}(t_1) + B_n \mu_{n-2}(t_1) \right] dt_1 \right\} ;$$

previous considerations extend to this case and the knots' contributions become

$$A_k \int_{t_0}^{t_i} \frac{e^{-(F_k - F_{k-1}) \int_{t_0}^{t_{i+1}} \frac{1}{g(t')} dt'}}{g(t_{i+1})} \int_{t_0}^{t_{i+1}} \dots \, dt_{i+2} \, dt_{i+1} \dots$$
$$B_k \int_{t_0}^{t_i} \frac{e^{-(F_k - F_{k-2}) \int_{t_0}^{t_{i+1}} \frac{1}{g(t')} dt'}}{g(t_{i+1})} \int_{t_0}^{t_{i+1}} \dots \, dt_{i+2} \, dt_{i+1} \dots$$

After changing to the rescaled time variable $\tau$, we recover the more familiar expressions

$$A_k \int_0^{\tau_i} e^{-(F_k - F_{k-1})\tau_{i+1}} \int_0^{\tau_{i+1}} \dots d\tau_{i+2} \, , d\tau_{i+1} \dots$$
$$B_k \int_0^{\tau_i} e^{-(F_k - F_{k-2})\tau_{i+1}} \int_0^{\tau_{i+1}} \dots d\tau_{i+2} \, , d\tau_{i+1} \dots$$

and taking into account that the internal $\tau_i$ are mute variables, we guess that the expansion (2.5) still applies with the substitution $t \to \tau(t) = \int_{t_0}^t \frac{1}{g(t')} dt'$

$$\mu_n(t) = e^{F_n \tau(t)} \left\{ \sum_{j=0}^n \left\langle X_0^{n-j} \right\rangle \left[ \sum_{N_A + 2N_B = j} \sum_{\Pi_{N_A N_B}} \Delta_n(\pi_{N_A N_B}; \tau(t)) \right] \right\} . \quad (2.8)$$

With the same procedure we can deal with the more general case of a time dependent $e(t)$. For instance, in [44] it has been shown that U.S. dollar - German mark exchange rate variations over different time scales can be regarded as a Markov process described by Eq. (2.2) with $g(t) = 1$, $e(t) = e + \tilde{e} \exp(\varepsilon t)$ and constant $a, b, c, d, \tilde{e}$. Walking through the previous steps, it is easy to convince

ourselves that this exponential time dependence makes every type-$B$ knot split into the sum of the two following contributions

$$\mathbb{B}_k = B_k \int_0^\tau e^{-(F_k - F_{k-2})\tau'} d\tau' \quad \text{and} \quad \tilde{\mathbb{B}}_k = \tilde{B}_k \int_0^\tau e^{-(F_k - F_{k-2} - \varepsilon)\tau'} d\tau' \ ,$$

where it has been defined $\tilde{B}_k = \frac{1}{2}k(k-1)\tilde{e}\exp(\varepsilon t_0)$. For example, the pattern $\mathbb{ABAB}$

$$\overset{\mathbb{A}_n}{\bullet} \to \overset{\mathbb{B}_{n-1}}{\circ} \to \overset{\mathbb{A}_{n-3}}{\bullet} \to \overset{\mathbb{B}_{n-4}}{\circ} \ ,$$

is now supplemented by the strings

$$\overset{\mathbb{A}_n}{\bullet} \to \overset{\tilde{B}_{n-1}}{\circ} \to \overset{\mathbb{A}_{n-3}}{\bullet} \to \overset{\mathbb{B}_{n-4}}{\circ}$$

$$\overset{\mathbb{A}_n}{\bullet} \to \overset{\mathbb{B}_{n-1}}{\circ} \to \overset{\mathbb{A}_{n-3}}{\bullet} \to \overset{\tilde{B}_{n-4}}{\circ}$$

$$\overset{\mathbb{A}_n}{\bullet} \to \overset{\tilde{B}_{n-1}}{\circ} \to \overset{\mathbb{A}_{n-3}}{\bullet} \to \overset{\tilde{B}_{n-4}}{\circ}$$

plus all their possible permutations, and the two sums in expansion (2.8) coherently generalize to

$$\sum_{N_A + 2(N_B + N_{\tilde{B}}) = j} \ \sum_{\Pi_{N_A N_B N_{\tilde{B}}}} \Delta_n(\pi_{N_A N_B N_{\tilde{B}}}, t - t_0) \ .$$

## 2.3 String amplitude computation

A careful analysis of the quantities $\Delta_n(\pi_{N_A N_B}, \tau(t))$ shows that they can always be computed analytically by means of the algorithmic strategy we detail below.

The total number of knots $N = N_A + N_B$ of a string is equal to the number of nested integrations and the string numerical contribution always has the form

$$I(\tau; c_1, \ldots, c_N) = \int_0^\tau e^{c_1 \tau_1} \int_0^{\tau_1} e^{c_2 \tau_2} \ldots \int_0^{\tau_{N-1}} e^{c_N \tau_N} d\tau_N \ldots d\tau_2 d\tau_1 \qquad (2.9)$$

where, in the notation of the previous section, $c_i = -(F_k - F_{k-1})$ or $c_i = -(F_{k'} - F_{k'-2})$ for some $k, k'$. Let us suppose [3] that $c_i \neq 0$ for every $i$; the innermost integration amounts to the sum of a constant and an exponential with the coefficient $c_N$ corresponding to the current integration index $(N)$

$$-\frac{1}{c_N}(1 - e^{c_N \tau_{N-1}}) \ . \qquad (2.10)$$

In this sense every integration produces a "splitting" in two terms. This combination is weighted (apart from the sign) with the inverse of the exponential

[3] It is worth noticing that due to the definition of $F_k$, only one coefficient $c_n$ may vanish, which occurs when the ration $|a|/c$ is integer or semi-integer. This special case requires nested integrations by parts, but it is otherwise dealt with in a similar manner.

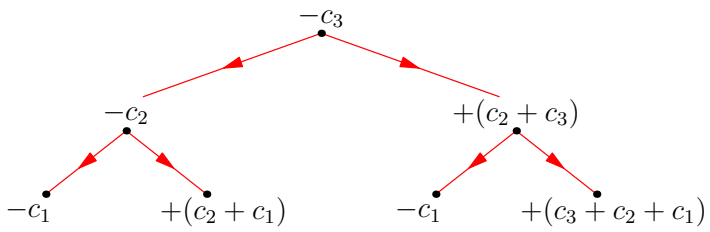



Figure 2.2: Binomial tree corresponding to $N = 3$ integrations; its output is the result (2.11).

coefficient. When this quantity is plugged into the second integral, $-1/c_N$ becomes an overall coefficient, the integration of the constant and the exponential terms give respectively

$$1 \to -\frac{1}{c_{N-1}}(1 - e^{c_{N-1}\tau_{N-2}})$$
$$\exp \to -\frac{1}{c_N + c_{N-1}}(1 - e^{(c_N+c_{N-1})\tau_{N-2}})$$

that is, the constant term gives back the basic combination (2.10) with the coefficient $c_i$ linked to the integration index $i$ ($i = N - 1$ here), while the exponential term gives the basic combination with the coefficient given by the sum of $c_i$ and the coefficient of the previous integration. The combinatorics of this cascade integrations can be visualized by means of a binomial tree, in which every level of knots (top down) corresponds to a different integration in (2.9), from the innermost to the outermost. The absolute value of the label attached to every knot is the coefficient involved in the corresponding integration, while its sign is the overall sign of the basic combination produced by the integrations performed up to the current level. At the end, we have $2^{N-1}$ basic combinations corresponding to the $2^{N-1}$ branches of the tree. The sign of a given basic combination is obtained by multiplying the signs of all the nodes of the branch, and the its characteristic coefficient is the one attached to the final node. As an example, let's consider the case $N = 3$. It is easy to verify that applying the above rules to the binomial tree of Fig. (2.2), we obtain the same result following from the explicit calculation of the expression (2.9)

$$\begin{aligned} I = &- \frac{1}{c_3}\frac{1}{c_2}\frac{1}{c_1}(1 - e^{c_1\tau}) + \frac{1}{c_3}\frac{1}{c_2}\frac{1}{c_1 + c_2}(1 - e^{(c_1+c_2)\tau}) \\ &+ \frac{1}{c_3}\frac{1}{c_2 + c_3}\frac{1}{c_1}(1 - e^{c_1\tau}) - \frac{1}{c_1 + c_2 + c_3}\frac{1}{c_2 + c_3}\frac{1}{c_3}(1 - e^{(c_1+c_2+c_3)\tau}) \ . \end{aligned} \tag{2.11}$$

Equivalently, the analytical expression of $I$ can be obtained by setting up a *signs matrix* $S$ and a *coefficient matrix* $C$ as follows. Both of them are $N \times (2^{N-1})$ matrices, in which every row $i = 1, \ldots, N$ has a number of "phases"

$n_{ph}(i) = 2^{i-1}$ of length $l_{ph}(i, N) = 2^{N-i}$. The elements $(i, j)$ for fixed $i$ share the same value for all the $j$s in the same phase. The $S$ matrix elements can take values $\pm 1$ and for $p = 1, \ldots, n_{ph(i)}$ the common value is given by $(-1)^p$. The construction of the $C$ matrix is a bit more tricky. Considering a fixed row $i$, the common value is $C(i, j) = 1$ when the phase number $p$ is odd and $j$ is in the $p-$th phase, while it is equal to $C(i, j) = C(i-1, j) + 1$ if $p$ is even, with $C(1, j) = 1$ for every $j$ by definition. An explicit example of the $S$ and $C$ matrices is reported below for the case $N = 4$

$$S = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{bmatrix} \begin{matrix} i=1 \\ i=2 \\ i=3 \\ i=4 \end{matrix}$$

$$\Downarrow$$

$$s = \begin{bmatrix} + & - & - & + & - & + & + & - \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 & 1 & 1 & 3 & 3 \\ 1 & 2 & 1 & 3 & 1 & 2 & 1 & 4 \end{bmatrix} \begin{matrix} i=1 \\ i=2 \\ i=3 \\ i=4\,. \end{matrix}$$

Once $S$ has been computed, the signs of the $2^{N-1}$ basic combinations of the final result are obtained by multiplying the elements of $S$ column by column, thus obtaining a sign vector $s$. The coefficients $c(i, j)$ of the basic combinations produced up to the $i$-th integration can be recovered by reading the elements of the $C$ matrix, through the following formula:

$$c(i, j) = \sum_{k=N-(i-1)}^{N-i+C(i,j)} c_k \ .$$

The basic combinations have to be multiplied by the inverse of the characteristic coefficient corresponding to the $N-1$ previous integrations, so that the final analytical expression of $I$ is given by

$$I(\tau; c_1, \ldots, c_N) = \sum_{j=1}^{2^{N-1}} s(j) K(j) \left(1 - e^{c(N,j)\tau}\right) \ , \quad K(j) \doteq \prod_{i=1}^{N} \frac{1}{c(i, j)} \ . \quad (2.12)$$

## 2.4 Scaling and power law distributions

With the formulas (2.12) it is possible to implement a numerical algorithm[4] computing exactly all the contributions $\Delta(\pi_{N_A,N_B};\tau)$. From the previous expressions it is clear they are combinations of exponentials in the time variable $\tau$. It follows that Eq. (2.8) can be rewritten in the form

$$\mu_n(t) = \sum_{j=0}^{n} c_j^n \, e^{F_{n-j}\tau(t)} \,, \tag{2.13}$$

the $c_j^n$ being real possibly vanishing functions of the $A_k$'s, $B_k$'s, $F_k$'s and $\langle X_0^k \rangle$. The functional form of $g(t)$ determines the type of scaling emerging from the multiplicative noise process. For a constant $g(t) = 1$, as in [41], we have $\tau = t - t_0$ and the $n$-th order moment is characterized by the superposition of $n$ exponentials with time constants $\{1/|F_n|, \ldots, 1/|F_1|\}$. When $g(t) = t$, as in [53, 54], the moments exhibits a power law scaling

$$\mathrm{e}^{F_{n-j}\tau(t)} = t^{F_{n-j}} \; t_0^{-F_{n-j}} \;.$$

More generally, for $g(t) = t^\beta$ ($\beta \neq 1$) the time dependence turns out to be a stretched exponential with stretching exponent $1 - \beta$:

$$\mathrm{e}^{F_{n-j}\tau(t)} = \mathrm{e}^{F_{n-j}\frac{1}{1-\beta}\left(t^{1-\beta} - t_0^{1-\beta}\right)} \;.$$

Eq. (2.13) also allows us to gain insight into the nature of the stochastic process described by the model (2.2), both at finite $t$ and for $t$ approaching $t_{\lim}$. We limit the discussion to the case of constant coefficients and observe that in terms of the rescaled time $\tau(t)$ the PDF $\tilde{p}(x,\tau) \doteq p(x,t(\tau))$ satisfies the FP equation

$$\frac{\partial}{\partial\tau}\tilde{p}(x,\tau) = -\frac{\partial}{\partial x}\left[D_1(x)\tilde{p}(x,\tau)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[D_2(x)\tilde{p}(x,\tau)\right] \;, \tag{2.14}$$

with $D_1(x) = ax + b$ and $D_2(x) = cx^2 + dx + e$ and initial condition $\tilde{p}(x,0) = \tilde{p}_0(x)$. In general, from the analysis of the previous equation we can't draw any conclusion about the shape of $p_{st}(x)$. Nevertheless, whenever $\tau(t) \to \infty$ for $t \to t_{lim}$, the behavior of $\tilde{p}_{\mathrm{st}}(x)$ emerging from the analysis of Eq. (2.14) applies to $p(x,t)$ asymptotically. Indeed, $\tau$ being a smooth function of $t$ we have that $\lim_{t\to t_{\lim}^-} p(x,t) = \lim_{\tau\to+\infty} \tilde{p}(x,\tau) = \tilde{p}_{\mathrm{st}}(x)$. Following [39, 41], the stationary solution $\tilde{p}_{\mathrm{st}}(x)$ of Eq. (2.14) can be computed analytically in the form

$$\log\frac{\tilde{p}_{st}(x)}{N} = -\left(1 - \frac{a}{2c}\right)\log\frac{2D_2(x)}{e} - \frac{\alpha}{\vartheta}\arctan\left(\frac{\vartheta x}{e + \frac{d}{2}x}\right) \tag{2.15}$$

[4] All the numerical results presented in the following figures have been obtained with routines written in C++ implementing the diagrammatic technique developed here.

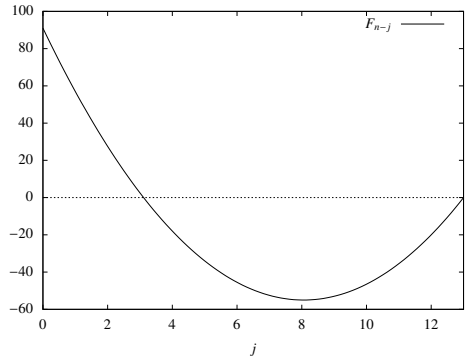



Figure 2.3: Typical shape of the functions $F_{n-j}$. The presence of positive values for a fixed $n$ determines the divergence of every $\mu_{n'}$ for $n \geq n'$.

where $N$ is a normalization factor, $\vartheta = \sqrt{c\,e - d^2/4}$ and $\alpha = a\,d/2c - b$. This solution provides evidence of the possible emergence of power law tails and Eq. (2.13) precisely characterizes the way the moments converge to the stationary level.

The previous results can be employed to characterize the stochastic process described by the SDE (2.2) for different choices of $a, b, c, d$, and $e$. Indeed, $F_{n-j}$ is a convex function of $j$ vanishing for $j = n$ and $j = n - n_1$, with $n_1 = 1 - 2a/c$, see Fig 2.3; it follows that if $F_n \leq 0$, all the contributions in the sum (2.13) do converge, otherwise $\mu_n(t)$ diverges for $t \to t_{lim}$. Moreover, if $F_n > 0$ then $F_{n'} > 0$ for $n' > n$ and the divergence of the $n$-th moment implies the divergence of the higher orders. Equivalently, if $n_1 < 0$ all the moments diverge while if $n_1 > 0$, all $\mu_n$'s for $n < n_1$ are convergent, otherwise not. In the following we detail the analysis of all possible cases, based on the solution (2.15) and on the signs of the coefficient $F_{n-j}$ in (2.13).

**Case $a > 0$ and $c > 0$.** All the moments diverge, since $F_n > 0\ \forall n > 0$. In order for the solution (2.15) to be properly normalized it has to be $0 < a < c/2$ and $e > 0$ and $p_{st}(x)$ is a generalized Student-$t$ with tail exponent $\nu = 1 + n_1$. For $a \geq c/2$ and finite $t$, all the moments are well-defined but no conclusions can be drawn about the exact form of the PDF. An example of the latter case is shown in Fig. 2.4, with all the $\mu_n$ diverging for large $t$.

**Case $a = 0$ and $e > 0$.** If $c = 0$, then also $d$ is 0 and Eq. (2.2) describes an ABM with time dependent coefficients. If $c > 0$, then $a_1 = F_1 = 0$ and $\mathbb{A}_1 = b\tau$, but integration by parts reveals that no moment converge, while the stationary solution is a power law with tail exponent $\nu = 2$.

**Case $a < 0$, $c > 0$, and $e > 0$.** $F_n > 0$ for $n \geq n_1$, thus only the first $n < n_1$ moments converge to a stationary level. The special case $a = -c/2$ implies $n_1 = 2$, $b_2 = F_2 = 0$ and $\mathbb{B}_2 = e\tau$ and previous conclusions are unchanged.

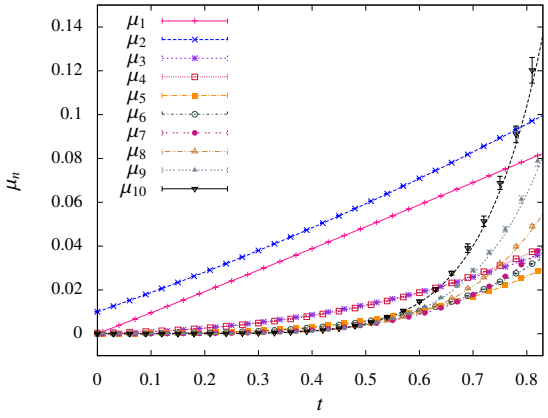



Figure 2.4: Scaling of the moments for $a = b = 9.5 \times 10^{-2}$ and $c = d = e = 8.3 \times 10^{-2}$; $\tilde{p}_0(x)$ is a zero mean Gaussian with $\langle X_0^2 \rangle = 0.01$. The Monte Carlo estimates of the $\mu_n$'s within 68% error bars are superimposed to the analytical curves, exhibiting full agreement.

Coherently the solution of the FP equation predicts the emergence of power law tails with $\nu = 1 + n_1$. In the left panel of Fig. 2.5 a case corresponding to $n_1 \simeq 9.9$ is shown for a stretched exponential time scaling, while the right panel corresponds to the case $g(t) = (t - t_{\rm lim})^2$, for a finite $t_{lim}$.

**Case $a \neq 0$, $c = 0$, and $e > 0$.** Eq. (2.2) describes an Ornstein-Uhlenbeck process [55]. $F_n$ becomes a linear function of $n$ and the moments reach a stationary value only if $a < 0$. For $a > 0$ the Gaussian PDF has time dependent unbounded mean and variance.

**Case $e = 0$ and $c > 0$.** As above the boundedness of the moments can be deduced from the value of $n_1$ and for $a < 0$ the stationary solution is an Inverted Gamma law with shape parameter $n_1 > 0$ and scale parameter $2|b|/c > 0$. If $b > 0$ the Inverted Gamma is defined for $x \in [0, +\infty)$, while for $b < 0$ the support is $(-\infty, 0]$. A similar situation occurs for $d - 4ce = 0$ and $d > 0$, $c > 0$, and $e > 0$, but the support boundary point corresponds to $-d/2c$.

## 2.5 A limit case

A different but interesting case is that corresponding to $g(t) = t$, $a = b = d = 0$, $e(t) = e\, t^\beta$, assumed in [51, 52] to model the dynamics of detrended log-returns $X_t = \log(S_t/S_0) - \langle \log(S_t/S_0) \rangle$. Indeed Eq. (2) in [51] exactly corresponds to those choices provided to express $c = (q-1)/[(2-q)(3-q)]$ and $\beta = 2/(3-q)$ in terms of the Tsallis entropic index $q$, and to impose

$$e = \sigma^2 \left[ c(2-q)(3-q) \right]^{\frac{q-1}{3-q}} .$$

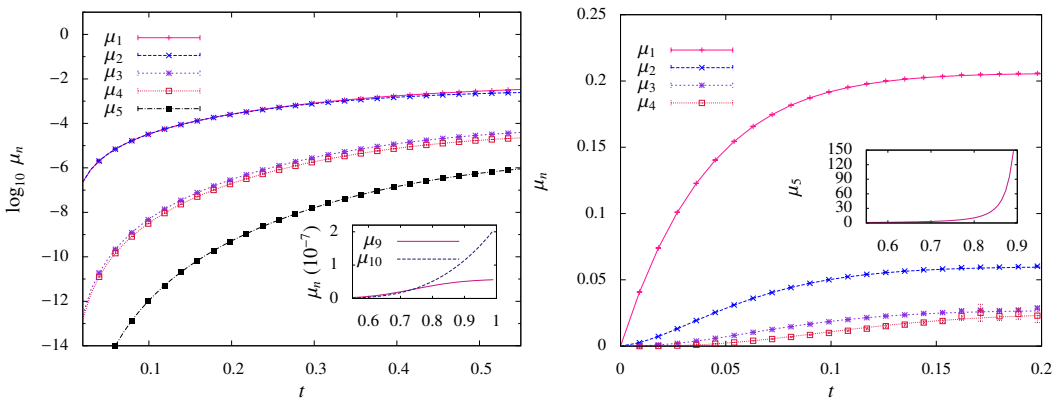



Figure 2.5: Left panel: lowest order moments for $a = -20$, $b = d = e = 0.1$, $c = 4.5$, with $g = t^\beta$, $\beta = 2$ and $\tilde{p}_0(x) = \delta(x)$. In the inset the last converging moment is compared to the first diverging one. Right panel: lowest order moments for $a = -24.3$, $b = 5$, $c = 12.2$, $d = e = 0.1$, for $g = (t - t_{lim})^2$ with $t_{lim} = 1$ and $\tilde{p}_0(x) = \delta(x)$. The first diverging moment $\mu_5$ is shown in the inset.

The quantities $\sigma^2$ and c are positive constants and the index $q$ necessarily belongs to $(1, 5/3)$ to ensure the existence of mean and variance [56]. The condition $b = d = 0$ implies that every type-$A$ knot is identically zero and so are all the odd moments, while the type-$B$ knot reads

$$\mathbb{B}_k = \frac{1}{2}k(k-1)e\ \mathrm{e}^{\frac{2}{3-q}\log t_0}\int_0^\tau \mathrm{e}^{(b_k+\frac{2}{3-q})\tau'}\mathrm{d}\tau' \ .$$

Moreover, if the initial condition is a Dirac delta centered in 0, we have $\langle X_0^n\rangle = 0$ and the only $j$ contributing to the expansion (2.8) is $j = n$. In this case, the expansion reduces to $\mathrm{e}^{F_n\tau(t)}\Delta_n(\pi_{0\frac{n}{2}}, \tau(t))$ with $\tau = \log t - \log t_0$ and $n = 2p$, $p > 1$ and the same analysis leading to (2.13) allows us to conclude that

$$\mu_{2p}(t) = \sum_{j=0}^{p} c_j^{2p} t^{F_{2(p-j)}+\frac{2j}{3-q}} t_0^{-\left[F_{2(p-j)}+\frac{2(j-p)}{3-q}\right]} \ ,$$

where the $c_j^{2p}$ are all non-vanishing constants. The authors in [51] assume the delta condition at $t_0 = 0$ but, from the point of view of our analysis, $t = 0$ is a singular point for $1/g(t)$ and a limit for $t_0 \to 0^+$ is mandatory in order to obtain sensible results. As a function of $j$, $F_{2(p-j)} + 2(j-p)/(3-q)$ is convex and for $j = p$ its value is zero, so that we only have to check the behavior for $j = 0$: if $F_{2p} + 2p/(q-3) > 0$ then $\lim_{t_0\to 0^+}\mu_{2p} = +\infty$, otherwise all the exponents of $t_0$ are non-negative and no divergence is possible. Recalling the expression of $c$, the former condition amount to $q > (2p+3)/(2p+1)$, which is precisely the one quoted in [51] to obtain a divergent $2p$-th order moment.

Chapter 3

# Modeling with stochastic volatility

In the previous chapter, we reviewed some evidences of the random nature of the volatility and the main correlations which can be investigated in the empirical analysis of financial data. In particular, the existence of long term memory effects in those correlations also supports the idea that the volatility itself can be considered a stochastic process as fundamental as that of price changes. These ideas can be formalized in a consistent way in terms of SDEs, giving rise to stochastic volatility models[1] (SVMs). They can be viewed as a continuous time version of a specific Generalized Autoregressive Conditional Heteroskedastic Model (see e.g. the introduction provided in [17]), but the continuous time framework makes it possible to exploit the power of Îto stochastic calculus. SVMs generalize the GBM dynamics assuming that the volatility itself can fluctuate randomly. The price SDE is coupled to a secondary Îto process and the volatility depends functionally on the latter. In this way, a second source of randomness enters the stochastic dynamics, and the correlation between the Brownian motions of the two SDEs reflects in correlations between the price and the volatility processes. This feature allows to reproduce the exponential decay of the leverage function and the existence of non trivial correlations between the volatility at different times. Moreover, a volatility which evolves stochastically generates non Gaussian tails in the return PDFs, accounting for the excess of kurtosis observed in the distribution of high frequency data.

In this chapter, we will review the general structure of a SVM focusing on the case in which the process driving the volatility follows an Ornstein-Uhlenbeck mean reverting process, and we will revise briefly some of the models most studied in the literature, especially the realistic exponential Ornstein-Uhlenbeck model. Then we will consider an important limit of this model, which is derived when the volatility is weakly random, and for which we will

[1] A more extensive introduction as well as advanced topics on stochastic volatility models can be found in [57].

provide an analytical characterization in terms of the process characteristic function and of the first moments of its probability density function. We will conclude with some recent results about a different linear model in which the volatility is driven by a specific mean reverting limit of the general diffusion process analyzed in Chapter 2 and whose stationary distribution is an Inverse Gamma law. We will characterize analytically the moments of the return PDF providing information about the power law decay of its tails, the return-volatility correlation and the volatility autocorrelation functions. We will also detail the empirical analysis of this model, estimating its parameters from the daily returns of the Standard & Poor 500 index and we will test to what extent the proposed dynamics is able to capture the scaling properties of the empirical return distribution.

## 3.1 SVMs: general framework

The models we are going to consider take the moves from the GBM dynamics (1.3), relaxing the assumption of a constant volatility and considering it to be, on the contrary, a Markov stochastic process itself [2]. This way, such a model can represented by the following system of coupled SDEs

$$dS_t = \mu S_t\, dt + \sigma(Y_t) S_t\, dW_{1,t} \tag{3.1}$$

$$dY_t = a(Y_t)\, dt + b(Y_t)\, dW_{2,t} \tag{3.2}$$

where we introduced a *secondary* stochastic process $Y_t$ driving the volatility $\sigma(Y_t)$ and assumed its drift and diffusion coefficients to be generic functions of $Y$ but not of $t$ explicitly, in order to limit the discussion to time homogeneous processes. Here, $W_1$ and $W_2$ are two Wiener processes with the following correlation structure

$$d\,\langle W_{1,t}, W_{2,t'}\rangle = \rho\,\delta(t-t')\,dt\ ,$$

with $\rho \in [-1,1]$; this assumption introduces a correlation between prices and the volatility which, as will be clear in the following, is necessary to account for the leverage effect. Unlike prices, it is quite reasonable to assume that the volatility can not diffuse indefinitely and the financial belief of a "normal level" of the volatility is often incorporated assuming the secondary process to follow a *mean reverting* process

$$dY_t = -\alpha(\gamma - Y_t)\, dt + b(Y)\, dW_{2,t}\ , \tag{3.3}$$

where $\alpha$ is a positive constant. In this case, the mean of $Y_t$ converges exponentially toward its long run value $\gamma$ thus justifying the denomination of the process (3.3). A common choice for the diffusion coefficient is to take it

[2] The Markovian assumption is not mandatory and $\sigma_t$ could be modeled with jump process, or in general non Markovian processes as well.

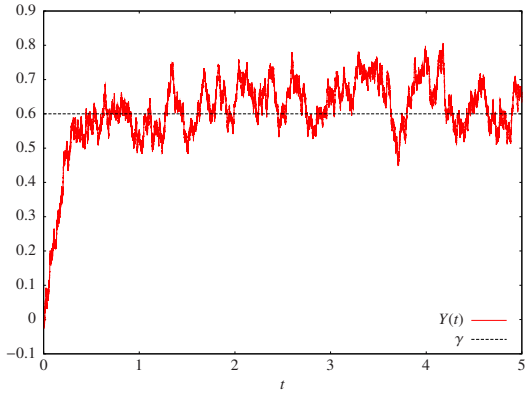



Figure 3.1: Sample path for an Ornstein-Uhlenbeck process, corresponding to parameters $\alpha = 7.8$, $\gamma = 0.6$ and $k = 0.3$ (arbitrary units).

constant, $b(Y) = k$, recovering the Ornstein-Uhlenbeck process [55]. Direct integration of Eq. (3.3) for this specific case shows that the PDF of the process is the one of a Normal distribution with time dependent mean and variance

$$\begin{aligned} \langle Y_t \rangle &= Y(t_0)e^{-\alpha(t-t_0)} + \gamma \left[1 - e^{-\alpha(t-t_0)}\right] \\ \mathrm{Var}[Y_t] &= \frac{k^2}{2\alpha}\left[1 - e^{-2\alpha(t-t_0)}\right] \ ; \end{aligned}$$

over a time interval of the order of $1/\alpha$ the process relaxes toward the value $\gamma$, keeping to fluctuate around this value with an amplitude given by the value of the stationary variance $k^2/2\alpha$, an example of which is shown in Fig. 3.1. This process, well known in physics for providing the realistic description of the Brownian motion, in this financial context is chosen as a prototype of a noisy relaxation process.

## 3.2 SVMs: three examples

Given the Eq.s (3.1)-(3.2), different models are obtained for every possible choice of the functions $a(Y)$, $b(Y)$ and $\sigma(Y)$. In the financial literature, the choice is often guided by the analytical tractability of the corresponding model, rather that its degree of realism and here we would like to compare briefly three deeply studied examples, focusing on their ability or inability in reproducing the empirical evidences discussed in Chapter 1.

**The Heston model (H).** One of the most studied models, widely exploited by practitioners, is the Heston model [58]. The dynamics chosen for $Y$ is a Cox-Ingersoll-Ross (CIR) process [59], originally applied to interest rates modeling and defined by the SDE

$$dY_t = -\alpha(Y_t - \gamma)\, dt + \kappa\sqrt{Y_t}\, dW_{2,t} \ , \tag{3.4}$$

while the volatility entering Eq. (3.1) is defined as $\sigma(t) = \sqrt{Y_t}$. Feller [60] proved that for $\gamma > 0$ the Fokker-Planck equation associated to (3.4) is well defined over $Y \in [0, +\infty)$ and admits as a stationary solution a Gamma PDF

$$\Pi_{\rm st}(Y) = \frac{A^A}{\Gamma(A)} \frac{Y^{A-1}}{\gamma^A} e^{-\frac{A}{\gamma}Y} \;, \quad A = \frac{2\alpha\gamma}{\kappa^2} \,. \tag{3.5}$$

A closed form expression for the PDF of the returns $p(x;t)$ is not available for this model, but the analytical expression of its characteristic function $f\omega;t$ has been worked out in [61], so that $p(x;t)$ can be evaluated numerically through

$$p(x;t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega \, e^{{\rm i}\omega x} f(\omega;t) \,.$$

A careful expansion [61] of $f(\omega;t)$ for large $x$ shows that the tails of the PDF are exponential at all times

$$p(x;t) = \begin{cases} e^{-xq^+(t)} & x > 0 \\ e^{xq^-(t)} & x > 0 \end{cases}, \quad q^\pm(t) = \mp i p_1^\pm(t) \,,$$

the slopes $q^\pm(t)$ of the exponential tails, real and positive, being determined by the singularities $p_1^\pm(t)$ closest to the real axis of the extension of $f(\omega;t)$ to the complex plane of $\omega$. The exponential decay corresponds to an excess of kurtosis which decreases with time; indeed, it can be shown that the fraction of probability contained in the non Gaussian tails becomes negligible for long times, in agreement with the Normal shape of return distributions over large time scales. As long as $\rho \neq 0$ those slopes are different each other, making $p(x;t)$ asymmetric and corresponding to non zero skewness.

**The Stein-Stein model (S2).** This model assumes a volatility linear in $Y$, $\sigma(t) = Y_t$, with the secondary process obeying an Ornstein-Uhlenbeck dynamics; it was first studied by Stein and Stein [62] for the case $\rho = 0$, and then generalized in [63] for non zero correlation. More recently, its statistical characterization has been addressed in [28], where, in particular, the model Fokker Planck equation is solved in the Fourier space for the conditional characteristic function $f_t(\omega|\sigma_0)$.

Much as for the Heston model, the PDF turns out to have exponential decay which is asymmetric as long as $\rho \neq 0$. Skewness and kurtosis vanish for long times with a scaling in agreement with the Central Limit Theorem (CLT)

$$\zeta \sim \frac{1}{\sqrt{\alpha t}} \;, \quad \kappa \sim \frac{1}{\alpha t} \quad (\text{for } \alpha t \gg 1) \;;$$

nevertheless the stationary distribution converges to a Normal one only for $k/\alpha \ll 1$, that is when the volatility is only weakly random, see [28].

**The exponential Ornstein-Uhlenbeck model (ExpOU).** Keeping on to describe $Y$ through the Ornstein-Uhlenbeck dynamics (3.3), we may explore choices for $\sigma(Y)$ more complicated than the linear assumption made by the Stein-Stein model. An interesting one is the exponential form

$$\sigma(t) = m\, e^{Y_t} \tag{3.6}$$

where $m$ is a positive strength parameter providing an additional degree of freedom. The corresponding model, first pioneered by Scott [64], has found renewed interest from the Econophysicists' community [34, 65, 66] where it is referred to as the exponential Ornstein-Uhlenbeck model. As it will be clear in a while, the reason of that interest is its ability to take into account realistically almost all the stylized facts mentioned in Chapter 1.

Since the process (3.3) is Gaussian at all times, the volatility (3.6) is Log-Normal; in particular the stationary distribution of the volatility reads

$$\Pi_{st}(\sigma) = \frac{1}{\sigma\sqrt{2\pi\beta}} \exp\left\{ -\frac{\left[\log(\sigma) - \log(\sigma_0)\right]^2}{2\beta} \right\} \tag{3.7}$$

where $\beta \doteq k^2/2\alpha$ is the stationary variance of the $Y$ process and $\sigma_0 \doteq me^{\gamma}$.

Unlike the previous models, analytical information about the return distribution (or at least its CF) of this model is still lacking and so far no analytical approximation allows to exactly say whether the ExpOU tails decay exponentially rather than as a power law. Nevertheless Monte Carlo simulations clearly show [67, 7] positive/negative skewness depending on $\rho \gtrless 0$ and an excess of kurtosis governed by the level $\beta$ of stationary variance.

**Correlations.** The return-volatility correlation (1.11) is known in closed form for all the previous models, while the volatility autocorrelation function (1.12) can be computed only for the S2 and ExpOU models [68]. They all succeed in predicting an exponential decay of $\mathcal{L}(\tau)$, proportional to $\rho$, with time scales of order ten days but the ExpOU turns out to be the more realistic of the three, featuring multiple time scales [34] for $\mathcal{A}(\tau)$. Indeed, for long times with respect to $1/\alpha$, it exhibits an exponential scaling governed by the characteristic relaxation time of the secondary process

$$\mathcal{A}_{ExpOU}(\tau) \approx \frac{4\beta}{3e^{4\beta} - 1} e^{-\alpha\tau} \quad (\alpha\tau \gg 1) \ ,$$

while on the long run the decay exponent depends on the "strength of the randomness" $k$ of the $Y$ process

$$\mathcal{A}_{ExpOU}(\tau) = \frac{1}{3e^{4\beta} - 1} \left[ e^{4\beta - 2k^2\tau} - 1 \right] + \mathcal{O}(\alpha^2\tau^2) \quad (\alpha\tau \ll 1) \ .$$

A calibration of the ExpOU model over DJIA index [34] reveals $\tau_{long} = 1/\alpha \approx$ 550 days and $\tau_{short} = 1/2k^2 \approx 35$ days, consistently with the regime separation for $\mathcal{A}(\tau)$ discussed at the end of Chapter 1.

We avoid resorting to a more detailed review of empirical analysis (available in the cited references) and we limit ourselves to claim that the ExpOU model provides the more realistic description of stocks or stock indexes returns among the three considered cases, since it

- provides a realistic Log-Normal description for high frequency volatility;
- reproduce the fast exponential decay of $\mathcal{L}(\tau)$;
- features a multiple time scale decay of $\mathcal{A}(\tau)$ as observed empirically.

## 3.3 The linearized ExpOU model

Now we will describe limit of low volatility fluctuations of the ExpOU model, studied in [67, 7, 8]; it corresponds to requiring that the stationary variance of the secondary process would be small, $\beta = k^2/2\alpha \ll 1$.

In terms of centered log-returns, $X_t = \log(S_t/S_0) - \mu(t - t_0)$, the asset dynamics for the ExpOU model can be rewritten as

$$dX_t = -\frac{1}{2}m^2 e^{2Y_t}\, dt + m e^{Y_t} dW_{1,t} \tag{3.8}$$

which follows by application of Itô lemma to the variable $X_t$. As long as $Y$ is weakly random, the two exponential functions can be approximated by a first order Taylor expansion around its stationary mean $\gamma$. Introducing the shifted process $Z = Y - \gamma + 1$ and the rescaled constant $\bar{m} = m\, e^{\gamma}$, the linearized dynamics, referred to as the LinExpOU model in the following, reads

$$dX_t = -\frac{\bar{m}^2}{2}(2Z(t) - 1)\, dt + \bar{m} Z(t)\, dW_{1,t} \tag{3.9}$$

$$dZ_t = \alpha(1 - Z(t))\, dt + k\, dW_{2,t}\,, \tag{3.10}$$

with initial conditions $X_{t_0} = X_0 = 0$ and $Z_{t_0} = Z_0 = Y_0 - \gamma + 1$. Following the original technique by Heston, we can express the PDF of $X$ in terms of its CF

$$p(x, \tau | x_0, z_0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\omega x} f(\omega, \tau; x_0, z_0) d\omega \tag{3.11}$$

with $\tau \doteq t - t_0$ and, if assuming the final condition $e^{i\omega x}$, the backward Fokker Planck equation [39] associated to Eq.s (3.8)-(3.10) is readily translated into a partial differential equation for $f(\omega, \tau; x_0, z_0)$; as shown in a detailed way in [67], the latter can be solved exactly by means of the trial guess

$$f(\omega, \tau; x_0, z_0) = \exp\left\{\mathcal{A}(\omega, \tau) + \mathcal{B}(\omega, \tau) z_0 + \mathcal{C}(\omega, \tau) z_0^2 + i\omega x_0\right\}\,. \tag{3.12}$$

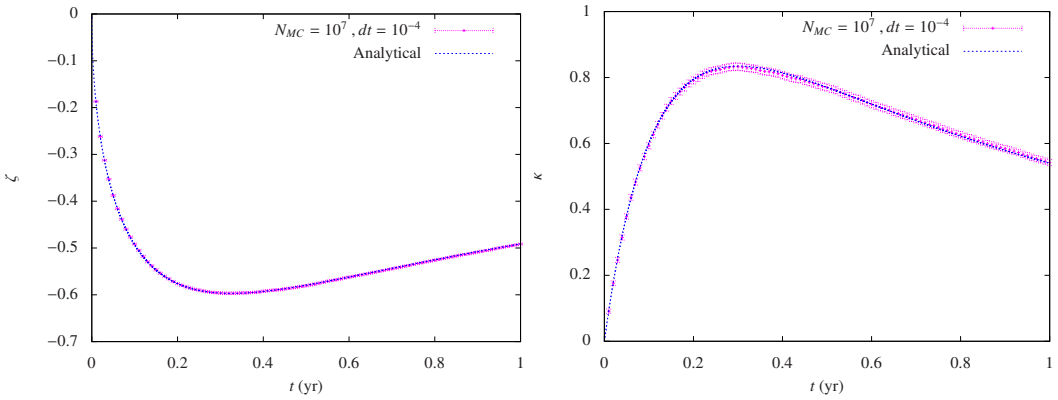



Figure 3.2: Analytical expressions for the skewness (left) and kurtosis (right) for the LinExpOU model, compared with the corresponding cumulants from an intensive Monte Carlo simulation ($10^7$ paths) of Eq.s (3.9),(3.10). Parameters: $\alpha = 6.2$, $\beta = 0.13$, $\gamma = 0$, $\bar{m} = 0.266$ and $y_0 = 0$.

The explicit expressions [8] of the three functions $\mathcal{A}, \mathcal{B}, \mathcal{C}$, representing an important and in applications useful result, read

$$\begin{aligned} A(\tau,\omega) = & \left[\frac{h}{2} + 2\alpha\frac{n-h}{d} + 2k^2\left(\frac{n-h}{d}\right)^2 + \frac{b-d}{4}\right]\tau \\ & - \frac{1}{2}\left[\log\left(1-ge^{-d\tau}\right) - \log\left(1-g\right)\right] \\ & - 2k^2\frac{e^{-d\tau}-1}{(1-g)\left(1-ge^{-d\tau}\right)}\left\{\frac{g}{d^3}\left[\frac{\alpha}{2k^2}(b+d)-h\right]^2\right. \\ & \left. + \frac{((g+1)h-2n)^2 + 2(n-gh)(n-h)}{d^3} + \frac{g}{d^3}(n-h)^2\right\} \\ & - 4k^2\,\frac{(g+1)h-2n}{d^3}\left(\frac{\alpha}{k^2}b-2h\right)\frac{\left(1+ge^{-\frac{d}{2}\tau}\right)\left(e^{-\frac{d}{2}\tau}-1\right)}{(1-g)\left(1-ge^{-d\tau}\right)} \end{aligned} \tag{3.13}$$

$$B(\tau,\omega) = 2\,\frac{e^{-\frac{d}{2}\tau}\left[(g+1)h-2n\right] + n + e^{-d\tau}(n-gh) - h}{d(1-ge^{-d\tau})} \tag{3.14}$$

$$C(\tau,\omega) = \frac{b-d}{4k^2}\frac{1-e^{-d\tau}}{1-ge^{-d\tau}}\,, \tag{3.15}$$

where the auxiliary dimensionless functions $b \doteq 2\alpha(1-i\rho\Phi)$, $d \doteq \sqrt{2\alpha^2\Phi^2 + b^2}$, $g \doteq (b-d)/(b+d)$, $h \doteq i\alpha\bar{m}\Phi/k$, $n \doteq \alpha(b-d)/(2k^2)$, and $\Phi \doteq k\bar{m}\omega/\alpha$ have been introduced. It is worth noting that the difference of principal logarithms in (3.13) has not been contracted; indeed, this operation can be performed only if taking into account a suitable phase correction (see Eq. (2.4) in [69]). Given the previous analytical expressions, and remembering the relation (1.5) linking the CF and the order $n$ cumulant $k_n(\tau)$, it is possible to compute the cumulants for the LinExpOU model exactly up to arbitrary orders. However,

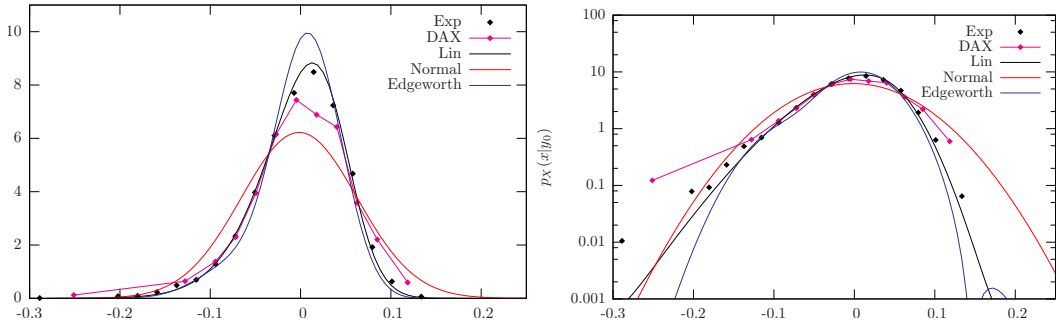



Figure 3.3: Linear and semi-log plots of the PDF of the returns over 25 days of the DAX30 index, compared with the PDFs of the LinExpOU model and the ExpOU model with parameters calibrated from the data, and with the Gaussian fit. Also reported an Edgeworth expansion approximation, (see [67] for more details).

most of the interesting information about the PDF requires the knowledge of just the first four cumulants. Their quite involved expressions are reported in Appendix A, and they have been obtained performing a 4-th order Taylor expansion of $f$ by means of MATHEMATICA® software, and then extracting its four coefficients [8] and multiplying them by the appropriate constant factors. Typical shapes for $\zeta$ and $\kappa$ as functions of $\tau$ are reported in Fig 3.2, where the obtained analytical expressions are confirmed by the Monte Carlo simulation. In the limit of small times, $\tau \to 0$, they vanish with the following scalings

$$\zeta \sim 3\frac{k\rho}{Z_0}\sqrt{\tau}\,, \quad \kappa \sim 4\frac{k^2(1+2\rho^2)}{Z_0^2}\tau\,,$$

while for $\tau \to \infty$ it is the CLT which regulates convergence to zero [7]

$$\zeta \sim -6\frac{k^2}{\alpha^2}\frac{\left[\frac{1}{\alpha} - k\rho\left(\frac{1}{4} + \frac{m^2}{\alpha} + \frac{\alpha}{k^2}\right) + m\,\rho^2\right]}{\left[\left(\frac{k\,m}{\alpha}\right)^2 + \frac{1}{2}\frac{k^2}{\alpha} - 1 - 2\rho\,\frac{k\,m}{\alpha}\right]^{3/2}}\frac{1}{\sqrt{\tau}}$$

$$\kappa \sim 3\frac{k^2}{\alpha^2}\frac{\left[4\left(\frac{k\,m}{\alpha}\right)^2 + \frac{k^2}{2\alpha} + 4 - 24\frac{k\,m}{\alpha}\,\rho + 2\rho^2\left(6\left(\frac{k\,m}{\alpha}\right)^2 + \frac{k^2}{\alpha} + 6\right) - 8\frac{k\,m}{\alpha}\,\rho^3\right]}{\left[\left(\frac{k\,m}{\alpha}\right)^2 + \frac{1}{2}\frac{k^2}{\alpha} - 1 - 2\rho\,\frac{k\,m}{\alpha}\right]^2}\frac{1}{\tau}\,.$$

So, we conclude that $p_{LinExpOU}(x)$ is leptokurtic at all finite time horizons ($\kappa(t) \geq 0$), and it converges toward a Gaussian regime asymptotically.

In [67, 7], it has been shown that the approximation leading to (3.9) is justified for $\beta \lesssim 10\%$ and that this value roughly matches what is found empirically for stock indexes. Fig. 3.3 shows the distribution of the returns $X_t$ of the DAX30 index for a time horizon of 25 days, along with the PDFs of the LinExpOU and ExpOU models with parameters calibrated with the procedure explained in [67] and corresponding to $\beta = 0.116$. It is clear that the

agreement between $p_{LinExpOU}$ and $p_{ExpOU}$ gets worse on the tails; nevertheless, the LinExpOU model is still a much better approximation with respect to the Gaussian fit, also reported in the figure, which grossly underestimate the central region of the data and is unable to account for the different weights of the tails. The better performing model is the ExpOU, giving a higher level of kurtosis with this value of $\beta$; despite of its good properties and as indeed deducible from the log-linear plot, it is not able to reproduce extreme kurtosis as those observed in high frequency (intraday) data, which, as discussed in Chapter 1 are likely to be described by a power law distribution.

## 3.4 A minimal linear model

We now aim to establish a link between SVMs and the multiplicative noise diffusion process extensively studied in Chapter 2. There we saw explicitly that the Ornstein-Uhlenbeck process (3.3) is a specific case of the dynamics (2.2) corresponding to $e > 0$, $c = 0$ with $a < 0$ (necessary for the process to be mean-reverting). On the other hand, when $a < 0$, for $d = e = 0$ we also recover the CIR process (3.4). In some sense, the MNDP seems to generalize the standard assumptions for the dynamics of the secondary process $Y$ and we could ask what are the properties of the SVM assuming it to describe the volatility

$$\begin{aligned} dX_t &= \sigma(Y_t)\, dW_{1,t}\ , \quad \sigma(Y_t) = \sqrt{c}\, Y_t \\ dY_t &= (aY_t + b)\, dt + \sqrt{cY_t^2 + dY_t + e}\, dW_{2,t}\,, \end{aligned}$$

where $X_t$ is the detrended log-return and we are going to require $a < 0$ with the usual conditions on the other coefficients. Unfortunately, the analysis of the previous model is highly complicated, unless to proceed in a full numerical way. On the other hand we could ask if it exhibits any realistic and sensible limit which is analytically tractable. Actually, in Chapter 2 we also analyzed the case $d = e = 0$, and we found that for $b > 0$ the stationary distribution of the process is an Inverse Gamma whose support is $[0, +\infty)$, which, from Chapter 1 we know to be probably the more realistic description of the empirical distribution of the high frequency volatility. So, provided to identify the definition (1.8) with $\sigma(Y_t)$ we are led to consider the following linear model

$$\begin{aligned} dX_t &= \sigma(Y_t)\, dW_{1,t}\,, \quad \sigma(Y_t) = \sqrt{c}\, Y_t \\ dY_t &= (aY_t + b)\, dt + \sqrt{c}\, Y_t\, dW_{2,t} \end{aligned} \tag{3.16}$$

with the volatility $\sigma_t$ relaxing for $t \gg |a|^{-1}$ toward an Inverse Gamma

$$\Pi_{st}(\sigma) = \frac{\nu^\lambda}{\Gamma(\lambda)} \frac{e^{-\nu/\sigma}}{\sigma^{\lambda+1}}\,, \tag{3.17}$$

with shape parameter $\lambda$ and scale parameter $\nu$ given by

$$\lambda = 1 - \frac{2a}{c} \quad \nu = \frac{2b}{\sqrt{c}}\,. \tag{3.18}$$

It follows that in this model the stationary volatility exhibits power law tails, the moments $\mu_{n,st}$ of $Y$ diverging for $n \geq \lambda$. Even though under investigation, we are going to show that the above model features rather interesting properties, despite its simplicity and linearity, making it a potentially good candidate to model high frequency data, as we hope it will be confirmed by our ongoing empirical analysis.

### 3.4.1 Power law tails

The structure of the model (3.16) allows to exactly compute the moments of the distribution of $X_t$, in the same spirit of what made in the previous chapter for the MNDP, and the power law tail of the volatility distribution (3.17) induces a power law scaling in the distribution of $X_t$. The proof of this statement relies in the following equation relating the moments of $X$ and those of $Y$

$$\langle X_t^n \rangle = \langle X_{t_0}^n \rangle + \frac{1}{2} n(n-1)c \int_{t_0}^{t} ds \, \langle X_s^{n-2} Y_s^2 \rangle \tag{3.19}$$

which is easily proved by application of Îto lemma to $X_t^n$ and taking its expected value. On their side, the correlations entering previous equation satisfy the following recursive equation

$$\begin{aligned} \frac{d}{dt} \langle X^m Y^n \rangle = F_n \langle X^m Y^n \rangle + A_n \langle X^m Y^{n-1} \rangle \\ + c\rho \, m \, n \langle X^{m-1} Y^{n+1} \rangle + \frac{1}{2} m(m-1)c \langle X^{m-2} Y^{n+2} \rangle \, . \end{aligned} \tag{3.20}$$

It is not trivial to write down the general solution $\langle X^m \sigma^n \rangle$, but since the moments $\mu_n(t)$ of $\sigma$ are known analytically for every $n$ and $t$, Eq. (3.20) can be solved recursively from the lowest order up. To convince ourselves let us consider the first non trivial case, that is the third moment; its expression involves the correlation between $X$ and $\sigma^2$

$$\langle X_t^3 \rangle = \langle X_{t_0}^3 \rangle + 3c \int_{t_0}^{t} \langle X_s Y_s^2 \rangle \, ds \, . \tag{3.21}$$

The quantity $\langle X_s \sigma_s^2 \rangle$ is solution of the following differential equation

$$\frac{d}{dt} \langle X_t Y_t^2 \rangle = F_2 \langle X_t Y_t^2 \rangle + A_2 \langle X_t Y_t \rangle + 2\rho c \mu_3(t) \; ; \tag{3.22}$$

in turn, $\langle X_t \sigma_t \rangle$ satisfy its own ODE

$$\frac{d}{dt} \langle X_t Y_t \rangle = F_1 \langle X_t Y_t \rangle + \rho c \, \mu_2(t) \tag{3.23}$$

which is readily solved

$$\langle X_t Y_t \rangle = e^{a(t-t_0)} \left\{ \langle X_{t_0} Y_{t_0} \rangle + \rho c \int_{t_0}^{t} e^{-a(s-t_0)} \mu_2(s) ds \right\} \, .$$

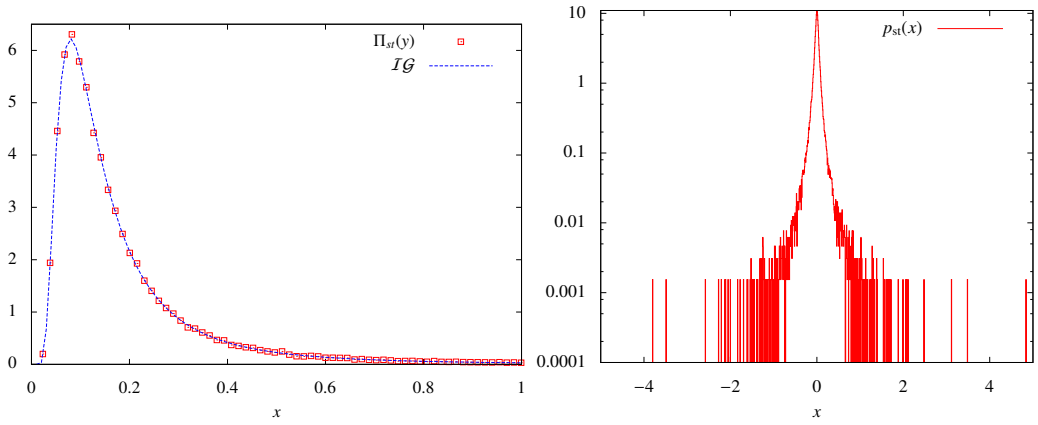



Figure 3.4: Simulation corresponding to parameters $a = -0.665$, $b = 0.133$, $c = 1$, $\rho = -0.3$ with $\Pi_0(y) = \delta(y)$. Left panel: simulated Inverse Gamma stationary distribution of $Y$ (points) and analytical curve (line); right panel: simulated PDF of $X$ for $t = 1$ (arbitrary units).

Once this solution is substituted back, Eq (3.22) becomes itself an ODE whose explicit solution could be plugged in (3.21). Reminding that the $\mu_n$s are simple combinations of exponential functions, we see that the last integral could be solved for fixed $t$, for instance with the algorithmic methodology discussed in Chapter 2. Even without performing this tedious calculation, we see the key point: the asymptotic behavior of $\langle X^n \rangle$ descends on that of moments $\mu_j$ for $j \leq n$ of the volatility process. In particular, power law tails in the volatility distribution can induce power law tails in the returns distribution: if $\mu_n$ is not finite, the same will be true for $\langle X^n \rangle$. When the starting distribution $\Pi_{t_0}(\sigma)$ of the volatility is not a power law distribution, for finite $t$ all the moments of $p(x;t)$ will be finite, but as $t$ increase, $p(x;t)$ will approach a power law regime more and more. Let us see that with a numerical example. In the left panel of Fig. 3.4 it is shown the stationary PDF of $Y$ for $\Pi_0(y) = \delta(y)$ and for a choice of parameters corresponding to

$$\Pi_{st}(Y) \sim \frac{1}{y^{3.33}} \quad \text{for } y \gg 1 \ ;$$

from the above reasoning, we expect asymptotically a power law scaling for $X$ too, and we are able to give a range for its tail index

$$p(x) \sim \frac{1}{x^{1+\nu}} \quad \text{with } 2 < \nu \leq 3 \ .$$

The right panel of Fig. 3.4 shows $p(x;t)$ for $t = 1$ which exhibits, despite the stationary state of $Y$ is far from having been reached ($1/|a| \simeq 1.5$), a high degree of leptokurtosis. On the other hand, if we choose a power law initial condition for $\Pi_{t_0}(y)$, this would reflect in a power law distribution for $X$ at finite times.

### 3.4.2 Return-volatility correlation

Now we will compute exactly the return-volatility correlation for the linear model (3.16). The basic tool we will use is the Novikov theorem, which we briefly review here. Let $\boldsymbol{\zeta}(t) = (\zeta_1(t), \ldots, \zeta_n(t))$ be an $n$-dimensional Gaussian white noise, with correlation structure given by

$$\langle \zeta_i(t)\zeta_j(t')\rangle = \rho_{ij}\delta(t-t') \ .$$

The Novikov theorem is useful in calculating expectations of functionals $f(t, |\boldsymbol{\zeta}|)$ of $\boldsymbol{\zeta}$ and one of its alternative forms [70, 33] is the following

$$\langle f(t,|\boldsymbol{\zeta}|)\zeta_j(t')\rangle = \sum_{i=1}^{n} \rho_{ij} \left\langle \frac{\delta f(t,|\boldsymbol{\zeta}|)}{\delta\zeta_i(t')} \right\rangle \ . \tag{3.24}$$

Taking into account that it is always possible to express a Wiener variation as $dW_t = \zeta_t\, dt$, and that $\langle dW_t^2\rangle = dt$, the numerator of the definition (1.11) can be written as

$$\begin{aligned} \left\langle dX_{t+\tau}^2 dX_t \right\rangle &= \left\langle \sigma_{t+\tau}^2 \sigma_t\, dW_{1,t} \right\rangle \ \left\langle dW_{t+\tau}^2 \right\rangle \\ &= \left\langle \sigma_{t+\tau}^2 \sigma_t \ \zeta_{W_1}(t) \right\rangle \ dt^2 \\ &= c^{3/2} \ \left\langle Y_{t+\tau}^2 Y_t \ \zeta_{W_1}(t) \right\rangle \ dt^2 \ ; \end{aligned} \tag{3.25}$$

while the denominator is equal to $\langle dX_t^2\rangle^2 = c^2\mu_2(t)^2\, dt^2$. Due to the Novikov theorem, the expectation can be computed as

$$\begin{aligned} \left\langle Y_{t+\tau}^2 Y_t \ \zeta_{W_1}(t) \right\rangle &= \left\langle \left[ \frac{\delta(Y_{t+\tau}^2 Y_t)}{\delta\zeta_{W_1}(t)} \right] \right\rangle = \left\langle \left[ \frac{\delta\zeta_{W_2}(t)}{\delta\zeta_{W_1}(t)} \frac{\delta(Y_{t+\tau}^2 Y_t)}{\delta\zeta_{W_2}(t)} \right] \right\rangle \\ &= \rho \ \left\langle \left[ \frac{\delta(Y_{t+\tau}^2 Y_t)}{\delta\zeta_{W_2}(t)} \right] \right\rangle = \rho \ \left\langle \left[ Y_{t+\tau}^2 \frac{\delta Y_t}{\delta\zeta_{W_2}(t)} + 2Y_{t+\tau}Y_t \frac{\delta Y_{t+\tau}}{\delta\zeta_{W_2}(t)} \right] \right\rangle \ . \end{aligned} \tag{3.26}$$

The functional derivative of $Y$ reads [33]

$$\frac{\delta Y_{t+\tau}}{\delta\zeta_{W_2}(t)} = \sqrt{c}\ H(\tau) e^{a\tau} Y_t \exp\left\{ \sqrt{c} \int_t^{t+\tau} dW_{2,s} \right\} \ , \tag{3.27}$$

where $H(\tau)$ is the Heaviside step function defined as

$$H(\tau) = \begin{cases} 0 & \text{for } \tau \leq 0 \\ 1 & \text{for } \tau > 0 \, . \end{cases}$$

Due to $H$, the functional derivative (3.27) gives no contribution for $\tau = 0$ and, after defining $\int_t^{t+\tau} dW_{2,s} \doteq \Delta_t W_2(\tau)$, we are left with

$$\left\langle dX_{t+\tau}^2 dX_t \right\rangle = 2c^2 H(\tau)\rho \left\langle e^{a\tau} Y_t^2 Y_{t+\tau} e^{\sqrt{c}\Delta_t W_2(\tau)} \right\rangle \ dt^2 \ .$$

A tricky calculation (see Appendix B) shows that the function $f(Y;t,\tau) \doteq \left\langle Y_t^2 Y_{t+\tau} e^{\sqrt{c}\Delta_t W_2(\tau)} \right\rangle$ is solution of a Volterra equation of the second kind and it reads

$$f(Y;t,\tau) = \mu_3(t) e^{(\frac{3}{2}c+a)\tau} + \frac{b}{a+c}\mu_2(t)\left[e^{(\frac{3}{2}c+a)\tau} - e^{\frac{c}{2}\tau}\right] \tag{3.28}$$

so that the leverage correlation finally reduces to

$$\mathcal{L}(t;\tau) = \frac{2\rho}{\mu_2(t)^2}\, H(\tau)\left\{\left[\mu_3(t) + \frac{b}{a+c}\mu_2(t)\right] e^{\alpha_1\tau} - \frac{b}{a+c}\mu_2(t) e^{\alpha_2\tau}\right\} \tag{3.29}$$

with $\alpha_1 \doteq (4a+3c)/2$ and $\alpha_2 \doteq (2a+c)/2$. This expression explicitly depends on the time $t$ starting from which we compute the correlation; nevertheless, when analyzing time series, we usually make the assumption of stationarity of the series, which implies to replace $\mu_2(t)$ and $\mu_3(t)$ in the previous expressions with their values at the stationary state. In order for $\mathcal{L}$ to be a finite quantity we have to assume that the stationary distribution of the volatility have finite third moment at least, that is $|a|/c > 1$ (see Chapter 2). The stationary values $\mu_{2,st}$ and $\mu_{3,st}$ are obtained as the limit for $t \to \infty$ of the expansion (2.5):

$$\mu_{2,st} = \frac{2b^2}{a(2a+c)} \tag{3.30}$$

$$\mu_{3,st} = -\frac{2b^3}{(2a+c)(a+c)a}\ . \tag{3.31}$$

When substituted in Eq. (3.29) we obtain the expression of the leverage correlation to be compared with real data:

$$\mathcal{L}(\tau) = -\rho H(\tau)\frac{a(2a+c)}{b(a+c)} e^{-\tau/\tau^{\mathcal{L}}}\ , \tag{3.32}$$

where we defined the leverage time scale $\tau^{\mathcal{L}} = 1/(|a| - c/2)$. So we see that the model we are considering is able to reproduce the exponential time decay of the leverage correlation for positive correlation times and its vanishing for $\tau < 0$.

### 3.4.3 Volatility autocorrelation

Turning to the volatility autocorrelation (1.12), we have

$$\begin{aligned}\left\langle dX_t^2 dX_{t+\tau}^2 \right\rangle &= \left\langle \sigma_t^2 \sigma_{t+\tau}^2\, dW_{1,t}\zeta_t \right\rangle\ dt^2 \\ &= \left\{2\rho\ \left\langle \left[\sigma_t^2 \sigma_{t+\tau} \frac{\delta\sigma_{t+\tau}}{\delta\zeta_{Z,t}}\, dW_{1,t}\right]\right\rangle + \left\langle \left[\sigma_t^2 \sigma_{t+\tau}^2 \frac{\delta\, dW_{1,t}}{\delta\zeta_t}\right]\right\rangle\right\}\ dt^2\ ,\end{aligned}$$

where previous equality follows after application of the Nokivov theorem and recalling that $\delta\sigma_t/\delta\zeta_{Z,t} = 0$. Recalling the relation (3.27) and the following result

$$\frac{\delta\, dW_{1,t}}{\delta\zeta_t} = \frac{\delta}{\delta\zeta_t}\int_t^{t+dt} \zeta_s ds = 1$$

we get to

$$\left\langle dX_t^2 dX_{t+\tau}^2 \right\rangle = \left\{ 2\rho\ c^{5/2} H(\tau) e^{a\tau} \left\langle \left[ Y_t^3 Y_{t+\tau} e^{\sqrt{c}\Delta_t W_2(\tau)}\, dW_{1,t} \right] \right\rangle + c^2\ \left\langle Y_t^2 Y_{t+\tau}^2 \right\rangle \right\} dt^2 \,. \tag{3.33}$$

Due to the presence of $dW_{1,t}$, the first term is of order $\mathcal{O}(dt^3)$ and can be neglected, leaving us with

$$\left\langle dX_t^2 dX_{t+\tau}^2 \right\rangle = c^2\ \left\langle Y_t^2 Y_{t+\tau}^2 \right\rangle\, dt^2 + \mathcal{O}(dt^3) \,.$$

The autocorrelation $\left\langle Y_t^2 Y_{+t+\tau}^2 \right\rangle$ can be computed exactly as shown in Appendix B, and the complete expression of the volatility autocorrelation becomes

$$\begin{aligned} \mathcal{A}(t;\tau) = \frac{1}{3\mu_4(t) - \mu_2(t)^2} \Bigg\{ & -e^{a\tau} \frac{2b}{a+c} \left[ \mu_3(t) - \mu_1(t)\ \mu_2(t) \right] \\ & + e^{(2a+c)\tau} \left[ \mu_4(t) + \frac{2b}{a+c}\mu_3(t) - \mu_2(t) \left( \mu_2(t) + \frac{2b}{a+c}\mu_1(t) \right) \right] \Bigg\} \,, \end{aligned} \tag{3.34}$$

where we approximated $\left\langle dX_{t+\tau}^2 \right\rangle$ with $\langle dX_t^2 \rangle$ in the denominator of (1.12) from stationarity arguments. As expected, $\mathcal{A}(t;\tau)$ depends on the $\mu_n$ up to $n = 4$, which we are going to assume to be finite at all times, imposing the consistency relation

$$\frac{|a|}{c} > 3/2 \,. \tag{3.35}$$

The fourth moment of $Y$ is readily computed as

$$\mu_{4,st} = \frac{4b^4}{(2a+3c)(a+c)(2a+c)a} \tag{3.36}$$

and after substitution of all the stationary $\mu_n$ we get to

$$\mathcal{A}(\tau) = \frac{1}{D} \left[ N_1\, e^{-\tau/\tau_1^{\mathcal{A}}} + N_2\, e^{-\tau/\tau_2^{\mathcal{A}}} \right] \tag{3.37}$$

where we have defined the three coefficients

$$\begin{aligned} D &= \frac{\left(4a^2 - 2ac - 3c^2\right)(a+c)}{c^2} \\ N_1 &= -\frac{(2a+3c)\,(2a+c)}{c} \\ N_2 &= a \end{aligned}$$

and the two volatility autocorrelation time scales

$$\tau_1^{\mathcal{A}} = \frac{1}{|a|} \,, \quad \tau_2^{\mathcal{A}} = \frac{1}{2|a| - c} \,. \tag{3.38}$$

At this point we foresee a potential drawback of the model: due to the condition $|a|/c > 3/2$ we have the following strict ordering among correlation times:

$$\tau_2^{\mathcal{A}} < \tau_1^{\mathcal{A}} < \tau^{\mathcal{L}} \ , \quad \text{with } \tau_1^{\mathcal{A}} > \frac{2}{3}\tau^{\mathcal{L}} \ ;$$

that is, the volatility autocorrelation would decay faster than the leverage correlation, contradicting the empirical evidence of a long range memory effect of the former (see Chapter 1). Nevertheless, we are considering a model which is possibly able to suitably describe high frequency data, still being linear, and we do not expect it to be able to capture such long term effects. On the other hand, the second inequality above reproduces the empirical evidence of a short range correlation time scale which is approximately the same for the leverage, $\tau^{\mathcal{L}}$, and the volatility autocorrelation, $\tau_1^{\mathcal{A}}$.

### 3.4.4 Estimation of parameters

Now we will provide a systematic methodology for estimating the parameters of our model, which are the constants $a$, $b$, $c$ entering the dynamics of $Y_t$, plus the correlation coefficient $\rho$. We will perform the estimation of the Standard & Poor 500 index daily returns from 1970 to 2010, approximating $dX_t$ with $\Delta X_t = X_{t+\Delta t} - X_t$, namely

$$dX_t \approx \Delta X_t = \log\left(\frac{S_{t+\Delta t}}{S_t}\right) - \left\langle \log\left(\frac{S_{t+\Delta t}}{S_t}\right)\right\rangle \ , \tag{3.39}$$

where $\Delta t = 1/250$ yr (one trading day). Taking into account that $dW_{1,t}$ is independent of $\sigma_t$ and that $|\Delta W_1|$ is distributed accordingly to a Folded Normal law, the following relations hold for the model (3.16):

$$A \doteq \frac{\langle|\Delta X|\rangle}{\langle|\Delta W_1|\rangle} = \sqrt{\frac{\pi}{2\Delta t}}\,\langle|\Delta X|\rangle = -\sqrt{c}\,\frac{b}{a} \tag{3.40}$$

$$B \doteq \frac{\langle\Delta X^2\rangle}{\langle\Delta W_1^2\rangle} = \frac{\langle\Delta X^2\rangle}{\Delta t} = c\,\frac{2b^2}{(2a+c)a} \tag{3.41}$$

$$C \doteq \frac{\left\langle|\Delta X|^3\right\rangle}{\left\langle|\Delta W_1|^3\right\rangle} = \sqrt{\frac{\pi}{(2\Delta t)^3}}\,\left\langle|\Delta X|^3\right\rangle = -\frac{2b^3\,c^{3/2}}{(a+c)\,(2a+c)\,a} \tag{3.42}$$

The constants $A$ and $B$ can be estimated directly from the data, providing us an estimation of the ratio $a/c$ through the relation

$$D \doteq \frac{B}{2\,(A^2-B)} = \frac{a}{c}\,. \tag{3.43}$$

The value of these estimators extracted from the series of the daily log-returns of the S&P500 index are reported in Table 3.1. It is crucial to observe that the value obtained for the ratio $|a|\,/c$ is compatible with the constraint (3.35),

| Estimators | S&P500 daily returns |
|---|---|
| A | 0.1457 $\mathrm{yr}^{-1/2}$ |
| B | 0.0295 $\mathrm{yr}^{-1}$ |
| C | 0.0107 $\mathrm{yr}^{-3/2}$ |
| $\lvert a\rvert/c$ | 1.7895 |

Table 3.1: Estimates from return sample averages. We compute the value of the estimators $A$, $B$, $C$ and $D$ for the daily log-returns of the S&P500 index during the period 1970-2010, exploiting the averages of $|\Delta X|$, $\Delta X^2$ and $|\Delta X|^3$. Yearly units ($\Delta t = 1/250$ yr).

| Estimators | S&P500 daily returns |
|---|---|
| $\tau^{\mathcal{L}}$ | 0.0864 yr |
| $\mathcal{L}(0^+)$ | -30.9515 |

Table 3.2: Estimation of the leverage time scale and the leverage amplitude obtained from the fit of the empirical leverage correlation (1.11) for the daily log-returns of the S&P500 index with the model predicted expression (3.32).

supporting the consistency of our model and the convergence of the volatility autocorrelation. Moreover, if we define $j^* = \max\{j \in \mathbb{N}, j \geq 1 \mid j < 1 + 2\lvert a\rvert/c\}$, we have $j^* = 4$, and in the light of the analysis made in Chapter 2 we conclude that $\mu_4(t)$ is the highest converging moment of $Y_t$. As explained early, the divergence of $\mu_5$ induces the divergence of $\langle X^5 \rangle$ and tail index $\beta$ of the return PDF must stay in the following range:

$$p(x) \underset{|x|\to\infty}{\sim} \frac{1}{|x|^{1+\beta}}\,, \quad 4 < \beta \leq 5\,.$$

The leverage correlation, as given by Eq. (3.32), provides a way to obtain the two further relations needed to fix the four free parameters of the model. Indeed, a two parameters fit of the function $\mathcal{L}(\tau)$ provides an estimation of the time scale $\tau^{\mathcal{L}}$ and the coefficient

$$\mathcal{L}(0^+) \doteq -\rho\,\frac{a(2a+c)}{b(a+c)}\,. \tag{3.44}$$

with the results reported in Table 3.2 and left panel of Fig. 3.5. In particular, the value obtained for the leverage time scale, $\tau^{\mathcal{L}} \approx 21$ days, and for its amplitude $\mathcal{L}(0^+)$, are consistent with those quoted in past analysis of stock indexes such as the Dow Jones Industrial Average [28, 33], and they confirm the short range nature of this effect.

At this point all the parameters can be recovered through the following

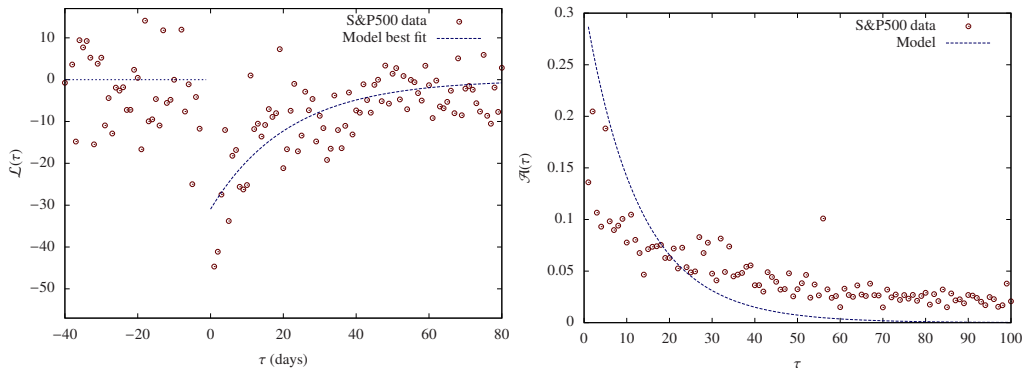



Figure 3.5: Left panel: best fit of the empirical leverage correlation with the model prediction (3.32) as a function of the two parameters $\tau^{\mathcal{L}}$ and $\mathcal{L}(0^+)$. Right panel: theoretical prediction for the volatility autocorrelation function of the daily returns of the S&P500 index 1970-2010, Eq. (3.37).

| Parameter | Estimate from S&P 500 | |
|---|---|---|
| $a$ | $-16.0608$ | $\mathrm{yr}^{-1}$ |
| $b$ | $0.8627$ | $\mathrm{yr}^{-1}$ |
| $c$ | $8.9749$ | $\mathrm{yr}^{-1}$ |
| $\rho$ | $-0.5089$ | |

Table 3.3: Model parameters estimated from the daily log-returns of the S&P500 index during 1970-2010 through the relations (3.48).

relations with the results reported in Table 3.3.

$$c = -\left[\tau^{\mathcal{L}}\left(D + \frac{1}{2}\right)\right]^{-1} \tag{3.45}$$

$$a = c \times D \tag{3.46}$$

$$b = -\frac{a+c}{\sqrt{c}}\frac{C}{B} \tag{3.47}$$

$$\rho = -\frac{b\,(a+c)}{a\,(2a+c)}\,\mathcal{L}(0^+)\,. \tag{3.48}$$

The correlation coefficient is negative, in agreement with the known left-skewed shape of daily return distributions and the relaxation time of the volatility process is $\tau^{\sigma} \doteq -1/a \approx 15$ days. This somewhat large value supports the lacking of evidence of mean reversion of the volatility at high frequencies (see e.g. [71, 72]) and, as a consequence, the need of assuming that it has already reached the stationary state at the time of observation.

While the fitted values of $\tau^{\mathcal{L}}$ and $\mathcal{L}(0^+)$ provides a good description of real data, as shown by the left panel of Fig. 3.5, the theoretical volatility autocorrelation for the estimated values of the parameters, Eq. (3.37), has not

a satisfactory agreement with the data. Indeed, as clear from the right panel of Fig. 3.5, it decays too rapidly and it is not able to capture the long range memory exhibited by the empirical curve. As observed early, this was expected because of the constraint (3.38). However, Eq. (3.38) was derived in the limit of $dt \to 0$ while here we approximated an infinitesimal time difference with a finite one and the logarithmic variation $\Delta X$ with $\sigma\Delta t$:

$$dX_t \approx \Delta X_t = \int_t^{t+\Delta t} dX_t \, .$$

The rational behind this approximation is that 1 day is small in yearly units, but it is not so with respect to ultra high frequency scales (minutes or less). In particular, the exact expression of $\Delta X_t$ would reveal the ($\rho$-dependent) effects of the correlated evolution of $X$ and $Y$ form $t$ to $t+\tau$ which, *a priori*, may be non negligible in determining the correlations (3.32),(3.37).

At this point, it is important to compare the return PDF predicted by the model with the data sample from which the model parameters have been estimated. Since the model is dynamic, it is even more important to asses to which extent the diffusion process (3.16) is able to capture the scaling properties of the empirical distribution over different time horizons. At this aim, with the parameters fixed from the daily S&P500 series, we reconstructed the theoretical PDFs simulating the process at different time scales, $\Delta t = 1, 3, 7, 14$ days, then comparing them with the corresponding empirical distributions obtained aggregating the daily returns. This comparison is shown in Fig. 3.6. The daily distribution is very well reproduced by the theoretical PDF, which is able to fully capture the leptokurtic nature of the daily data. Most of all, the plots show that the diffusive dynamics of our model, once the parameters have been fixed at the daily scale, follows closely the evolution of the empirical curves for larger $\Delta t$, capturing the progressive increasing of the asymmetry and the convergence of the central region of the distribution to a Gaussian regime, as dictated by the Central Limit Theorem. In this regard, it is especially important to notice that the theorem does not state that tails become Gaussian, which is particularly clear in this case as the variance of $p(x)$ grows linearly as the distribution becomes wider and wider while the tails decay as a as a power law at all finite time scales.

Concluding, the analytical characterization of the minimal model (3.16) reveals quite interesting features. It captures the Inverse Gamma nature of the high frequency volatility, which is responsible for the emergence of power law tails in the distribution of $X_t$, and it correctly forecasts the exponential decay of the leverage correlation. A possibly unpleasant feature is the existence of strict ordering relations between time scales which makes the volatility autocorrelation function decay faster than the return-volatility correlation. A simple procedure was described which allows to estimate the model parameters from the empirical series. The results of this estimation for the case of the S&P500 support the model consistency and they highlight its ability to capture the scaling properties of the distribution of the empirical data over different

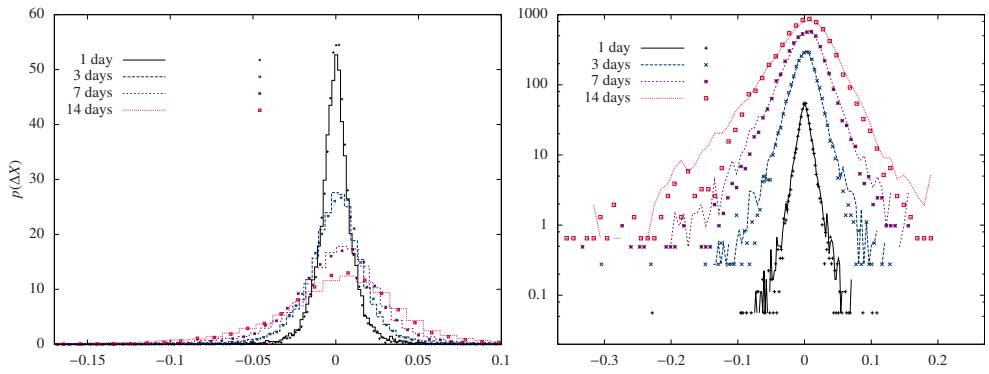



Figure 3.6: Linear plot showing the comparison between the return PDFs predicted by the model (lines) and the data for the S&P500 index, for different time scales.

time horizons. We conclude that the diffusione process (3.16) provides a quite realistic representation of the actual dynamics of the instantaneous volatility and of price returns at the daily time scale, also being able to account for the correlation between these processes, and, at the same time, it proves to model with good accuracy the return distributions for longer time horizons.

Chapter 4

# Modeling with Product Partition Models

In the previous chapters we considered the modeling of financial time series by means of SDEs, especially the class of SVMs. This kind of description focuses on the time evolution of the random paths of the volatility and the returns over a given time horizon; the corresponding PDFs aim to be non Gaussian functions able to account for the fat tailed, possibly power law, decay observed in high frequency returns and for the random behavior of the volatility. This approach overcomes the rough Gaussian hypothesis and it brings in a coupled dynamics of returns and volatility, this way introducing correlations which can not be taken into account easily in a static framework. However, when we look at the return empirical time series we are forced to assume that the series is stationary and that the returns are identically distributed variables in order to justify the identification of the return PDF at a given time horizon with the sample distribution. Loosely speaking, this approach is supported by the fact that returns, even though not independent, are nearly uncorrelated, as testified by the analysis of the return autocorrelation function (see Chapter 1); moreover, from the point of view of the modeling, it leads to the usual assumption that the volatility process has already reached its stationary state, as we did in performing the empirical analysis for the minimal linear model presented at the end of the previous chapter. Indeed, only if $\sigma_t$ has thermalized, meaning that its distribution has become invariant ($t$-independent), we can say that the increments $dX_t = \sigma_t \, dW_{1,t}$ are uncorrelated and identically distributed random variables.

A different approach is to consider returns as independent but not identically distributed and focus only on the statistical properties of the single observations, without investigating the possible dynamics linking observations at different times. From this viewpoint, we may look for similarities in these observations at the aim of identifying clusters of data sharing some statistical features. This would allow to gain insight into the underlying clustering structure, if any, and to test for the presence of outlying points. A useful statistical

framework to implement such a strategy is represented by the Bayesian Hierarchical Models. The Bayesian parametric approach makes an assumption about the form of distributions of the observations, and the parameters of these distributions, which determine the statistical properties of the data, are not fixed but they are drawn from *prior* distributions incorporating our degree of belief in that assumption. When these priors depend on other parameters to which a further prior is assigned, the Bayesian model becomes hierarchic and more flexible in reflecting the specific problem at hand. In this sense, Parametric Product Partition Models (PPM) are a special case of parametric Bayesian model in which a prior probability is assigned to the existence of clusters of observations whose distributions share the same values of some parameters. The aim of the inference process is to compute the *posterior* distributions, that is the distributions of the parameters conditioned to the observed sample of data, which should reflect the response of the model to the experimental evidence.

In this chapter we will describe the general form of a PPM, and we will introduce the Gibbs sampler, a special case of Markov Chain Monte Carlo algorithm which is needed to generate the posterior distributions. Then we will consider two specific models we devised for application to financial return series. Both of them assume *a priori* that returns are Normally distributed but they relax the assumption of identical distribution for the entries of the series; in the first one the returns share the same value of the variance and a partition structure is considered for their means, while in the second one clusters of observations can be identified depending on the value of their variance and the mean is assumed to be the same for all the observations. We will describe the sampling algorithms implementing these models and we will detail a data analysis performed over an Italian stock index and some of its components. In particular, the posterior distributions of the parameters will be shown, and we will see that the Bayesian estimates of these parameters usually modify the frequentist values, especially for variances. Finally, we will comment on the partition structures emerging during the sampling and we will introduce an original algorithm for the efficient identification of outliers in the series.

## 4.1 PPMs: general framework

Given a vector of observations $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)$, the process of parametric Bayesian inference consists of assuming the data to be distributed according to a PDF $p(\boldsymbol{y}|\boldsymbol{\theta})$ conditionally on a specific realization of some parameters $\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_m)'$. This conditional density is called the *likelihood* of observations. The parameters are assigned a *prior* distribution $g_\Theta(\boldsymbol{\theta})$, representing our degree of belief in the hypothesis: a different hypothesis corresponds to every possible value of $\boldsymbol{\Theta}$. Ideally, the prior should reflect some rational and coherent a priori information owned by the investigator. Basically, the aim of inference is the parameters estimation; in the Bayesian approach this is

provided in terms of a *posterior distribution* (briefly the *posterior*) $p(\boldsymbol{y}|\boldsymbol{\theta})$ of the parameters conditionally on the observed data, through the Bayes theorem

$$p(\boldsymbol{\theta}|\boldsymbol{Y}=\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\Theta}=\boldsymbol{\theta})g_{\Theta}(\boldsymbol{\theta})}{\int p(\boldsymbol{y}|\boldsymbol{\Theta}=\boldsymbol{\vartheta})g_{\Theta}(\boldsymbol{\vartheta})\,d\boldsymbol{\vartheta}}\ ; \tag{4.1}$$

in much the same way as the prior is considered a coherent degree of belief in the hypothesis, the posterior by Eq. (4.1) is considered a sort of response to evidence by an ideally rational observer.

If the prior $p(\boldsymbol{\theta})$ depends on some other parameters $\boldsymbol{\phi}$ and we are able to incorporate a priori information for them, the Bayesian model can be made *hierarchical* by posing a likelihood for $\boldsymbol{\Theta}$ conditionally on a given realization $\boldsymbol{\Phi}=\boldsymbol{\phi}$ and assigning to every possible scenario a prior $g_{\Phi}(\boldsymbol{\phi})$. This time, the joint posterior density for the parameters specifying the problem reads

$$p(\boldsymbol{\theta},\boldsymbol{\phi}|\boldsymbol{Y}=\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\Theta}=\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\Phi}=\boldsymbol{\phi})g_{\Phi}(\boldsymbol{\phi})}{\int p(\boldsymbol{y}|\boldsymbol{\Theta}=\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\Phi}=\varphi)g_{\Phi}(\boldsymbol{\varphi})\,d\boldsymbol{\varphi}}\ . \tag{4.2}$$

Product Partition Models were introduced in the nineties by Barry and Hartigan [73, 74], and can be thought of as Bayesian models in which the hypothesis which is given a prior belief is the existence of a given clustering structure on the vector of the data. Given a set of objects $S_0=\{1,2,\dots,n\}$ we define a partition $\rho=\{S_1,S_2,\dots,S_k\}$ by the two properties

- $S_i\cap S_j=0$ if $i\neq j$
- $\bigcup_{i=1}^{k} S_i = S_0$

and we assign the partition a *product probability* (PD) specified as follows

$$P\left(\rho=\{S_1,S_2,\dots,S_k\}\right) = K\prod_{i=1}^{k} C(S_i) \tag{4.3}$$

where $K$ is a normalization factor and the functions $C(S_i)>0$, named *cohesion functions*, represent a prior probability on the formation of a given cluster and they can be chosen in different ways. Let us suppose that a random observation $Y_i$ is associated to every object $i$ and let $p_{S_i}(\boldsymbol{Y}_{S_i}|\rho)$ the conditional probability (likelihood) of the observations associated to the cluster $S_i$ given that $S_i\in\rho$. We will also assume the observations $\boldsymbol{Y}_{S_1},\boldsymbol{Y}_{S_2},\dots,\boldsymbol{Y}_{S_k}$ to be independent conditionally on $\rho$ and that the conditional probability of the observations in a given cluster do not depend on how the other cluster of $\rho$ configure; in this case $\boldsymbol{Y}$ has a conditional product probability itself

$$p(\boldsymbol{Y}|\rho=\{S_1,S_2,\dots,S_k\}) = \prod_{i=1}^{k} p_{S_i}(\boldsymbol{Y}_{S_i}|\rho)\ . \tag{4.4}$$

Given Eq.s 4.3-4.4 the posterior probability of partitions is given by

$$p(\rho = \{S_1, S_2, \ldots, S_k\}|\boldsymbol{Y}) = \frac{K}{\nu(\boldsymbol{Y})} = \prod_{i=1}^{k} C(S_i) p_{S_i}(\boldsymbol{Y}_{S_i}|\rho) \tag{4.5}$$

where $\nu(\boldsymbol{Y})$ stands for the marginal density of the observations. When the partitions refers to the vector of parameters $\boldsymbol{\Theta}$ of a hierarchical model, these models are referred to as *parametric Product Partition Models*. We will assume that the distribution of the observations depends on a parameters vector $\boldsymbol{\Theta}$ and some parameter $\Phi$ which is common to all experimental units and we can consider the partition structure

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n) \doteq (\theta^*_{S_1}, \theta^*_{S_2}, \ldots, \theta^*_{S_k}; \rho)\,,$$

indicating with $\theta^*_{S_d} = \theta^*_d$ for $d = 1, \ldots, k$ the common value for the parameters in the cluster $S_d$ and with $\boldsymbol{\theta^*}$ the vector $(\theta^*_1, \ldots, \theta^*_{|\rho|})$. We will consider models in which the observations are independent conditionally on the partition $\rho$, and the parameters $\theta^*_d$ are independent and identically distributed (*i.i.d.*) with likelihood depending on $\phi$. Such a model can be summarized as follows

$$\begin{aligned} y_i|(\rho, (\theta^*_{S_1}, \theta^*_{S_2}, \ldots, \theta^*_{S_k}), \phi) &\overset{\text{ind}}{\sim} p(y_i|\theta_i, \phi) \\ \theta^*_{S_1}, \theta^*_{S_2}, \ldots, \theta^*_{S_k}|(\rho, \phi) &\overset{i.i.d.}{\sim} p(\cdot|\phi) \\ \rho &\sim \text{ product distribution (4.3)} \\ \phi &\sim g(\phi)\ . \end{aligned} \tag{4.6}$$

The marginal posteriors of the parameters can be computed analytically only in simple cases, but powerful Markov Chain Monte Carlo (MCMC) algorithms exist which allow to sample from them. Especially useful for our applications is the Gibbs sampling scheme reviewed in the following section.

## 4.2 The Gibbs sampler

The Gibbs sampler was introduced in [75, 76] and it is a specific instance of a MCMC. These algorithms have their roots in mathematical physics, dating back to the algorithm of Metropolis [77] and its later generalization by Hastings [78]. Let us consider the problem of generating a sample of realizations of the random variable $X$ from a density $f(x)$ obtained marginalizing out $p$ variables $y_1, \ldots, y_p$. In many problems the marginal distribution

$$f(x) = \int \ldots \int f(x, y_1, \ldots, y_p)\, dy_1 \ldots dy_p$$

is not known analytically either because $p$ is large or the above integral cannot be solved exactly. The Gibbs sampler solves this problem allowing to generate

a sample $X_1, X_2, \ldots, X_m \sim f(x)$ without requiring the knowledge of $f(x)$, provided the conditional distributions are known or it is possible to generate from them at least.

In order to illustrate the algorithm logic, we consider the simple bivariate case of two discrete variable $X$ and $Y$ with known conditional probability density functions $p(x|Y = y)$ and $p(y|X = x)$. A *Gibbs sequence* is a sequence of random variables

$$Y'_0, X'_0, Y'_1, X'_1, \ldots, Y'_k, X'_k$$

where, given the starting values $Y'_0 = y'_0$, the rest of the sequence is obtained sampling alternatively according to the scheme

$$\begin{aligned} X'_j &\sim p(x|Y'_j = y'_j) \\ Y'_{j+1} &\sim p(y|X'_j = x'_j) \ . \end{aligned}$$

The iteration $X'_j \longrightarrow X'_{j+1}$ defines a Markov chain to which the following transition probability is associated

$$P(x_{j+1}|X'_j = x_j) = \sum_y P(x_{j+1}|Y'_{j+1} = y) \times P(y|X'_j = x_j) \ .$$

We can therefore define a transition matrix $A_{x|x} = A_{x|y}A_{y|x}$ such that for every $k$ it holds the relation

$$f_k = f_{k-1}A_{x|x} \tag{4.7}$$

where $f_k$ is the marginal probability distribution of $X'_k$. Regardless of what the initial distribution $f_0$ is, it can be proved [79] that $f_k$ tends to a unique distribution as $k \to \infty$, that this distribution is a stationary point of Eq. (4.7), namely $f_k = f_k A_{x|x}$, and coincides with the true marginal $f(x)$.

With reference to a model of the type (4.6), the marginal densities of interest are the posteriors $p(\theta_i|\boldsymbol{Y})$ and $p(\phi|\boldsymbol{Y})$ of parameters given the observations and the Gibbs sampler proceed by sampling from the conditional distributions of the problems, which are $p(\phi|\boldsymbol{\theta}, \boldsymbol{y})$ and $p(\theta_j|\boldsymbol{\theta}_{-j}, \phi, \boldsymbol{y})$ for $j = 1, \ldots, k$, where $\boldsymbol{\theta}_{-j}$ is the vector obtained removing the $j$-th entry from $\boldsymbol{\theta}$. It has to be noticed that the partition $\rho$ is not sampled itself, but it is the updating of the values $\theta_i$ which possibly generates new clusters or induces a shuffling of the observations in the existing ones. This relies on a quite subtle result regarding the deep connection existing between PPMs and the class of non parametric Bayesian models with Dirichlet process prior [80, 81]. More precisely, a non parametric model can be viewed as a specific case of a PPM with the cohesions $C(S)$ specified as

$$C(S_i) = c \times (|S_i| - 1)! \tag{4.8}$$

where $|S_i|$ stands for the cardinality of the cluster $S_i$ and $c$ is a positive constant. This definition of $C(S)$ is quite general, supporting the formation of a small number of clusters and avoiding excessive fragmentation. This connection allows to adapt directly algorithms developed for the class of non parametric models such as that in [82]. In particular, it can be proved [83, 84]

that the full conditional density $p(\theta_i|\boldsymbol{\theta}_{-i}, \phi, \rho)$ can be expressed as a mixture of point masses at the values $\theta_j$ and a distribution that is proportional to $p(y_i|\theta_i, \phi)p(\theta_i|\phi)$

$$p(\theta_i|\boldsymbol{\theta}_{-i}, \phi, \theta) \propto \sum_{j \neq i} q_{ij}\delta_{\theta_j}(\theta_i) + q_{i0}p(y_i|\theta_i, \phi)p(\theta_i|\phi)\,, \tag{4.9}$$

where, after defining $p_0(y_i|\phi) \doteq \int p(y_i|\theta_i, \phi)p(\theta_i|\phi)\,d\theta_i$, the probabilities $q_{ij}$ for $i \neq j$ and $q_{i0}$ read

$$\begin{aligned} q_{ij} &= \frac{p(y_i|\theta_j, \phi)}{cp_0(y_i|\phi) + \sum_{l \neq i} p(y_i|\theta_l, \phi)} \\ q_{i0} &= \frac{c\,p_0(y_i|\phi)}{cp_0(y_i|\phi) + \sum_{l \neq i} p(y_i|\theta_l, \phi)}\ . \end{aligned} \tag{4.10}$$

They represent, respectively, the probability of overwriting $\theta_i$ with an existing value, and the probability of drawing a fresh new value for $\theta_i$, thus generating a new cluster. Given the representation (4.9), the other full conditional density needed to run the Gibbs scheme is $p(\phi|\boldsymbol{y}, \boldsymbol{\theta})$. This one can be worked out after computing the joint density of the observations and the parameters $f\left(\boldsymbol{y}, \boldsymbol{\theta}, \rho = \{S_1, \ldots, S_{|\rho|}\}, \phi\right)$ and exploiting the Bayes rule [1].

## 4.3 The $\boldsymbol{\mu}$-PPM and $\boldsymbol{\sigma^2}$-PPM models

Now we will describe a possible application of parametric PPMs to financial series as in [5, 6], where the elicitation of a partition structure on the vector of variances allows to stay in a Normal setting while relaxing the hypothesis of identical distribution. Let $\boldsymbol{y}$ be the vector of returns of a financial asset; we will assume the returns to be Normally distributed with parameters $(\boldsymbol{\theta}, \phi)$ and compare two different models. In the first one we will induce a partition structure on the vector of the means $\boldsymbol{\theta} = \boldsymbol{\mu}$, while $\phi$ will be the common variance of the returns, $\phi = \sigma^2$. In the second one the partition will refer to the vector of the variances $\boldsymbol{\theta} = \boldsymbol{\sigma^2}$, $\phi = \mu$ being the returns common mean. We will refer to the two cases as the $\boldsymbol{\mu}$-PPM model and the $\boldsymbol{\sigma}^2$-PPM model respectively. As far as the priors are concerned, we consider a Normal prior for the means and an Inverse Gamma prior for the variances. The latter choice can be motivated by observing that marginalizing a Gaussian likelihood with respect to an Inverse Gamma upon its variance gives a Student-$t$ PDF for the observations, which is known to provide a good fit of return market data, as discussed in Chapter 1.

[1] In the following we will omit the tedious derivation of the explicit expressions of the full conditional distributions. They directly follow from the structure (4.6) exploiting the Bayes rule and integrating out when marginalization is required.

The hierarchical structure for the $\boldsymbol{\mu}$-PPM model reads

$$\begin{aligned} y_t|(\rho,(\mu^*_1,\ldots,\mu^*_{|\rho|})\sigma^2) &\overset{ind.}{\sim} N(\mu_t,\sigma^2) \\ \mu^*_1,\ldots,\mu^*_{|\rho|}|(\rho,\sigma^2) &\overset{i.i.d.}{\sim} N(m,\tau_0^2\sigma^2) \qquad (4.11) \\ \rho &\sim \text{PD, with } C(S_d) = c\times(|S_d|-1)! \\ \sigma^2 &\sim IG(\nu_0,\lambda_0)\ , \qquad (4.12) \end{aligned}$$

where $IG(\nu_0,\lambda_0)$ is an Inverse Gamma distribution with scale parameter $\lambda_0$ and shape parameter $\nu_0$. We sample from the marginal posterior distributions of the above model with the following Gibbs sampling scheme.

**1.** Sample $\sigma^2$ from its full conditional distribution

$$\sigma^2|\boldsymbol{\mu},\boldsymbol{y} \sim IG\left\{\nu_0+\frac{T}{2}+\frac{|\rho|}{2},\lambda_0+\frac{1}{2\tau_0^2}\sum_{d=1}^{|\rho|}(\mu^*_d-m)^2+\frac{1}{2}\sum_{t=1}^{T}(y_t-\mu_t)^2\right\}. \qquad (4.13)$$

**2.** Update each $\mu_t$, $t=1,\ldots,T$, by sampling from the mixture

$$\mu_t|\boldsymbol{\mu}_{-t},\sigma^2,\boldsymbol{y} \sim \sum_{j\neq t} q_{tj}\delta_{\mu_j}(\mu_t) + q_{t0}\times N\left(\frac{y_t\tau_0^2+m}{1+\tau_0^2},\frac{\sigma^2\tau_0^2}{1+\tau_0^2}\right), \qquad (4.14)$$

where the weights are given by Eq.s (4.10)

$$\begin{aligned} q_{tj} &\propto \exp\left\{-\frac{1}{2\sigma^2}(y_t-\mu_j)^2\right\} \\ q_{t0} &\propto \frac{c}{\sqrt{1+\tau_0^2}}\exp\left\{-\frac{(y_t-m)^2}{2\sigma^2(1+\tau_0^2)}\right\} \\ &\sum_{j\neq t} q_{tj}+q_{t0}=1\ . \qquad (4.15) \end{aligned}$$

**3.** Before proceeding to the next Gibbs iteration we update the vector $\boldsymbol{\mu}^*$, given the partition $\rho$, sampling from

$$\mu^*_d \sim N\left(\frac{\sum_{t\in S_d} y_t + m/\tau_0^2}{|S_d|+1/\tau_0^2},\frac{\sigma^2}{|S_d|+1/\tau_0^2}\right) \quad d=1,\ldots,|\rho|. \qquad (4.16)$$

This step is not strictly required by the Gibbs sampling algorithm. It was introduced in [82] to speed up the convergence avoiding the Markov chain to be trapped in sticky patches of the Markov space.

For the case $\boldsymbol{\sigma}^2$-PPM we consider the following parametric PPM

$$\begin{aligned} y_t \mid \left(\mu, (\sigma^{2^*}_1 \dots \sigma^{2^*}_{|\rho|}), \rho\right) &\overset{ind.}{\sim} N(\mu, \sigma_t^2) \\ \mu \mid \left(\sigma^{2^*}_1 \dots \sigma^{2^*}_{|\rho|}\right), \rho &\sim N\left(m, \frac{\lambda_0}{T(\nu_0 - 1)}\right) \\ \sigma^{2^*}_1 \dots \sigma^{2^*}_{|\rho|} \mid \rho &\overset{i.i.d.}{\sim} IG(\nu_0, \lambda_0) \\ \rho &\sim \text{ PD, with } C(S_d) = c \times (|S_d| - 1)! \,, \end{aligned} \tag{4.17}$$

where we fixed the variance of the likelihood of $\mu$ in order to match the mean of the $\sigma_d^{2*}$ prior, $\langle \sigma^{2*} \rangle = \lambda_0/(\nu_0 - 1)$ and rescaling it with the dimension of the sample. Even in this case the full conditional distribution of the problem can be computed exactly, leading to the following Gibbs scheme:

**1.** Sample $\mu$ from its full conditional distribution

$$\mu | \boldsymbol{\sigma^2}, \boldsymbol{y} \sim N\left( \frac{m + \sum_{d=1}^{|\rho|} \frac{\lambda_0}{T(\nu_0-1)\sigma^{2*}_d} \sum_{i \in S_d} y_i}{1 + \sum_{d=1}^{|\rho|} \frac{|S_d|\lambda_0}{T(\nu_0-1)\sigma^{2*}_d}}, \frac{\frac{\lambda_0}{T(\nu_0-1)}}{1 + \sum_{d=1}^{|\rho|} |S_d| \frac{\lambda_0}{T(\nu_0-1)\sigma^{2*}_d}} \right) . \tag{4.18}$$

**2.** Update each $\sigma_t^2$, $t = 1, \dots, T$, by sampling from the mixture

$$\sigma_t^2 | \boldsymbol{\sigma^2}_{-t}, \boldsymbol{y} \sim \sum_{j \neq t} \tilde{q}_{tj} \; \delta_{\sigma_j^2}(\sigma_t^2) + \tilde{q}_{t0} \times IG\left(\nu_0 + \frac{1}{2}, \lambda_0 + \frac{(y_t - \mu)^2}{2}\right) . \tag{4.19}$$

The distribution in Eq. (4.19) corresponds to a mixture of point masses and an Inverse Gamma distribution, with weights

$$\begin{aligned} \tilde{q}_{tj} &\propto \frac{1}{\sqrt{\sigma_j^2}} e^{-\frac{(y_t - \mu)^2}{2\sigma_j^2}} \\ \tilde{q}_{t0} &\propto c \times \frac{\Gamma\left(\nu_0 + \frac{1}{2}\right)}{\Gamma(\nu_0)} \frac{2^{\nu_0 + \frac{1}{2}} (\lambda_0)^{\nu_0}}{\left[(y_t - \mu)^2 + 2\lambda_0\right]^{\nu_0 + \frac{1}{2}}} \\ &\sum_{j \neq t} \tilde{q}_{tj} + \tilde{q}_{t0} = 1 \,, \end{aligned} \tag{4.20}$$

where $\Gamma$ is the Euler Gamma function.

**3.** Resample the unique values of $\boldsymbol{\sigma^2}$ from

$$\sigma_d^{2*} \sim IG\left(\nu_0 + \frac{\mid S_d \mid}{2}, \lambda_0 + \sum_{t \in S_d} \frac{(y_t - \mu)^2}{2}\right) \quad d = 1, \dots, |\rho| \,. \tag{4.21}$$

## 4.4 Posteriors and cluster dynamics

In this section we will apply the models described before to the analysis of real financial series, describing the posterior distributions of the parameters obtained as the output of the Gibbs schemes.

We considered four time series, each one of depth $T = 1000$, corresponding to the daily log-returns of the MiB30 index (Milan Stock Exchange) and its three components Lottomatica (LTO.MI), Mediobanca (MB.MI) and Snam Rete Gas (SRG.MI), from April 2004 to March 2008. These components were the ones exhibiting the highest excess of kurtosis, for which we know a priori that usual approaches based on Normal distributions do not provide a satisfactory description. We carried out $N = 10000$ sweeps of the Gibbs algorithms, discarding the first 1000 realizations to guarantee the Markov chain thermalization. The prior parameters were fixed as in Table 4.1. The choice of $m$ can be justified by the fact that the mean of daily returns is usually of order $10^{-3}$ and it can be neglected a priori, while the somewhat large value of $\tau_0^2$ was chosen to avoid making the likelihood of $\mu$ given $\sigma^2$ too informative. Moreover, since the variability of stock returns is usually extremely small, we fixed $\lambda_0$ and $\nu_0$ in order to have prior expectation and variance for $\sigma^2$ equal to 0.01. Finally, since the cohesion parameter $c$ controls the prior probability of generating new clusters, after a numerical sensitivity analysis, we chose $c = 1$, since we found that such a value allows an efficient updating of the partition structure during the sampling, still preventing an excessive fragmentation. A major consequence of this choice is that the posterior partitions always exhibit a dominant cluster including the largest part of the observations, thus allowing to define coherently the concept of outlying experimental points (see next section). Before considering the posterior distributions, it is worth looking at

| $N$ | $m$ | $\tau_0^2$ | $\lambda_0$ | $\nu_0$ | $c$ |
|---|---|---|---|---|---|
| 10000 | 0 | 1000 | 0.0101 | 2.01 | 1 |

Table 4.1: Prior parameters for the Gibbs sampling of the $\boldsymbol{\mu}$-PPM (4.12) and $\boldsymbol{\sigma}^2$-PPM (4.17) models.

how the clusters distribute during the sampling. From the expression (4.3), the prior expectation of the number of clusters can be computed

$$\langle NC \rangle = \sum_{|\rho|=1}^{T} \frac{c}{c + |\rho| - 1} \, .$$

It can be compared with the posterior average cardinality of the partitions $NC_B$ [2] as in Table 4.2, from which we see that the prior expectation is modified

[2] We put a subscript $B$ to indicate the posterior (Bayesian) value of the corresponding quantity.

| | $\langle NC \rangle$ | Number of Clusters | | Largest Cluster Weight | |
|---|---|---|---|---|---|
| | | $\boldsymbol{\mu}$-PPM | $\boldsymbol{\sigma^2}$-PPM | $\boldsymbol{\mu}$-PPM | $\boldsymbol{\sigma^2}$-PPM |
| MIB30.MI | 7.4855 | 3.11 | 3.39 | 0.986 | 0.990 |
| LTO.MI | 7.4855 | 5.02 | 4.52 | 0.963 | 0.944 |
| MB.MI | 7.4855 | 4.11 | 3.72 | 0.968 | 0.970 |
| SRG.MI | 7.4855 | 3.44 | 3.59 | 0.984 | 0.978 |

Table 4.2: Posterior average of the number of clusters and relative weight of the largest cluster for $\boldsymbol{\mu}$-PPM and $\boldsymbol{\sigma^2}$-PPM.

by the inference process, eliciting a clustering structure with a number of clusters between 3 and 4 (on average). Moreover the posterior dimension of the largest cluster is always about 99% of the sample, justifying the concept of outlying observations (see next section). In Fig. 4.1 we report the histograms of the parameters posterior distributions of the MiB30 index (the ones for the other components being similar) for the case of the $\boldsymbol{\mu}$-PPM model. We plot the distribution of the first entry of the $\boldsymbol{\mu}$ vector, its average $\bar{\mu} = \sum_{t=1}^{T} \mu_{t,B}/T$, and of the variance $\sigma^2$. Possible considerations here are highly qualitative, since the exact forms of the posterior distributions are not known analytically. We

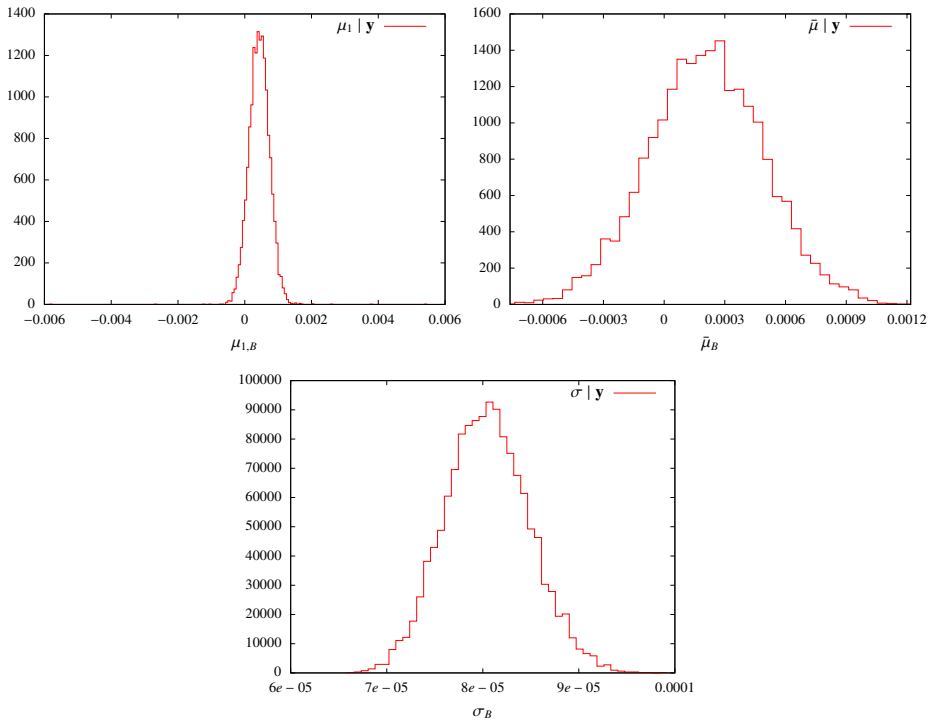



Figure 4.1: Posterior distributions in the $\boldsymbol{\mu}$-PPM model for the MiB30 index: clockwise, the firs entry of $\boldsymbol{\mu}_B$, its mean value $\bar{\mu}_B$ and the variance $\sigma_B^2$.

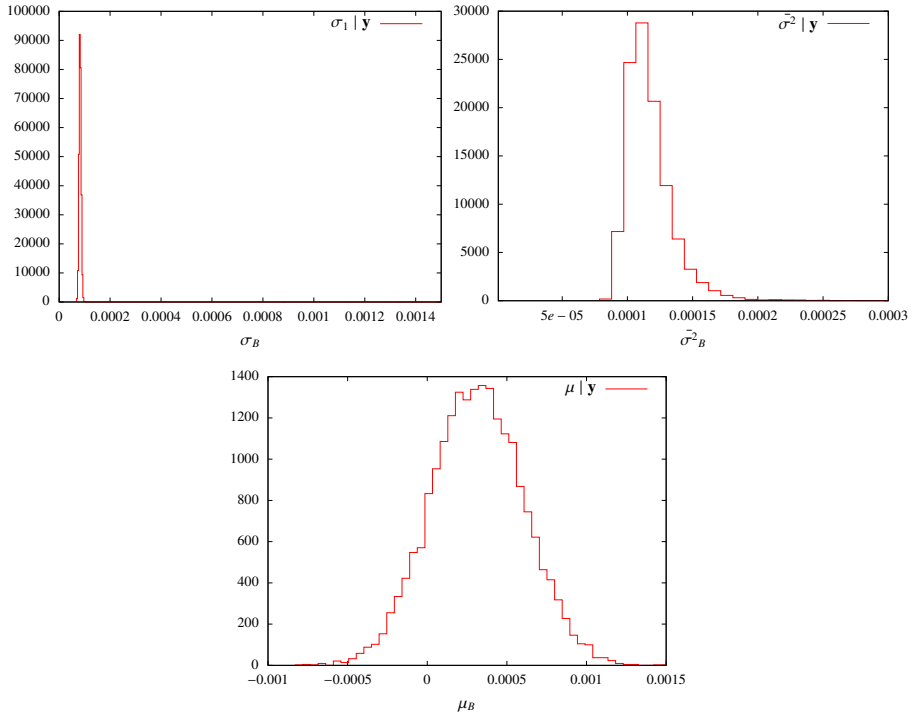



Figure 4.2: Posterior distributions in the $\boldsymbol{\sigma}^2$-PPM model for the MiB30 index: clockwise, the firs entry of $\boldsymbol{\sigma^2}_B$, its mean value $\bar{\sigma^2}_B$ and the return mean $\mu_B$.

observe the bell shaped profile of the distribution of $\bar{\mu}$, which could be expected since for every iteration of the Gibbs sampler $\bar{\mu}$ is a sum of Gaussian variables as stated in (4.14) and (4.16). We also note that the distribution of $\sigma_B^2$ lose its prior Inverse Gamma shape, despite the fact that the likelihood (4.13) was Inverse Gamma itself.

Fig. 4.2 shows the posteriors of $\sigma_{1,B}^2$, its average $\bar{\sigma^2}_B = \sum_{t=1}^{T} \sigma_{t,B}^2/T$ and $\mu_B$ for the case of the $\boldsymbol{\sigma}^2$-PPM model with reference to the MiB30 index. Here, the posteriors of the variances retains a clearly non Gaussian, Inverse Gamma like shape. This can be easily understood as the large value of $T$ makes the likelihood $\mu \mid (\sigma_1^{2*}, \ldots, \sigma_{|\rho|}^{2*})$ very narrow and peaked around $m$, so that the $\mu$ entering the likelihood of $y_t$ in (4.12) is nearly constant. Since the Inverse Gamma prior of $\sigma_t^2$ is conjugated [3] to that likelihood, we expect the posterior to be approximately Inverse Gamma either. Table 4.3 shows the posterior estimates for the MiB30 index, as obtained taking the expected value of the corresponding posterior distributions, to be compared with the frequentist values $\hat{\mu}$ and $\hat{\sigma^2}$. Errors on the Bayesian estimates have been computed by searching for the most symmetric interval in the support of the posterior distributions which encloses the assigned confidence level (CL), here 68%, and they have to be considered as "credibility" intervals from the point of view of

[3] When the posterior distribution $p(\phi|y)$ is in the same class of the prior $p(\phi)$, they are said to be *conjugate distributions* and the prior is called a *conjugate prior* for the likelihood.

| | $\hat{\mu}(10^{-4})$ | $\hat{\sigma^2}(10^{-5})$ | $\mu_B(10^{-4})$ | $\bar{\mu}_B(10^{-4})$ | $\sigma^2_B(10^{-5})$ | $\bar{\sigma^2}_B(10^{-5})$ |
|---|---|---|---|---|---|---|
| $\boldsymbol{\mu}$-PPM | $2.05^{+2.68}_{-2.68}$ | $7.20^{+0.32}_{-0.32}$ | - | $2.08^{+2.80}_{-2.80}$ | $8.05^{+0.43}_{-0.37}$ | - |
| $\boldsymbol{\sigma}^2$-PPM | $2.05^{+2.68}_{-2.68}$ | $7.20^{+0.32}_{-0.32}$ | $3.14^{+2.89}_{-2.86}$ | - | - | $11.76^{+1.38}_{-1.63}$ |

Table 4.3: Posterior estimates of parameters for the MiB30 index, compared with the frequentist values $\hat{\mu}$ and $\hat{\sigma^2}$ (errors at 68% confidence level).

Bayesian statistics. Even though the Bayesian estimates for the mean of the data are in agreement with the sample statistics, this is knot true for the variances, whose posterior values are noticeably larger than the frequentist one, especially for the $\boldsymbol{\sigma}^2$-PPM case. This evidence may be interpreted as a clue of the fact that the partition structure on the variances is statistically relevant and that the assumption of identically distributed random variables for the sample at hand is not fully justified. This is in agreement with the evidence of regime switching in the high frequency volatility [16, 17].

## 4.5 Outliers identification

The modeling described in this chapter is greatly motivated by the ability of PPMs to identify clusters of observations sharing some statistical properties, such as the mean or the variance for the $\boldsymbol{\mu}$-PPM and $\boldsymbol{\sigma}^2$-PPM models respectively. The concept of a partition structure brings in the notion of outlying observation as an experimental point belonging to a small cluster when the majority of the observations belong to a large dominant group.

A natural choice is to model outliers as a shift in the mean of the data and for this reason, we focus on the $\boldsymbol{\mu}$-PPM model. Indeed, if we take a look at the expressions of the probabilities $q_{tj}$ and $q_{t0}$ in (4.15) we see that the extent of that shift is the criterion used in the $\boldsymbol{\mu}$-PPM approach to induce a new cluster on the vector of returns: the larger the difference $(y_t - \mu_j)$ the smaller the probability $q_{tj}$ of a reshuffling and, from normalization arguments, the larger the probability $q_{t0}$ of generating a genuine new value. Along this line, we present an efficient algorithm [5, 6] allowing to select the partition that best separates the main group of standard observations from one or more groups of atypical points.

Every partition corresponds to a different scenario, and we can think of the best model as the one maximizing a suitably chosen score function. In [83] it is suggested that a good choice for the specific problem at hand is the following

$$SC(\rho) = \frac{k_1}{T} \parallel \hat{\boldsymbol{\mu}}_B(\boldsymbol{y}) - \hat{\boldsymbol{\mu}}_\rho(\boldsymbol{y}) \parallel^2 + k_2 \left[\hat{\sigma}^2_B(\boldsymbol{y}) - \hat{\sigma}^2_\rho(\boldsymbol{y})\right]^2 + (1 - k_1 - k_2)|\rho|, \quad (4.22)$$

where $\parallel \cdot \parallel$ is the Euclidean norm and $k_1$, $k_2$ are non-negative parameters with $k_1 + k_2 \leq 1$. The subscript "$B$" stands for the Bayesian (posterior) estimate of the corresponding parameter while the subscript "$\rho$" for its estimate

conditioned on a given partition. Formally, we have that $\hat{\boldsymbol{\mu}}_B(\boldsymbol{y}) = \langle \boldsymbol{\mu} | \boldsymbol{y} \rangle$, $\hat{\boldsymbol{\mu}}_\rho(\boldsymbol{y}) = \langle \boldsymbol{\mu} | \boldsymbol{y}, \rho \rangle$ and analogously for $\hat{\sigma}^2_B(\boldsymbol{y})$ and $\hat{\sigma}^2_\rho(\boldsymbol{y})$. The function (4.22) combines the estimation of parameters and partition selection problems, $k_{1,2}$ being the weights which control their priority.

The Bayesian estimates are obtained via the Gibbs sampling algorithm of Eq.s (4.13)-(4.16). The evaluation of $\hat{\boldsymbol{\mu}}_\rho(\boldsymbol{y})$ and $\hat{\sigma}^2_\rho(\boldsymbol{y})$ requires the use of the Gibbs scheme in a simpler form: a fixed partition is first selected by means of the search algorithm detailed below, and then we sample iterating step **1** and **3** but skipping step **2** since $\rho$ is fixed. An exhaustive search on the space of all possible partitions is infeasible. In fact, for a set with $T$ elements, the number of all possible partitions is equal to the *Bell number* of order $T$, recursively defined by $B(T+1) = \sum_{k=0}^{T} \binom{T}{k} B(k)$, with $B(0) = 1$, which is an extremely large number even for moderate values of $T$. For this reason we restrict to a tractable subset, performing an exhaustive search over the partitions with cardinality up to three, selected as follows.

i) Let $\boldsymbol{\mu}_B = (\mu_1, \ldots, \mu_T)$ be the vector of the Bayesian estimates of the return means, and $\widetilde{\boldsymbol{\mu}}_B = (\widetilde{\mu}_1, \ldots, \widetilde{\mu}_{\widetilde{T}})$ be the vector of the unique entries of $\boldsymbol{\mu}_B$ sorted in increasing order, with $\widetilde{\mu}_1 \doteq \min(\mu_t)$, $\widetilde{\mu}_{\widetilde{T}} \doteq \max(\mu_t)$, and $\widetilde{T} \leq T$.

ii) We perform our search of the optimal partition over the set of the partitions $\rho = \{S_1, S_2, S_3\}$, where $S_1 = \{t : \mu_t < \widetilde{\mu}_i\}$, $S_2 = \{t : \widetilde{\mu}_i \leq \mu_t \leq \widetilde{\mu}_j\}$, and $S_3 = \{t : \mu_t > \widetilde{\mu}_j\}$ with $i, j = 1, \ldots, \widetilde{T}$. The optimal partition is the one for which the score function (4.22) achieves its maximum.

When $i \neq 1$ and $j \neq \widetilde{T}$, $\rho$ has 3 distinct clusters and the indexes in $S_1$ and $S_3$ may be considered as representative of those returns being in the "left tail" and the "right tail" of the empirical distribution of $\boldsymbol{y}$. In a naive way, $S_2$ would correspond to elements occupying the central region of this distribution. When $i = 1$ and $j = \widetilde{T}$, the algorithm explores the trivial partition, $S_1 = S_3 = \{\emptyset\}$, while if $i = 1$ or $j = \widetilde{T}$ we have partitions with only two clusters. However, up to now our search is not exhaustive yet since it exists an alternative way of generating partitions of cardinality 2. Given any cardinality-3 partition $\rho$, we need also to consider the new partition $\hat{\rho} = \{S_1, S_2\}$, obtained identifying the two "tails" cluster, namely posing $S_1 \doteq S_1 \cup S_3$. Once the optimal partition has been found, we identify the outliers with those elements in $\boldsymbol{y}$ whose indexes belong to the sets with lowest cardinality.

In Fig. 4.3 we report the results of this outliers detection algorithm for the considered time series [4]. The triangles represent "positive" outliers, corresponding to extreme gains, while circles represent extreme losses. We see that the points detected as atypical always match spikes in the return series. At the same time, not all the spikes are detected as outliers. This reflect the very nature of parameter PPMs modeling: large returns which are extreme

[4] For a detailed table reporting the composition of the optimal partition see [5]

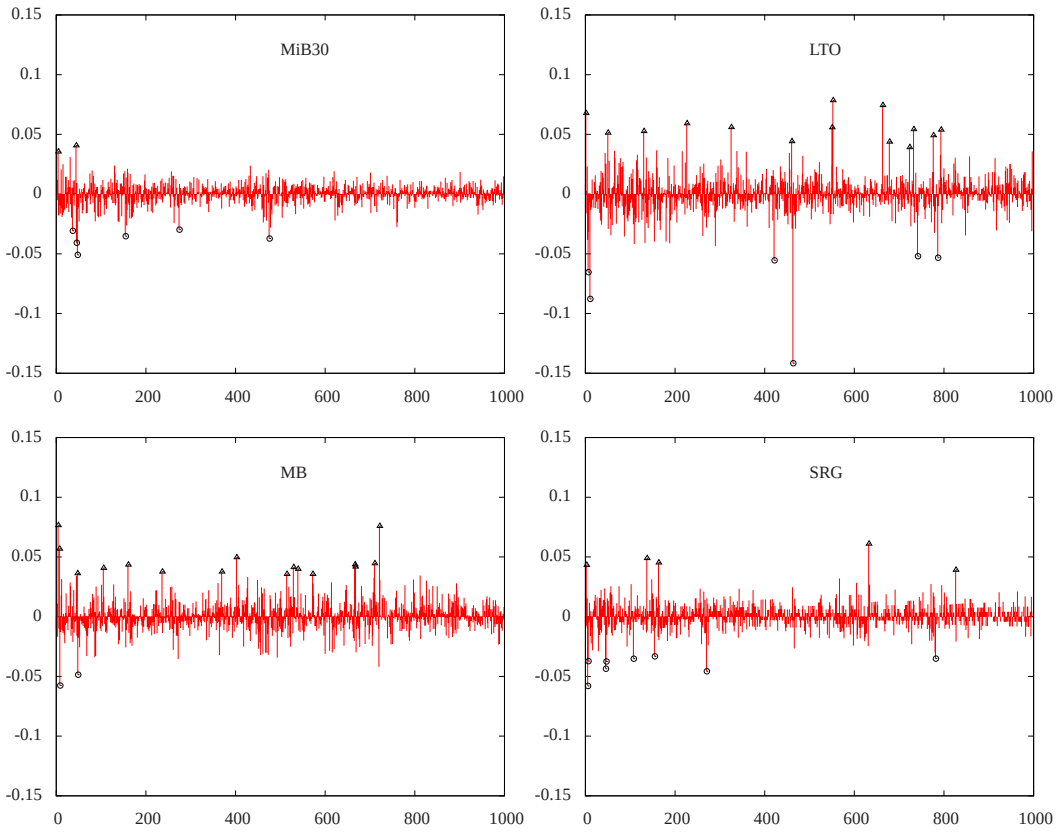



Figure 4.3: Detected outliers for the returns series of MiB30 index and its components Lottomatica, Mediobanca and Snam Rete Gas. From [5].

realizations of a Gaussian variable whose mean value is not atypical are not detected as atypical either.

Chapter 5

# Option pricing and market risk evaluation

So far we described two different, complementary approaches to the modeling of financial series focusing on some specific examples. First we considered SVMs, diffusion processes in which price returns are driven by a stochastic volatility, in agreement with empirical evidences about the non constant behavior of the amplitude of price fluctuations. These models focus on the price dynamics and describe its random time evolution, aiming at the more realistic description of the statistical properties of financial time series considered as a realization of the path of a stochastic process. Then, we explored a completely different approach, based on Bayesian inference which updates the degree of belief in some hypothesis reacting to the empirical evidence; within this approach, exploiting the specific framework of PPMs, it is possible to incorporate the hypothesis of a non homogeneity of the sample, eliciting the partition structure which is more likely to be realized given the observations, and to take into account this information when estimating the parameters of interest if needed. We also saw that the Bayesian PPM approach modifies the frequentist sample estimates, supporting the need to consider the different statistical properties of the data, especially as far as the volatility is concerned.

In this chapter we will discuss some original financial applications of these approaches to problems of major interest in finance: the rational evaluation of option prices and the estimation of market risk exposure. In the first part of the chapter we will review the basic definitions and results on these subjects along with some original methodological achievement. We will define an option contract, and the celebrated price formula by Black, Scholes and Merton will be recalled as well. We will introduce the concepts of arbitrage and market efficiency and completeness, the fundamental martingale pricing theorem providing a general framework for evaluating financial derivatives in efficient markets, the notion of implied volatility and its empirical features pointing toward the lacking of realism of the Black and Scholes (B&S) assumptions. We will see how it is possible to derive the martingale dynamics for SVMs, propos-

ing a useful way to exploit the fact that the market model is incomplete, and we will show that the knowledge of the model CF can be exploited usefully for pricing purposes. As far as risk estimation is concerned, we will define the most commonly used estimators for evaluating the risk associated to price fluctuations, the Value-at-Risk (VaR) and the Expected Shortfall (ES), and we will detail an original computational methodology for their efficient evaluation based on the Generalized Fourier Transforms (GFTs). This methodology applies both to the case of single assets and portfolios and it also allows for the evaluation of the sensitivities with respect to the portfolio composition. The non parametric historical approach will be also revised along with some useful backtesting procedures to test the robustness of the models used for risk assessment. Finally we will introduce the Delta-Gamma-Normal (DGN) model, a benchmark framework to account for risk of portfolios containing derivative instruments like options.

The second part of the chapter will be devoted to describe some original results about the previously introduced topics. Firstly we will consider the pricing problem for the linearized exponential Ornstein-Uhlenbeck model introduced in Chapter 3. We will describe the form of the martingale dynamics for this model and we will propose an efficient procedure to estimate its parameters from real market quotes exploiting the analytical expression of the CF. The implied volatility surface predicted by the linear model and those obtained for the exponential Ornstein-Uhlenbeck and the Stein-Stein models will be also compared in the light of an accurate propagation of the statistical uncertainty associated to the parameters estimation and to the numerical implementation. Then we will exploit the $\boldsymbol{\mu}$-PPM and $\boldsymbol{\sigma}^2$-PPM models of Chapter 4 to obtain Bayesian estimates for the VaR of single assets from the Milan Stock Exchange and for the MiB30 index. At this aim, we will devise suitable expressions for the VaR in order to account for the contribution of outlying clusters of observations. Our Bayesian estimates will be also compared with the ones obtained in the frequentist parametric approach proposed in [25] where the return PDF is fitted with a fat tailed generalized Student-$t$. As a conclusion, our GFT approach to risk, presented in the first part, will be applied to the case of the DGN model, whose CF is known in closed form. We will detail a case study, obtaining Fourier estimates of the VaR, the ES and their sensitivities for a fictitious portfolio and we will check the results against the historical estimates from a synthetic, Monte Carlo generated, sample of portfolio value variations.

## 5.1 Martingale pricing and risk measures

Derivative contracts are financial instruments whose value $O_t$ at time $t$ depends on the value $S$ of some fundamental asset. The latter is called the *underlying*, and it may be a stock, an exchange rate or something else, and the value of the contract possibly depend on the whole past history of the underlying, that is $O_t = O(t, S_{t' \leq t})$. Here we focus on *European plain vanilla options*.

An option gives its *holder* the right, not the due, to buy from or sell to the *writer* a unit (conventionally) of the underlying stock at the price $K$ specified by the contract. In the first case we have a *call* option, while a *put* in the second. The exercise price is named the *strike* price and the European option can be exercised only at the $T$, the *maturity*, which is also part of the contract specifications. From these definitions, it follows that the value of the contract at $t = T$, its *payoff*, only depends on the final price of the underlying

$$h(S_T) = \begin{cases} (S_T - K)^+ & \text{Call} \\ (K - S_T)^+ & \text{Put} \end{cases} . \tag{5.1}$$

In 1973, F. Black and M. Scholes [85], and independently R. Merton [86], published their theory of rational option pricing, based on the assumption of a GBM dynamics for the underlying asset and a few other assumptions about the financial market. Their famous formula for the price of a European option can be seen as a specific result of a more general approach to derivative pricing which is known as *martingale pricing* or *no arbitrage pricing*.

This approach find its roots in the hypothesis of efficient market, formulated in 1965 by P. Samuelson [87] which we state as follows: a market is efficient when all the information relevant for the transactions is gathered and processed instantaneously by all market participants. This is a reasonable approximation for liquid markets and it translates in the absence of *arbitrage opportunities*, that is a trading strategies able to guarantee a riskless profit with a null initial investment. The martingale pricing theory relies on the following fundamental theorem of mathematical finance, see for instance [88]. We call a *numéraire* $N(t)$ the value of any freely tradable and positively valued asset. The martingale pricing theorem states that, in absence of arbitrage, there exists a probability measure $\mathbb{Q}$ such that for all other securities, the value $O_t$ of the security divided by the numéraire is a martingale, that is

$$\frac{O_t}{N_t} = \left\langle \frac{O_T}{N_T} \middle| \mathcal{I}_t \right\rangle_{\mathbb{Q}} \quad \text{for all } T > t$$

where $\langle . \rangle_{\mathbb{Q}}$ denotes the expectation with respect to $\mathbb{Q}$ and $\mathcal{I}_t$ stands for the information available at time $t$ about the process. In our case $O_t$ is the value of the Call/Put option and $O_T = h(S_T)$ is its payoff. We will discuss the case when $S_t$ is modeled with a SVM, so that $S_t$ is a Markovian process. In this case, the theory of Markovian processes tells us that conditioning with respect to $\mathcal{I}_t$ is equivalent to conditioning only to the value of the underlying process at $t$. We can also introduce the concept of a bank account $B(t) = e^{r(t-t_0)}$ which is the value at $t$ of a unit of currency invested in a bank deposit at time $t_0$ when the continuous interest rate is constant and equal to $r$. This (deterministic) process can be conveniently chosen as our numéraire, so that $N_t = B_t$; if so, the corresponding measure $\mathbb{Q}$ is named *risk neutral measure* and the value of the option at present (conventionally $t = 0$) reads

$$O_0 = e^{-rT} \langle h(S_T) | S_0 \rangle_{\mathbb{Q}} . \tag{5.2}$$

From this point of view, the fundamental issue of option pricing consists of finding the risk neutral measure, or more generally a way to compute the expectation (5.2) under such a measure. When $\mathbb{Q}$ is unique, the market is said to be *complete*; in this case it can be proved [57] that the value of the option can be replicated at any time by a portfolio set up with only the underlying $S$ and the bank account $B$. When the market is incomplete, and we will see it is the case of SVMs, this is not true anymore. In the B&S theory, the underlying dynamics under the risk neutral measure reads

$$S_t = S_0 e^{\left(r-\frac{\sigma^2}{2}\right)t+\sigma(W_t^*-W_0^*)} \tag{5.3}$$

where $W_t^*$ is a standard Brownian motion under $\mathbb{Q}$. The previous expression is nothing but the solution of the GBM equation (1.3) with the drift coefficient $\mu$ replaced by the risk free interest rate. It follows that the risk neutral PDF is the Log-Normal with mean $\langle S_t \rangle_{\mathbb{Q}} = S_0 e^{rt}$ and variance $\mathrm{Var}_{\mathbb{Q}}[S_t] = S_0^2 e^{2rt}(e^{\sigma^2 t} - 1)$ and the expectation (5.2) evaluated for this PDF gives us the celebrated B&S formula for the price of a European Call option

$$C_0^{B\&S} = S_0 \Phi(d) - K e^{-rT} \Phi(d - \sigma\sqrt{T}) \,, \tag{5.4}$$

where $\Phi$ stands for the cumulative distribution function of the standard Normal distribution and $d$ is defined as

$$d = \frac{\log(S_0/K) + \left(r + \frac{\sigma^2}{2}T\right)}{\sigma\sqrt{T}} \,. \tag{5.5}$$

Historically, option contracts were introduced as a sort of insurance against adverse moves of prices: they give the right to trade a stock at a fixed price, which right is to be exercised only if the price difference (5.1) is favorable depending on whether we are willing to buy the stock or we own it. Fluctuations of financial prices are what we call *market risk* and a huge number of papers, either from the practitioners and the academics community, has been devoted to define suitable risk level estimators and efficient methodologies for computing them. A widely considered risk measure is the so called Value-at-Risk (VaR), defined to be the maximum potential loss in the value of an asset (or portfolio) at given confidence level $\alpha^* = 1 - \mathcal{P}^*$. The parameter $\mathcal{P}^*$ is called *significance level* and common choices are $\mathcal{P}^* = 1\%, 5\%$. Indicating with $p(V)$ the PDF of the price variation of an asset or portfolio whose present value is $W_0$, and with $\tilde{p}(R)$ that of the returns $R = V/W_0$, the VaR $\Delta^*$ is implicitly defined by

$$\mathcal{P}^* = \int_{-W_0}^{-\Delta^*} p(V)\, dV = S_0 \int_{-\infty}^{-\Delta^*/S_0} \tilde{p}(R)\, dR \,. \tag{5.6}$$

When $R \sim N(\mu, \sigma^2)$, we obtain the usual form of the Gaussian VaR

$$\Delta^* = -\mu W_0 + \sigma W_0 \sqrt{2}\, \mathrm{erfc}^{-1}(2\mathcal{P}^*) \,. \tag{5.7}$$

The VaR is important for regulatory purposes since Basel II [89] accords on bank supervision impose institutions to meet strict capital requirements based on VaR estimates, even though they do not specify any procedure to compute them, leaving financial actors free to employ their own internal models. Despite of the interest by regulators, the VaR gives no information about the extent of the loss when the threshold is exceeded; this motivated the introduction of a more informative estimator, called the Expected Shortfall (ES), which is defined as the normalized average of the losses over threshold:

$$E^*(\mathcal{P}^*) = -\frac{1}{\mathcal{P}^*}\int_{-W_0}^{-\Delta^*} Vp(V)\,dV\,. \tag{5.8}$$

There are situations in which the variation $V$ is modeled by means of unbounded distributions (Gaussian for instance), even though that quantity is limited from below by $-W_0$, and the limit $-W_0 \to -\infty$ is implicit. Since $\mathcal{P}^*$ is small in practical applications, we can think of $-W_0$ as standing on the far left tail of the distribution and the error induced by this limit to be small.

### 5.1.1 Implied volatility smile

From the B&S formula (5.4) it stems a concept of the volatility quite different with respect to that encountered so far, that of *implied volatility*. It is defined as the value $\sigma$ to be plugged into in the B&S formula to match the price of the option quoted by the market

$$C^{market} = C_0^{B\&S}(S_0, K, T, r; \sigma) \Rightarrow \sigma_{imp}\,.$$

If the B&S theory was meaningful we would expect all the options written on the same underlying to have the same value of $\sigma_{imp}$ and this value to be equal to the volatility $\sigma$ entering the GBM dynamics of the underlying. Indeed, this is not the case. It is well known that options with the same underlying stock but with different strikes or maturities are quoted differently by the market, and the corresponding surface of implied volatility has the typical shape shown in Fig. 5.1. In particular, the profile of $\sigma_{imp}$ as a function of the variable $\log(S_0/K)$ forms the so called *volatility smile*. While this dependency can be considered to reflect investor's expectation about the future trend and riskiness of the underlying, the dependency of $\sigma_{imp}$ on the expiry date $T$ reflects the non constant nature of the volatility itself, in agreement with the evidences outlined in Chapter 1.

It exists an interesting connection between the implied volatility smile and the variance $\sigma_\tau^2$, skewness $\zeta_\tau$ and kurtosis $\kappa_\tau$ of the risk neutral PDF at time $T = t_0 + \tau$, as provided by the following expression

$$\sigma_{\mathrm{imp},\tau}(d_1) \simeq \frac{\sigma_\tau}{\sqrt{\tau}}\left[1 - \frac{\zeta_\tau}{3!}d_1 - \frac{\kappa_\tau}{4!}(1 - d_1^2)\right], \tag{5.9}$$

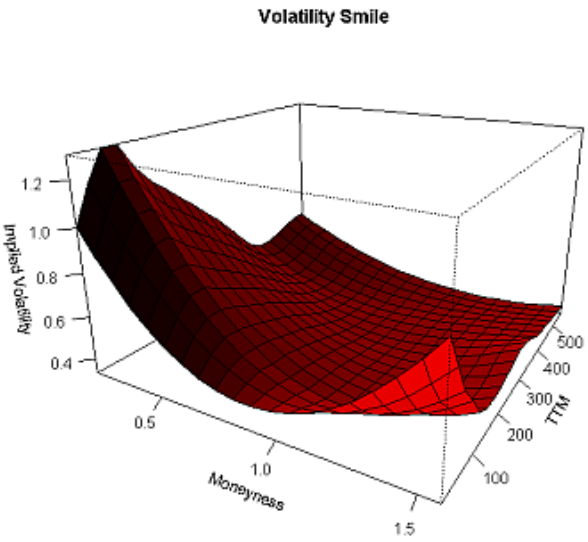



Figure 5.1: Typical profile of an implied volatility surface.

where $\sigma_{\text{imp},\tau}$ is the B&S implied volatility for the time to maturity $\tau$ and

$$d_1(\tau, K) \doteq \frac{\log(S_0/K) + r\tau + \sigma_\tau^2/2}{\sigma_\tau}\,.$$

The previous relation shows that the smile effect and its asymmetry, can be considered as a clue of the non Gaussianity of the risk neutral PDF. In particular, its excess of kurtosis is responsible for the convexity of the smile. The expression (5.9), derived in [90] (see also [16, 91]), is based on the approximation of the risk neutral PDF for the time to maturity $\tau$ in terms of a Gram Charlier cumulant expansion. It does not rely on the choice of any specific underlying dynamics and it is shown that it is effective only for $d_1 \sim 0$ and $\sigma_\tau \ll 1$, which is often realized in practice (for $\tau \simeq 1$ year, $\sigma_\tau$ ranges from 0.2 to 0.3 typically) [1].

### 5.1.2 Martingale pricing for SVMs

Now we will sketch the procedure by which the risk neutral process can be derived for the case of a SVM (for further details and a more formal treatment see for instance [57]). We will focus on the case in which the secondary process

[1] This limitation is due to the fact that the inversion of the B&S formula, required to write (5.9), is accomplished neglecting terms of higher order in $\sigma_T$ and $d_1$.

$Y$ follows an Ornstein-Uhlenbeck process:

$$\begin{aligned} dS_t &= \mu S_t\, dt + \sigma(Y_t) S_t\, dW_{1,t} \\ dY_t &= -\alpha(Y_t - \gamma)\, dt + k\, dW_{2,t}\ . \end{aligned} \tag{5.10}$$

The martingale pricing theorem states that under the risk neutral measure $\mathbb{Q}$ the ratio of the price over the bank account numéraire, $\tilde{S}_t \doteq e^{-rt} S_t$, is a martingale. From Îto lemma applied to $\tilde{S}$ we have

$$\begin{aligned} d\tilde{S}_t &= (\mu - r)\, \tilde{S}_t\, dt + \sigma(Y_t) \tilde{S}_t\, dW_{1,t} \\ &= \sigma(Y_t) \tilde{S}_t \left[ dW_{1,t} + \frac{\mu - r}{\sigma(Y_t)}\, dt \right] \doteq \sigma(Y_t) \tilde{S}_t\, dW^*_{1,t}\,. \end{aligned} \tag{5.11}$$

Since martingales are stochastic integrals with respect to Wiener processes, $\mathbb{Q}$ must be a measure under which the process $W^*_{1,t}$ is a Wiener process. Moreover, since a shift in $W_{2,t}$ does not influence the drift of $\tilde{S}$, we can also admit a transformation of the type

$$W^*_{2,t} = W_{2,t} + \int_0^t \eta(S_s, Y_s, s)\, ds\,, \tag{5.12}$$

where the process $\eta$ may depend on the underlying processes but not on any of parameters specifying the option contract and, apart from some integrability conditions, it is an arbitrary function. Under quite general assumptions about the form of $\sigma(Y)$, the existence of such a measure is guaranteed by the Girsanov theorem (see for instance [92]) so that the complete risk neutral dynamics can be written as

$$\begin{aligned} dS_t &= r S_t\, dt + \sigma(Y_t) S_t\, dW^*_{1,t} \\ dY_t &= -\left[\alpha(Y_t - \gamma) + \eta(S_t, Y_t, t)\right]\, dt + k\, dW^*_{2,t}\ . \end{aligned} \tag{5.13}$$

We see that under $\mathbb{Q}$, the $\mu$ is replaced by the risk free rate $r$ and the process $\eta$ modifies the drift term of $Y$, which is why $\eta$ is called the *market price of volatility risk*. It is important to notice that for every choice of $\eta$ we have a different risk neutral measure, $\mathbb{Q} = \mathbb{Q}^\eta$, so that SVMs are incomplete market models. By the way, for the B&S model we have a constant volatility, $\sigma_t = \sigma$, and the solution of the first of (5.13) is just the Log-Normal process (5.3) whose expectation leads to the B&S formula. In this case the only source of randomness is the $S$ process and no arbitrariness enters the risk neutral dynamics, leading to a unique $\mathbb{Q}$ measure, which makes the market complete.

The practical approach to the problem of fixing $\eta$ is to suppose that a unique choice is made by the market itself. Naively, we may think that the market price options under the measure corresponding to that choice and, even though the process $\eta$ never enters the "objective" dynamics (5.10), we expect it to be reflected by the values of options. In practical applications $\eta$ is often assumed as a constant to be estimated jointly with the other model parameters

from the implied volatility surface. However, choosing a constant $\eta$ may sound quite restrictive; a somewhat general assumption [8, 66] is to assume a linear dependence on $Y$

$$\eta(S_t, Y_t, t) = \eta_0 + \eta_1 Y_t \,. \tag{5.14}$$

In this case, the risk neutral dynamics of $Y_t$ becomes

$$dY_t = -\tilde{\alpha}\,(Y_t - \tilde{\gamma})\;dt + k\,dW^*_{2,t}\,, \tag{5.15}$$

where we defined the risk neutral versions of the relaxation coefficient $\tilde{\alpha}$ and stationary mean value $\tilde{\gamma}$:

$$\alpha \to \tilde{\alpha} = \alpha + k\eta_1\,, \quad \tilde{\gamma} = \frac{\alpha\gamma - k\eta_0}{\tilde{\alpha}}\,.$$

The crucial point to be observed is that this choice for $\eta$ leaves us with the same Ornstein-Uhlenbeck dynamics (5.15) for $Y$ under $\mathbb{Q}$ either.

### 5.1.3 Characteristic function machinery

As we saw in Chapter 3, it is frequent the case when the return PDF associated to a SVM is not available, but we know the analytical expression of its CF. Even though we would prefer the former, the latter is extremely useful too. First of all it allows to draw exact expressions for the cumulants, which characterize the shape of the PDF; secondly, we will see that it provides a quite efficient and general way to solve the computational problems of option pricing and risk measure evaluation.

Let us start from the former. In [93] it is proved that a large class of payoff functions admits a generalized Fourier transform (GFT) in the complex plane $z \in \mathbb{C}$ for $z$ limited in a suitable strip of regularity. Let's focus on the Call payoff $h(S_T) = S_0\, e^{rT}(e^{x_T} - \tilde{K})^+$, where $x_t = \log(S_t/S_0) - rt$ and $\tilde{K} = K\, e^{-rT}/S_0$; the GFT of the function $\tilde{h}_{\tilde{K}}(x) = (e^x - \tilde{K})$ reads

$$\hat{h}(z)_{\tilde{K}} = -\frac{\tilde{K}^{1+iz}}{z^2 - iz}\,, \quad z \in \mathcal{S}_h \tag{5.16}$$

where $\mathcal{S}_h \doteq \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 1\}$ and $\tilde{h}$ is recovered integrating along a line parallel to the real axis included in $\mathcal{S}_h$

$$\tilde{h}_{\tilde{K}}(x) = \frac{1}{2\pi}\int_{-\infty+i\nu}^{+\infty+i\nu} e^{-izx}\hat{h}_{\tilde{K}}(z)\,dz\,.$$

with $z = \omega + i\nu \in \mathcal{S}$ and $dz = d\omega$. Exploiting this relation we can express the value of a European option (5.2) in the form

$$O_0 = \frac{S_0}{2\pi}\int_{-\infty}^{+\infty} p(x_T; T)\left[\int_{-\infty+i\nu}^{+\infty+i\nu} e^{-izx_T}\hat{h}_{\tilde{K}}(z)\,dz\right]\,dx_T\,.$$

Formally, we can switch the integration order and write the following evaluation formula

$$O_0 = \frac{S_0}{2\pi} \int_{-\infty+i\nu}^{+\infty+i\nu} \hat{h}_{\tilde{K}}(z) f(-z;T)\, dz\,, \tag{5.17}$$

where we defined the GFT of $p(x;t)$ as

$$f(z;t) = \int_{-\infty+i\nu}^{+\infty+i\nu} e^{izx} p(x;t)\, dx \tag{5.18}$$

assuming the integral to be well defined. Now, let's suppose we know the exact expression of the CF $f(\omega;t)$ associated to the returns PDF $p(x;t)$ and that its extension to the complex plane $f(z;t)$, $z \in \mathbb{C}$, is analytical and single valued in the neighborhood of $z = 0$. If so, a theorem by Lukacs [94] guarantees that the GFT of $p$ is well defined in a regularity strips $\mathcal{S}_p \doteq \{z \in \mathbb{C} \mid \nu_- < \mathrm{Im}(z) < \nu_+\}$, and this GFT coincides with $f(z;t)$. The same theorem states that the limits of the strips $\nu_\pm$, with $\nu_- < \nu_+$, are the imaginary parts of the singularities of $f(z;t)$ closest to the real axis, if any. So, if we define the strip $\mathcal{S}_p^* \doteq \{-\nu_+ < \mathrm{Im}(z) < -\nu_-\}$, the expression (5.18) is well defined only for $z \in \mathcal{S}_h \cap \mathcal{S}_p^*$, if this intersection is not empty. Specializing (5.17) to the case of the Call option we obtain

$$\begin{aligned} C_0 &= -e^{-D}\, \frac{S_0}{2\pi} \int_{-\infty+i\nu}^{+\infty+i\nu} e^{-izD} \frac{f(-z;T)}{z^2 - iz}\, dz \\ &= -e^{D(\nu-1)}\, \frac{S_0}{2\pi} \left\{ \int_{-\infty}^{\infty} d\omega \cos(\omega D)\, \mathrm{Re}\left[W(\omega_+) f(-\omega_+) + W(\omega_-) f(-\omega_-)\right] \right. \\ &\qquad \left. + \int_{-\infty}^{\infty} d\omega \sin(\omega D)\, \mathrm{Im}\left[W(\omega_+) f(-\omega_+) - W(\omega_-) f(-\omega_-)\right] \right\} \end{aligned} \tag{5.19}$$

where $z = \omega \pm i\nu$, $D \doteq \log(S_0/K) + rT$ and $W(\omega) \doteq [\omega^2 - i\omega]^{-1}$, with $\omega_\pm \doteq \pm\omega + i\nu$.

Now we will show how a similar approach can be extended to risk evaluation. Fourier techniques were introduced since the work of Rouvinez [95], and more recently in [96, 97], and we generalized to the case of non linear portfolios in [10]. The expressions (5.6) and (5.8) defining the VaR and ES can be managed with the same formalism leading to Eq. (5.19). Indeed, if $f(z)$ is the GFT of the PDF for the variation $V$ of an asset value, we can express the VaR (5.6) as

$$\mathcal{P}^* = \frac{1}{2\pi} \int_{-\infty+i\nu}^{+\infty+i\nu} dz\; f(z) \left( \int_{-\infty}^{-\Delta^*} dV e^{-izV} \right)\,, \tag{5.20}$$

and the convergence of the innermost integral is guaranteed if we fix $\nu \in (0, \nu^+)$. With this choice, the previous expression readily reduces to

$$\mathcal{P}^* = \frac{i}{2\pi} \int_{-\infty+i\nu}^{+\infty+i\nu} dz\; \frac{f(z)}{z}\; e^{-iz\Delta^*} = \frac{e^{-\nu\Delta^*}}{\pi} \mathrm{Re}\left[ \int_0^{+\infty} d\omega\; \frac{f(\omega+i\nu)}{\nu - i\omega}\; e^{i\omega\Delta^*} \right]. \tag{5.21}$$

Analogous considerations allow us to write the following Fourier representation for the ES, see [10]

$$E^*(\mathcal{P}^*) = \Delta^* - \frac{e^{-\nu\Delta^*}}{\pi\mathcal{P}^*}\mathrm{Re}\left[\int_0^{+\infty} d\omega \frac{f(\omega+i\nu)}{(\omega+i\nu)^2}\, e^{i\omega\Delta^*}\right]. \tag{5.22}$$

Since the only information required is the CF, the above Fourier machinery is a powerful tool, allowing to generalize the evaluation of risk measures to dynamic models such as those SVMs [97] whose CF is known in closed form, such as the Heston or the S2 model. When the value variation $V$ involved in previous equations is associated to a combination of different assets, a portfolio, the associated CF will depend on the set $\beta = \{\beta_1, \beta_2, \dots\}$ of the parameters defining its composition (see the last section of this chapter). From a practical point of view, after estimating the portfolio VaR for a given $\mathcal{P}^*$, it is important to evaluate the risk measure sensitivities to variations of the weights $\beta_i$. The only ingredient required here are the derivatives $\partial f/\partial\beta_i$, and Fourier representations can be drawn also for sensitivities. Indeed, after differentiating (5.20) at constant $\mathcal{P}^*$ and reminding the definition of GFT of $p(V)$, the VaR sensitivities reduce to a single Fourier integral

$$\frac{\partial\Delta^*}{\partial\beta} = \frac{e^{-\nu\Delta^*}}{\pi p(-\Delta^*)}\mathrm{Re}\left[\int_0^{+\infty} \frac{d\omega}{\nu-i\omega}\frac{\partial f(\omega+i\nu)}{\partial\beta}\, e^{i\omega\Delta^*}\right], \tag{5.23}$$

where the value of the PDF in $\Delta^*$ is computed in terms of its GFT

$$p(-\Delta^*) = \frac{1}{2\pi}\int_{-\infty+i\nu}^{+\infty+i\nu} e^{iz\Delta^*} f(z)\, dz\,.$$

Finally, recalling again the relation (5.6) between $\mathcal{P}^*$ and $\Delta^*$, differentiation of the expression (5.22) gives us analogous expressions for the ES derivatives

$$\frac{\partial E^*(\mathcal{P}^*)}{\partial\beta} = -\frac{e^{-\nu\Delta^*}}{\pi\mathcal{P}^*}\mathrm{Re}\left[\int_0^{+\infty} \frac{d\omega}{(\omega+i\nu)^2}\frac{\partial f(\omega+i\nu)}{\partial\beta}\, e^{i\omega\Delta^*}\right]. \tag{5.24}$$

It is worth noticing that this approach someway unify the problems of option pricing and risk measure estimation, both involving expectations of suitable payoffs. Moreover, when the CF is the only information available it proves to be highly efficient, since it reduces to a single Fourier integration what would require a double numerical integration otherwise, the first over the probability space and the second over the Fourier variable.

### 5.1.4 Historical approach to risk measures

A widely exploited, non parametric approach to risk assessment is the historical one, consisting of estimating the empirical quantiles from a sample of observations of asset value changes. Let us suppose we have a sample (real or simulated) of length $T$ of price changes, $V_{i=1,\dots,T}$. If $\tilde{V}_t$ are the entries of

$V_t$ sorted in ascending order, and assuming $t^* = T \times \mathcal{P}^*$ to be integer, the historical VaR at the significance level $\mathcal{P}^*$ is defined as [2]

$$\Delta_H^* = -\tilde{V}_{t^*} \ . \tag{5.25}$$

The ES is obtained as the average on the left tail of the empirical distribution

$$E_H^* = -\frac{1}{t^*} \sum_{t=1}^{t^*} \tilde{V}_t \ . \tag{5.26}$$

Confidence intervals for VaR can be derived from a basic result of order statistics. Let us consider the sample of $T$ i.i.d. variables $V_i$ and indicate with $Q_p$ the $p$-th percentile of their distribution; then, the probability of $Q_p$ being enclosed in $(\tilde{V}_{t^-}, \tilde{V}_{t^+})$ is given by the following sum of binomial probabilities [98]

$$P(\tilde{V}_{t^-} < Q_p < \tilde{V}_{t^+}) = \sum_{k=t^-}^{t^+-1} \binom{T}{k} p^k (1-p)^{T-k} \ . \tag{5.27}$$

To find the confidence interval associated to $\Delta_H^*$ for a given CL, we find indices $t^{\mp}$ satisfying

$$P(\tilde{V}_{t^-+1} < Q_{\mathcal{P}^*} < \tilde{V}_t^+) \leq CL \leq P(\tilde{V}_{t^-} < Q_{\mathcal{P}^*} < \tilde{V}_{t^+})$$

where $Q_{\mathcal{P}^*} = -\Delta^*$ and the choice between possible different pairs $(t^-, t^+)$ satisfying the previous inequality is made requiring the confidence interval to be as symmetric as possible around $\Delta_H^*$. With these positions, the historical VaR is estimated by

$$(\Delta_H^*)_{-\delta^-}^{+\delta^+} \ , \quad \begin{cases} \delta^+ \doteq -\tilde{V}_{t^-} - \Delta_H^* \\ \delta^- \doteq \ \Delta_H^* + \tilde{V}_{t^+} \end{cases} \tag{5.28}$$

with confidence level $CL$.

Since $E^*$ is a monotonously increasing function of $\Delta^*$, a lower and upper bound for ES are easily obtained by evaluating the average in equation (5.26) for $t^* = t^-$ and $t^* = t^+$ respectively. At the same CL as above, we estimate

$$(E_H^*)_{-e^-}^{+e^+} \ , \quad \begin{cases} e^+ \doteq -\left(\frac{1}{t^-} \sum_{k=1}^{t^-} \tilde{V}_k\right) - E_H^* \\ e^- \doteq E_H^* + \left(\frac{1}{t^+} \sum_{k=1}^{t^+} \tilde{V}_k\right) \ . \end{cases} \tag{5.29}$$

[2] If $T \times \mathcal{P}^*$ is not integer, $\Delta^*$ can be defined as the average $(-V_s - V_u)/2$, where $s, u$ are the two integers closest to $t^*$.

When $V$ is the variation of a portfolio value whose composition depends on parameters $\beta$, the sensitivities of $\Delta^*_H$ can be estimated approximating its derivatives, for instance with a 3 points finite difference formula

$$(\partial_\beta \Delta^*)_H \doteq \left(\frac{\partial \Delta^*}{\partial \beta}\right)_H = \frac{\Delta^*_H(\beta + \Delta\beta) - \Delta^*_H(\beta - \Delta\beta)}{2\Delta\beta} \tag{5.30}$$

where $\Delta^*_H(\beta \pm \Delta\beta)$ denotes the historical estimate for the VaR obtained after giving a positive/negative shock to the parameter $\beta$ while keeping fixed all the others[3]. The error affecting $(\partial_\beta \Delta^*)_H$ is obtained by linear propagation of the errors affecting $\Delta^*_H(\beta \pm \Delta\beta)$ through equation (5.30). An analogous procedure is applied to estimate $(\partial_\beta E^*)_H$ and its confidence interval.

### 5.1.5 VaR backtesting procedures

The Basel II accords also prescribe to asset the robustness of internal models used to compute the VaR for daily variations through backtesting procedures taking into account the number of exceptions recorded over the last year with respect to the model forecast. The most widely exploited methods are the Kupiec unconditional coverage (UC) test [99] and the conditional coverage (CC) one by Christoffersen [100].

We briefly review the logic underlying these tests since they will be applied to the VaR estimates obtained by the PPMs described in Chapter 4. If we indicate with $p$ the probability of observing an exception out of a sample of $N$ observations of an asset returns, the number of registered exceptions follows a binomial distribution $\mathrm{Bin}(N, p)$. The null hypothesis we need to test for is that $p = \mathcal{P}^*$, where $\mathcal{P}^*$ is the significance level associated to our VaR estimate. Kupiec considers the following generalized likelihood ratio

$$LR_{UC} = -2\log\left[(1-\mathcal{P}^*)^{N-n}\mathcal{P}^{*n}\right] + 2\log\left[(1-n/N)^{N-n}(n/N)^n\right] \tag{5.31}$$

where $n$ are the registered exceptions. Asymptotically, $LR_{UC}$ distributes chi-square with one degree of freedom under the null hypothesis, which allows to reject that hypothesis at 95% CL when $LR_{UC} > 3.84$. The above test statistics can be extended to test for the serial correlation of violations, introducing a state indicator $j$ which is equal to 0 when VaR is not exceeded and 1 otherwise. The Christoffersen test considers the following combined statistics

$$LR_{CC} = LR_{UC} + LR_{IND} \tag{5.32}$$

$$\begin{aligned} LR_{IND} = &-2\log\left[(1-n/N)^{N_{00}+N_{10}}(n/N)^{N_{01}+N_{11}}\right] \\ &+ 2\log\left[(1-\pi_0)^{N_{00}}\pi_0^{N_{01}}(1-\pi_1)^{N_{10}}\pi_1^{N_{11}}\right] \end{aligned} \tag{5.33}$$

[3]Obviously, this procedure for computing sensitivities usually applies to simulation studies, when time series of portfolio values can be generated with Monte Carlo techniques for every possible portfolio composition.

where $N_{ij}$ stands for the number of days in the state $j$ conditioned to the state $i$ having been realized the previous day. With the same logic $\pi_i$ is the probability of observing an exception the day after state $i$ was realized, that is $\pi_0 = N_{01}/(N_{00} + N_{01})$ and $\pi_1 = N_{11}/(N_{10} + N_{11})$. Here, the null hypothesis for the $LR_{IND}$ statistics is that violations are independent from what occurred (violation or not) the previous day. Under this hypothesis, it can be shown that $LR_{CC}$ distributes chi-square with two degrees of freedom, and the VaR model should be rejected with 95% CL when $LR_{CC} > 5.99$ (see also [101]).

### 5.1.6 The Delta-Gamma-Normal model

While the non Gaussian nature of returns distributions of single assets can be considered a quite fundamental fact which is intrinsic of their underlying dynamics, a different source of non Gaussianity emerges when considering portfolios containing options or different derivative contracts. The modeling of these portfolios is mostly made by means of the so called Delta-Gamma-Normal (DGN) model. This approach is based on a second order Taylor expansion of the portfolio price variation $V$ with respect to the risk factors $X_{i=1,\dots,N}$ responsible for the portfolio fluctuations. Under this quadratic approximation, we can express $V$ as

$$V = \theta + \Delta^\top X + \frac{1}{2} X^\top \Gamma X \,, \tag{5.34}$$

where $V = W - W_0$ is the value variation over a time horizon $\Delta t$, $\theta \in \mathbb{R}$, $\Delta \in \mathbb{R}^{N\times N}$ and $\Gamma \in \mathbb{R}^{N\times N}$ contain the first and second order derivative of $V$ with respect to the risk factors. In particular $\Gamma$ accounts for possible non linearities. The DGN model assumes the risk factors to be drawn from a multivariate Normal distribution $N(0, \Sigma)$ with zero mean and covariance matrix $\Sigma$. After solving the generalized eigenvalue problem

$$CC^\top = \Sigma \,, \quad C^\top \Gamma C = \Lambda \,,$$

with $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$, the portfolio variation becomes

$$V = \theta + \sum_{i=1}^{N} \left( \delta_i Y_i + \frac{\lambda_i}{2} Y_i^2 \right) \tag{5.35}$$

where we defined $\delta \doteq C^\top \Delta$, $X \doteq CY$ and the transformed risk factors $Y_i$ are now independent Gaussian variables. The CF associated to the distribution of $V$ has been drawn in [102], reading

$$f(z) = \mathbb{E}[e^{izV}] = e^{i\theta z} \prod_{i=1}^{N} \frac{1}{\sqrt{1 - i\lambda_i z}} \exp\left\{ -\frac{1}{2} \frac{\delta_i^2 z^2}{1 - i\lambda_i z} \right\} \,; \tag{5.36}$$

the asymptotic properties of the corresponding PDF have been studied in [103] where it is shown that the behavior of the left (right) tail of the PDF is essentially dictated by the sign of the lowest (highest) eigenvalue $\lambda^*$. Focusing of the left tail, which is the one involved by risk estimation, the following three cases are possible.

1. $\lambda^* = 0$: the tail is asymptotically Gaussian;

2. $\lambda^* > 0$: the PDF is truncated on the left;

3. $\lambda^* < 0$: the left tail exhibits an exponentially damped power law decay

$$p(V) \sim |V|^{\nu} e^{\frac{V}{\lambda^*}} , \quad \text{for } V \to -\infty , \tag{5.37}$$

the index $\nu$, possibly negative, being a function of the multiplicity of $\lambda^*$.

The behavior of the right tail follows the antithetic pattern, being exponentially damped (Gaussian) for $\lambda^* > 0$ ($\lambda^* = 0$) and truncated on the right when $\lambda^* < 0$, where $\lambda^*$ is now the largest eigenvalue. So, here we see an example in which fat tails are induced by the inclusion of second order terms in the expansion leading to (5.34), even though the single risk factors $X_i$ are drawn Gaussian. The latter assumption may sound quite unrealistic in the light of the empirical facts analyzed so far, and actually it is. However, the term "risk factor" is quite generic, and the actual definition of $X_i$, a relative or logarithmic return rather than an absolute variation for instance, depends on which parametrization is chosen at first for the portfolio; its nature is generic either, possibly being related to an interest rate rather than a plain stock or a volatility itself, depending on the type of the derivative contract. So we argue that Normality is a working assumption, necessary to make the model (5.34) analytically manageable and enough general, even though sometimes it is justified by invoking the CLT (see [104, 105]).

## 5.2 Option pricing for the LinExpOU model

Now we will calibrate the LinExpOU model against real option quotes and we will compare the results with the ExpOU and the S2 model, all of which have been described in Chapter 3. With the linear choice (5.14) for the market price of volatility risk, the risk neutral dynamics for the centered log-return $X_t = \log(S_t/S_{t_0}) - r\tau$, with $\tau = t - t_0$, has the form

$$\begin{aligned} dX_t &= -\frac{1}{2}\sigma^2(Y_t)\, dt + \sigma(Y_t)\, dW^*_{1,t} \\ dY_t &= \tilde{\alpha}(\tilde{\gamma} - Y_t)\, dt + k\, dW^*_{2,t} \end{aligned} \tag{5.38}$$

with initial conditions $X(t_0) = X_0 = 0$ and $Y(t_0) = Y_0$, and $\tilde{\alpha} > 0$ that ensures the existence of a stationarity state of the $Y$ process. For $\sigma(Y_t) = me^{Y_t}$ or $\sigma(Y_t) = m\, Y_t$ we recover the ExpOU or the S2 models respectively. When the variance of $Y$, $\tilde{\beta} \doteq k^2/(2\tilde{\alpha})$ is small, a linear approximation of $\sigma$ and $\sigma^2$ around $\tilde{\gamma}$ can be performed and if we define $Z_t = Y_t + 1 - \tilde{\gamma}$, $\bar{m} = me^{\tilde{\gamma}}$, $\bar{k} = k$ for the ExpOU case, and $Z_t = Y/\tilde{\gamma}$, $\bar{m} = m\,\tilde{\gamma}$ and $\bar{k} = k/\tilde{\gamma}$ for the S2 case, the risk

neutral dynamics becomes

$$dX_t = -\frac{\bar{m}^2}{2}(2Z_t - 1)\,dt + \bar{m} Z_t\, dW_1^* \tag{5.39}$$

$$dZ_t = \tilde{\alpha}(1 - Z_t)\,dt + \bar{k}\, dW_{2,t}^* \ . \tag{5.40}$$

This dynamics, emerging here as the common limit of the ExpOU and the S2 models under the martingale measure $\mathbb{Q}$, has just the same form it had for the LinExpOU model under the objective measure (see Eq. (3.9)-(3.10)). This implies that the expression of the CF $f_{LinExpOU}(\omega; t)$, as given by Eq.s (3.12)-(3.15) is still valid, provided to replace the original parameters with their risk neutral, "tilded" versions. The linearization procedure hides a serious drawback: it breaks the martingality of the discounted price $\tilde{S}_t = S_{t_0} e^{X_t}$ which must hold for the original $\mathbb{Q}$-dynamics (5.38). That can be easily verified by observing that $\left\langle e^{X_t} | X_{t_0} \right\rangle_{LinExpOU} = f^*_{LinExpOU}(-i) \neq 1$. To account for this violation we can add by hand an extra term, a function $\mathcal{M}(\tau)$, to the drift term of $X_t$

$$dX_t = -\frac{\bar{m}^2}{2}(2Z_t - 1)\,dt + \mathcal{M}(\tau)\,dt + \bar{m} Z_t\, dW_1^* \ ; \tag{5.41}$$

reminding the exponential form (3.12) of $f_{LinExpOU}$ under the original objective measure, we guess the form of the risk neutral CF

$$f^*_{LinExpOU}(\omega; \tau) = \exp\left\{+A(\omega,\tau) + B(\omega,\tau) z_0 + C(\omega,\tau) Z_0^2 + i\omega x_0\right\} \times \exp\left\{-i\omega \frac{m^2}{2} \int_0^\tau \mathcal{M}(\tau')\, d\tau'\right\} \tag{5.42}$$

where the expressions (3.13)-(3.15) of $A, B, C$, remains unchanged. Requiring the martingale condition $\left\langle e^{X_t} \mid X_0 \right\rangle_{\mathbb{Q}} = f_{LinExpOU}(-i; \tau) = 1$ to be satisfied, we obtain the explicit expression of the extra term

$$\mathcal{M}(\tau) = \frac{2}{m^2} \frac{\mathrm{d}}{\mathrm{d}\tau} \left[A(-i,\tau) + B(-i,\tau) z_0 + C(-i,\tau) z_0^2 + x_0\right] \ ,$$

which motivates *a posteriori* the time homogeneity assumed for $\mathcal{M}$. In order to fit the model parameters, we will exploit the volatility surface reported in Table 5.1; it corresponds to the market quotes of Call options written on the asset Intesa San Paolo S.p.A. from the Milan Stock Exchange, for different time to maturities $\tau$, as of 22nd November 2007, spot price $S_0 = 5.16$ EUR. The term structure of risk free rates was retrieved from the market as well. We will also compare the results of the calibration of the LinExpOU model with those for the ExpOU and S2 models.

The fitting parameters are $k, m, \rho, \alpha, \gamma$ and $Y_0$ [4]; however, under the assumption that the the volatility has reached the stationary state at the time

[4] From now on we will drop the tilde from above the risk neutral parameters $\tilde{\alpha}$, $\tilde{\gamma}$ and the bar from above $\bar{m}$ and $\bar{k}$, unless necessary to avoid confusion.

| $\tau$ (yr) | $r_\tau$ ($\mathrm{yr}^{-1}$) | $\log(S_0/K)$ | $\sigma_{\mathrm{imp},\tau}$ ($\mathrm{yr}^{-1/2}$) |
|---|---|---|---|
| 0.0795 | 0.0425 | 0.0626 | 0.3354 |
| | | 0.0218 | 0.3089 |
| | | -0.0175 | 0.2839 |
| | | -0.0552 | 0.2599 |
| | | -0.0657 | 0.2822 |
| 0.1562 | 0.0465 | 0.1496 | 0.3427 |
| | | 0.0626 | 0.3114 |
| | | 0.0218 | 0.2823 |
| | | -0.0175 | 0.2700 |
| | | -0.0552 | 0.2566 |
| | | -0.0916 | 0.2592 |
| | | -0.1267 | 0.2630 |
| | | -0.1606 | 0.2686 |
| 0.2329 | 0.0474 | 0.0626 | 0.3347 |
| | | 0.0218 | 0.2874 |
| | | -0.0175 | 0.2704 |
| | | -0.0552 | 0.2726 |
| | | -0.0916 | 0.2681 |
| | | -0.1267 | 0.2593 |
| | | -0.1606 | 0.2643 |
| 0.0795 | 0.0425 | 0.0626 | 0.3354 |
| | | 0.0218 | 0.3089 |
| | | -0.0175 | 0.2839 |
| | | -0.0552 | 0.2599 |
| | | -0.0657 | 0.2822 |
| 0.1562 | 0.0465 | 0.1496 | 0.3427 |
| | | 0.0626 | 0.3114 |
| | | 0.0218 | 0.2823 |
| | | -0.0175 | 0.2700 |
| | | -0.0552 | 0.2566 |
| | | -0.0916 | 0.2592 |
| | | -0.1267 | 0.2630 |
| | | -0.1606 | 0.2686 |
| 0.2329 | 0.0474 | 0.0626 | 0.3347 |
| | | 0.0218 | 0.2874 |
| | | -0.0175 | 0.2704 |
| | | -0.0552 | 0.2726 |
| | | -0.0916 | 0.2681 |
| | | -0.1267 | 0.2593 |
| | | -0.1606 | 0.2643 |

Table 5.1: Implied volatilities market data.

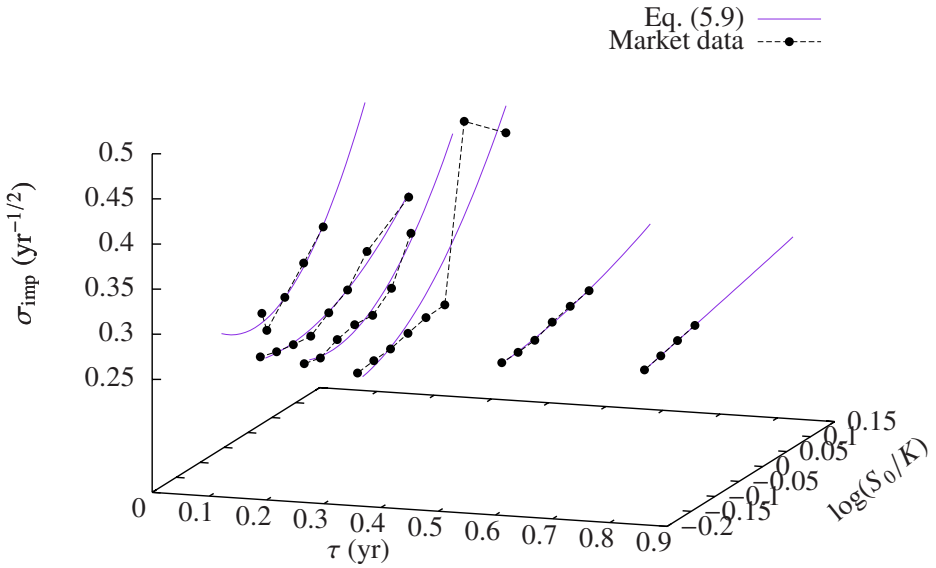



Figure 5.2: Implied volatilities smiles for Intesa San Paolo as fitted with the parabolic approximation (5.9).

| $\tau$ (yr) | $\sigma_\tau^{Mk} \pm \epsilon_{\sigma_\tau}^{Mk}$ | | $\zeta_\tau^{Mk} \pm \epsilon_{\zeta_\tau}^{Mk}$ | | $\kappa_\tau^{Mk} \pm \epsilon_{\kappa_\tau}^{Mk}$ | |
|---|---|---|---|---|---|---|
| 0.0795 | 0.0885 | 0.0063 | -0.80 | 0.20 | 2.0 | 2.1 |
| 0.1562 | 0.1145 | 0.0012 | -0.578 | 0.064 | 1.44 | 0.31 |
| 0.2329 | 0.164 | 0.013 | -1.11 | 0.16 | 4.6 | 1.8 |
| 0.3260 | 0.210 | 0.071 | -1.82 | 0.92 | 5.3 | 7.8 |
| 0.5781 | 0.235 | 0.011 | -0.587 | 0.066 | 1.7 | 1.2 |
| 0.8274 | 0.269 | 0.011 | -0.760 | 0.068 | 0.2 | 1.0 |

Table 5.2: Market calibrated normalized cumulants and their standard 68% errors.

| | $\alpha^* \pm \epsilon_\alpha$ (yr$^{-1}$) | | $k^* \pm \epsilon_k$ (yr$^{-1/2}$) | | $m^* \pm \epsilon_m$ (yr$^{-1/2}$) | | $\rho^* \pm \epsilon_\rho$ | | $\beta^* \pm \epsilon_\beta$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| ExpOU | 6.3 | 1.5 | 1.3 | 0.1 | 0.266 | 0.018 | -0.51 | 0.09 | 0.13 | 0.04 |
| S2 | 5.7 | 1.3 | 1.9 | 0.4 | 0.265 | 0.008 | -0.41 | 0.07 | 0.32 | 0.14 |
| LinExpOU | 5.6 | 1.3 | 1.9 | 0.4 | 0.264 | 0.008 | -0.41 | 0.07 | 0.34 | 0.15 |

Table 5.3: Optimal parameters for the three considered models, and standard 68% errors.

the quotes correspond to, we can fix arbitrarily $Y_0 = \gamma$. Moreover, it can be shown that the moments of the volatility driving $X_t$ always depend on the combination $\bar{m}$, not on $m$ and $\gamma$ separately (see also [67]); so we can also fix $\gamma$ and we choose $\gamma = 0$ for the ExpOU and LinExpOU cases and $\gamma = 1$ for the S2 model, in order to have the same value of the starting volatility. We are left with four three parameters and we fit them with the following simple two step procedure. First, we fit the market smiles with the parabolic approximation (5.9) (see Fig. 5.2), this way exctrating the values for the variance, the skewness and the kurtosis of the implied PDF for the available time to maturities, and associated standard errors, reported in Table 5.2. Then we fit the market cumulants, minimizing the square distance from the theoretical predictions of the models. For the LinExpOU model the cumulants time scaling is provided by the expressions in Appendix A and analogous expressions can be obtained for the S2 model [5] On the other hand, the cumulants of the ExpOU model have to be simulated via Monte Carlo. The optimal parameters $\alpha^*, k^*, m^*, \rho^*$ satisfy the equation

$$\alpha^*, k^*, m^*, \rho^* = \underset{\alpha,k,m>0,\ \rho\in[-1,1]}{\operatorname{argmin}} \sum_{\tau} \left[ \frac{(\sigma_\tau^{Mk} - \sigma_\tau)^2}{\epsilon_{\sigma_\tau}^2} + \frac{(\zeta_\tau^{Mk} - \zeta_\tau)^2}{\epsilon_{\zeta_\tau}^2} + \frac{(\kappa_\tau^{Mk} - \kappa_\tau)^2}{\epsilon_{\kappa_\tau}^2} \right]$$

where the standard error is $\epsilon^2 = \epsilon^{Mk\,2} + \epsilon^{MC\,2}$, the Monte Carlo error being zero for the LinExpOU and S2 cases.

The optimal parameters for the data set at hand, calibrated by implementing the previous procedure by means of Marquard-Levemberg and `MINUIT` minimization routines, are reported in Table 5.3. The value of $\beta$ for the ExpOU model supports its linearization and the assumption of the market being in a low volatility fluctuations regime. The correlation coefficient is always negative, in agreement with what is known for stock returns distributions. Moreover, no statistically significant differences are observed between S2 and LinExpOU, which was expected as the two models are quite similar in the structure. Once the models have been calibrated, we can reconstruct the volatility surface and check it against the original ones. For the LinExpOU and S2 models, this can be done evaluating the formula (5.19) by means of standard trapezoidal routines or Fast Fourier Transform (FFT) algorithms [6], while for ExpOU a further Monte Carlo simulation is mandatory.

In Fig. 5.3 we present the implied volatilities smiles reconstructed by the LinExpOU model, showing its ability to capture the correct shape of the volatility. The agreement slightly decreases for very large, $K \gg S_0$, or very small, $K \ll S_0$, strike values; from the third panel, corresponding to $\tau =$

[5] The expression of CF for the S2 model was already known in the literature [28] and since the choice (5.14) for the market prices guarantees that the dynamics is invariant in form, the same expression is valid under $\mathbb{Q}$.

[6] The analysis of the funcions $A, B, C$ in (3.13)-(3.15) shows that the strip $\mathcal{S}_h \cap \mathcal{S}_p^*$ is non empty for the calibrated values of the parameters, allowing to use the formula (5.19) for the LinExpOU model. An analogous result holds for the CF of the S2 model.

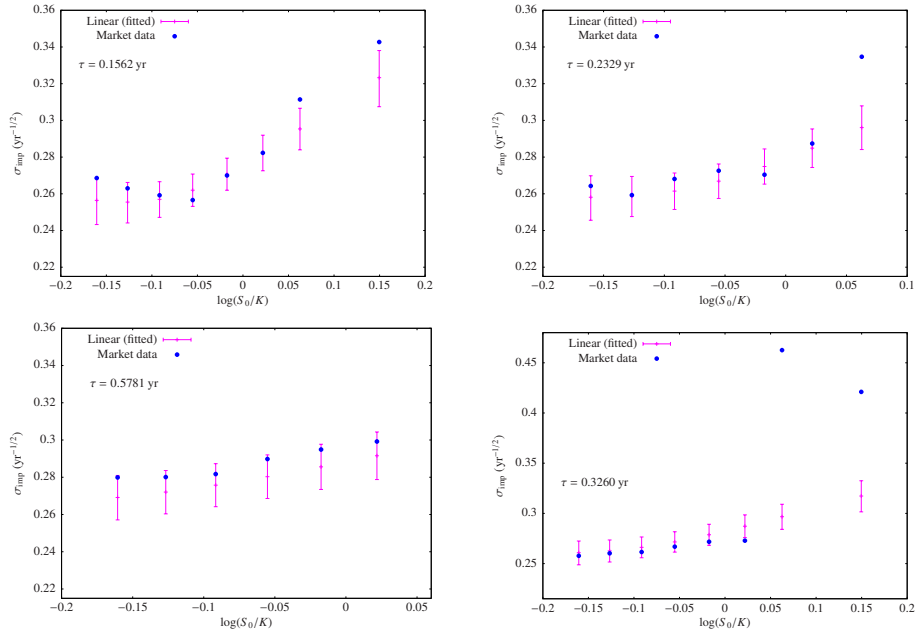



Figure 5.3: Comparison between market implied volatility smiles for Intesa San Paolo and those obtained from the LinExpOU model. Parameters values as in the third row of Table 5.3, as obtained from the calibration of the LinExpOU model.

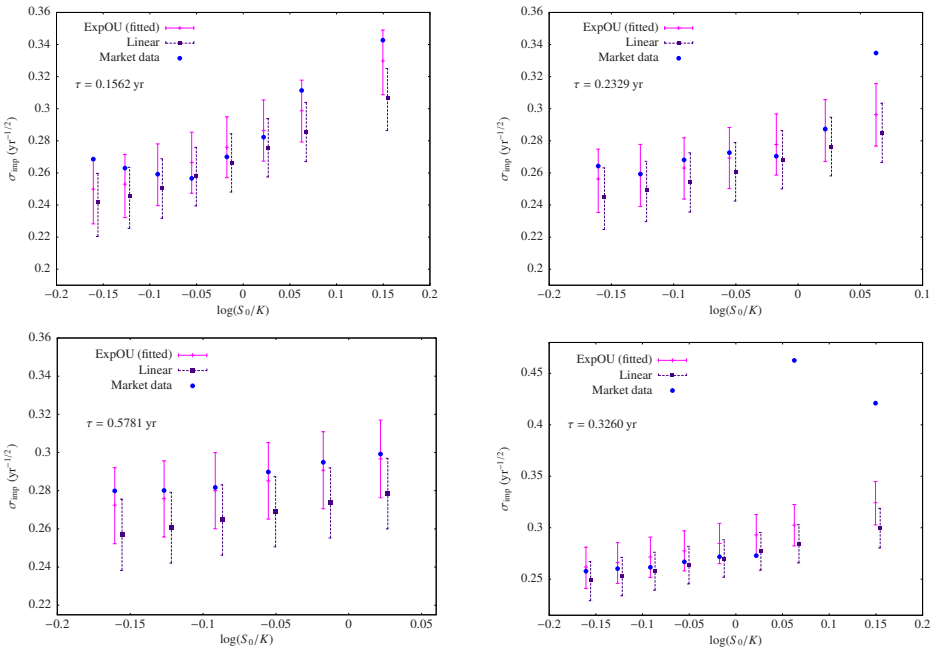



Figure 5.4: Comparison between market implied volatility smiles for Intesa San Paolo and those obtained from the ExpOU and LinExpOU models. Parameters values as in the first row of Table 5.3, as obtained from the calibration of the ExpOU model.

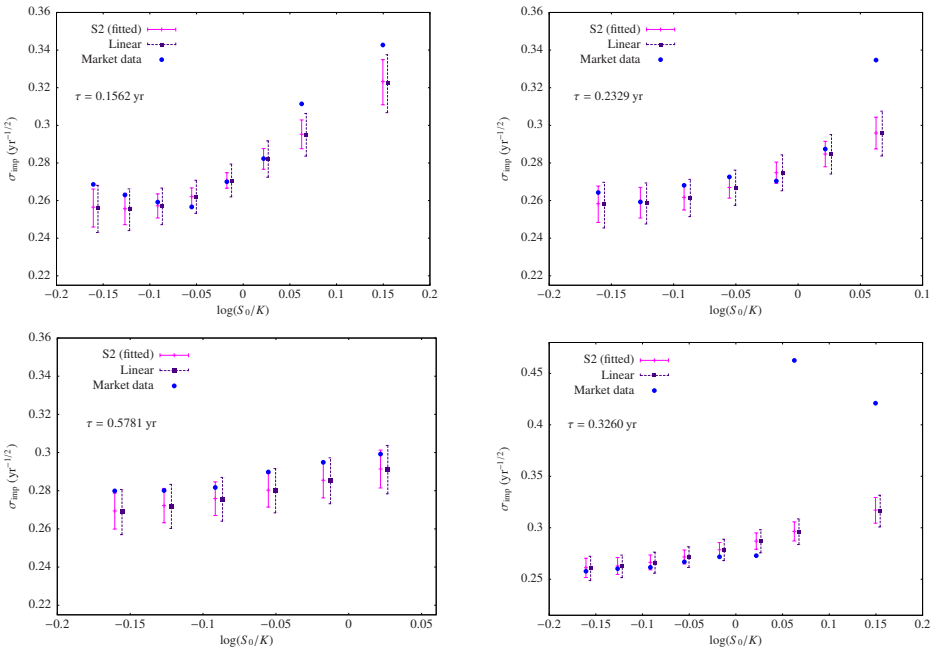



Figure 5.5: Comparison between market implied volatility smiles for Intesa San Paolo and those obtained from the S2 and LinExpOU models. Parameters values as in the second row of Table 5.3, as obtained from the calibration of the S2 model.

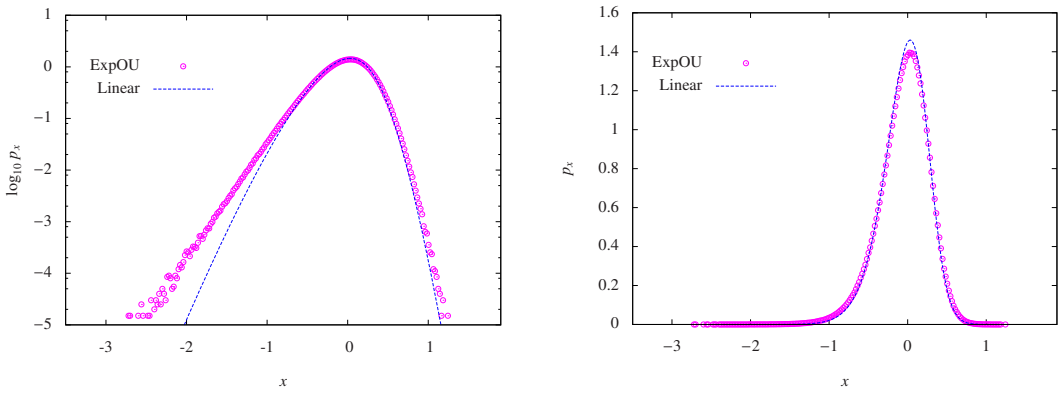



Figure 5.6: Comparison between the return PDF $p(x, \tau|X_0, Z_0)$ with $\tau = 1$ for the LinExpOU and ExpOU models in log-linear (left panel) and linear (right panel) scales. Parameters values as in the first row of Table 5.3.

0.326 yr, we notice that there are two anomalous quotes, whose values are aout of scale and can not reproduced as expected. The statistical uncertainty on the parameters values reflects in the standard errors associated to the curves. The error propagation has been performed making use of the numerical first order derivatives. It is worth pointing out that, following the guidelines depicted in the previous section, calibration and pricing are carried out in an efficient way; this is mainly due to the analytical characterization we provided for the CF and the cumulants, making the entire approach a real time procedure. In Fig. 5.4 we present the smiles for the set of parameters corresponding to the ExpOU calibration and we plot both the curves obtained from MC simulation ($T = 10^5$) of the exponential model and those computed integrating Eq. (5.19) for the LinExpOU model with the same parameters values (the curves corresponding to the LinExpOU case have been shifted rightward). In Fig. 5.6 we plot the PDF of the LinExpOU model against the one for the ExpOU model computed with trapezoidal integration and MC simulation, respectively. Both panels confirm the analysis performed in [67], showing fatter tails and a lower central peak for the histogram of the ExpOU with respect to the PDF of the LinExpOU model. Even though the value $\beta \sim 13\%$ (see Table 5.3) is at the border of the regime allowing the linearization, as far as the volatility smiles are concerned, we conclude that they are in a good statistical agreement. In Fig. 5.5 a comparison analogous to the one in Fig. 5.4 is reported for the S2 parameters , revealing again that the results for the two cases are statistically compatible. With respect to Fig. 5.4, the narrower error bars reflect the fact that parameters fitting has been performed exploiting the available analytical information. Actually, the MC simulation involved both in the calibration and the price computation for ExpOU introduces an additional statistical uncertainty.

We can conclude that the LinExpOU model, being analytically characterzied, can be fitted efficiently. Most of all, if a careful errror propagation is performed, the LinExpOU volatility smiles are statistically indiscernible from the S2 and ExpOU models.

## 5.3 Bayesian VaR

The Bayesian modeling with PPMs described in Chapter 4 finds an interesting application in evaluating the VaR (5.6) for single assets [5]. In the $\boldsymbol{\mu}$-PPM model, at the $l$-th iteration of the Gibbs sampling we obtain a peculiar partition structure in which all the returns share the same variance $\sigma^2_{(l)}$, but every cluster of observations corresponds to a different value of the mean $\mu^*_{1(l)}, \mu^*_{2(l)}, \ldots, \mu^*_{|\rho|(l)}$. In order to provide a single VaR estimates for each iteration we choose to combine the different entries of $\boldsymbol{\mu}^*_{(l)}$ by means of the following

| | $\mathcal{P}^* = 5\%$ | | | $\mathcal{P}^* = 1\%$ | | |
|---|---|---|---|---|---|---|
| | $\boldsymbol{\mu}$-PPM | $\boldsymbol{\sigma^2}$-PPM | $\boldsymbol{\sigma^2}$-CP | $\boldsymbol{\mu}$-PPM | $\boldsymbol{\sigma^2}$-PPM | $\boldsymbol{\sigma^2}$-CP |
| MIB30.MI | $1.45^{+0.05}_{-0.05}$ | $1.74^{+0.11}_{-0.12}$ | $1.76^{+0.01}_{-0.01}$ | $2.07^{+0.06}_{-0.06}$ | $2.48^{+0.16}_{-0.17}$ | $2.49^{+0.01}_{-0.01}$ |
| LTO.MI | $2.08^{+0.07}_{-0.07}$ | $2.78^{+0.15}_{-0.16}$ | $2.66^{+0.02}_{-0.02}$ | $2.95^{+0.09}_{-0.09}$ | $3.94^{+0.21}_{-0.21}$ | $3.78^{+0.03}_{-0.03}$ |
| MB.MI | $1.91^{+0.07}_{-0.08}$ | $2.40^{+0.12}_{-0.12}$ | $2.36^{+0.01}_{-0.01}$ | $2.72^{+0.11}_{-0.11}$ | $3.40^{+0.17}_{-0.17}$ | $3.35^{+0.02}_{-0.02}$ |
| SRG.MI | $1.58^{+0.05}_{-0.05}$ | $1.97^{+0.12}_{-0.13}$ | $2.01^{+0.01}_{-0.01}$ | $2.26^{+0.06}_{-0.06}$ | $2.81^{+0.17}_{-0.17}$ | $2.87^{+0.02}_{-0.02}$ |

Table 5.4: Estimated daily VaR (%) with 68% credibility intervals.

arithmetic average

$$\frac{\Delta^*_{(l)}}{S_0} = -\sum_{d=1}^{|\rho|} \frac{\left|S_{d(l)}\right|}{T} \mu^*_{d(l)} + \sigma_{(l)} \sqrt{2}\, \mathrm{erfc}^{-1}(2\mathcal{P}^*)$$

where we normalized the VaR to the spot value $S_0$ of the asset and $\mathrm{erfc}^{-1}$ is the inverse of the complementary error function. For the $\boldsymbol{\sigma}^2$-PPM model, the clustering structure emerges upon the vector of the variances $\boldsymbol{\sigma}^{2*}_{(l)} = (\sigma^{2*}_{1,(l)}, \ldots, \sigma^{2*}_{|\rho|(l)})$ while the observations share the same value of the mean $\mu_{(l)}$. In this case, the previous definition becomes

$$\frac{\Delta^*_{(l)}}{S_0} = -\mu_{(l)} + \sum_{d=1}^{|\rho|} \frac{\left|S_{d(l)}\right|}{T} \sigma^*_{d(l)} \sqrt{2}\, \mathrm{erfc}^{-1}(2\mathcal{P}^*) \ .$$

For the trivial partition, $|\rho| = 1$, previous formulas reduce to the well know expression (5.7) of the VaR for Gaussian *i.i.d.* returns; otherwise, they introduce corrections due to the contribution of the outlying clusters. Finally we combine all the previous values through an ergodic mean to obtain the final posterior VaR estimate

$$\frac{\Delta^*}{S_0} \doteq \frac{1}{L} \sum_{l=1}^{L} \frac{\Delta^*_{(l)}}{S_0} \ .$$

The VaR estimates (normalized to $S_0$) for the data set described at the end of Chapter 4, the MiB30 index, Lottomatica, Mediobanca and Snam Rete Gas, and for the two values of interest $\mathcal{P}^* = 5\%$ and $\mathcal{P}^* = 1\%$, are reported in Table 5.4. In the same table we also report the estimates obtained with a different PPM approach, developed by Loschi et al. [106] and referred to a $\boldsymbol{\sigma}^2$-CP model in the following, which still considers a partition structure over the variances, but constraining the clusters to be contiguous. This more sophisticated approach, besides requiring different sampling algorithms, better applies to the problems of detecting change points and modeling of the volatility clustering but here it is only considered as a benchmark for our VaR estimates. Nevertheless, we see that the VaR results for the $\boldsymbol{\sigma}^2$-PPM and $\boldsymbol{\sigma}^2$-CP models are in good agreement, allowing to conclude that, as far as the risk measure

| VaR(%) | $\mathcal{P}^*$=5% | | $\mathcal{P}^*$=1% | |
|---|---|---|---|---|
| | Normal | Student-$t$ | Normal | Student-$t$ |
| MIB30.MI | $1.38^{+0.05}_{-0.07}$ | $1.27^{+0.04}_{-0.06}$ | $1.95^{+0.07}_{-0.09}$ | $2.22^{+0.09}_{-0.10}$ |
| LTO.MI | $2.50^{+0.14}_{-0.15}$ | $2.15^{+0.08}_{-0.09}$ | $3.55^{+0.19}_{-0.20}$ | $4.05^{+0.19}_{-0.22}$ |
| MB.MI | $2.07^{+0.06}_{-0.09}$ | $1.89^{+0.05}_{-0.07}$ | $2.95^{+0.09}_{-0.11}$ | $3.37^{+0.12}_{-0.14}$ |
| SRG.MI | $1.62^{+0.05}_{-0.08}$ | $1.48^{+0.04}_{-0.07}$ | $2.32^{+0.08}_{-0.10}$ | $2.65^{+0.11}_{-0.12}$ |

Table 5.5: Daily ML estimated VaR(%) values at 5% and 1% significance level with 68% bootstrap intervals.

evaluation is concerned, even a simpler model discarding the contiguous character of clusters in the volatility series is also effective. On the other hand, the estimates obtained with the $\boldsymbol{\mu}$-PPM approach tend to underestimate VaR, which can be justified by noticing that the contribution of the variance to daily VaR is about ten times that of the mean, due to the typical values of the mean rate of return for daily horizons.

As said before, PPMs in our framework allow to stay in a Normal settings while renouncing to the hypothesis of identical distribution. From this point of view, it is of major interest to compare this approach with the one presented in [25]. There, a maximum likelihood (ML) approach is used to fit the empirical return distribution with a Generalized Student-t (GST) distribution with tail index $\nu > 2$, and an exact expression for the corresponding VaR is obtained. This GST approach assumes i.i.d. returns and explicitly takes into account the fat tails of daily returns by means of the GST assumption for their distribution. We assume it as a benchmark since it presents a good agreement with historical simulations. The corresponding estimates for the dataset at hand, and 68% bootstrap errors, are shown in Table 5.5 along with the Gaussian ML ones. The comparison of Bayesian and ML estimates reveals that for $\mathcal{P}^* = 1\%$ the results for $\boldsymbol{\sigma}^2$-PPM and $\boldsymbol{\sigma}^2$-CP are those in best agreement with the GST values while for for $\mathcal{P}^* = 5\%$ the $\boldsymbol{\mu}$-PPM approach seems to perform better. This may indicate that a PPM approach for the variances tends to overestimate the volatility of the data with respect to a ML static fit, which effect is less evident for lower $\mathcal{P}^*$ where the VaR estimate is mostly determined by far left tail of the distribution, conversely being less sensitive to its bulk. On the other hand, the Gaussian ML values always underestimates the Student-$t$ values for $\mathcal{P}^* = 1\%$; from this point of view, we may say that the PPM approach, renouncing to model identically distributed variables, is able to capture indirectly the effect of the power law tails of a GST distribution, even keeping to model returns as Gaussian variables.

The Basel II regulations [89] require the computation of the 10 day ahead VaR for fixing the regulatory capital for market risk, and suggest propagating the daily estimates to the 10 days horizon with the square root rule

$$\Delta^*(1 \text{ day}) \Rightarrow \Delta^*(10 \text{ days}) = \sqrt{10}\,\Delta^*(1 \text{ day}) \ ;$$

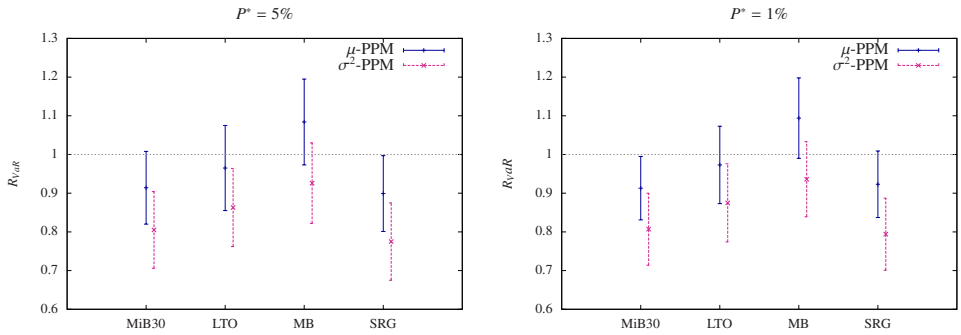



Figure 5.7: Ratio of the 10 days ahead over the daily VaR, and 68% error bars, computed for the $\boldsymbol{\mu}$-PPM and $\boldsymbol{\sigma}^2$-PPM models.

however, this scaling rule is valid only for standard VaR formula assuming Gaussian *i.i.d.* returns, otherwise it should be corrected. To test this scaling, we computed the 10 days ahead VaR forecast for the $\boldsymbol{\mu}$-PPM and $\boldsymbol{\sigma}^2$-PPM models, running the Gibbs schemes over the series of non overlapping 10 days returns obtained from the original daily sets. In Fig. 5.7 we plot the ratio

$$R_{VaR} = \frac{\Delta^*(10 \text{ days})}{\sqrt{10}\,\Delta^*(1 \text{ day})}\,,$$

and it shows statistically significant deviations from the square root law, confirming the ability of the PPM approach in capturing the non trivial time scaling of the risk measure over different time horizons.

To asses the robustness of the PPM approaches, we performed the UC and CC coverage tests for both the models. Using a rolling window of returns from the original daily series, we ran $N$ instances of the Gibbs sampler; for $J = 1, \ldots, N$ the algorithms compute the ex ante VaR estimate $\Delta^*_{MAX_J}$ using the returns $y_i$ with $i = J, \ldots, MAX_J$, where we fixed $MAX_J = J+744$ in order to use all the information from our original series of 1000 returns and to obtain $N = 255$ VaR estimates, roughly corresponding to one trading year as imposed by the Basel regulators. Then we check $\Delta^*_{MAX_J}$ against the ex post realized return $y_{MAX_J+1}$. An exception occurs when $y_{MAX_J+1} < -\Delta^*_{MAX_J}$. We set a state indicator $I_J$ set equal to 1 if we register an exception, 0 otherwise. This way we obtain the numbers $n$ and $N^{ij}$ needed to compute the $LR_{UC}$ and $LR_{CC}$ statistics defined in Eq.s (5.31)-(5.33). In Table 5.6 we report the results for the LTO.MI series, the ones for the other data sets being similar. Gloabally, the models pass both Kupiec's and Christoffersen's tests, with the only exception of the $\boldsymbol{\sigma}^2$-PPM model for $\mathcal{P}^* = 5\%$, which produced a fairly low number of exceptions with respect to the expected number $\langle n \rangle = 255 \times 0.05 \approx 13$, indicating a too conservative risk estimate. This exception can be justified by the trend of the considered time series, whose historical volatility exhibits the steep decreasing trend plotted in Fig. 5.8, meaning that, systematically, ex-ante VaR estimates computed by "training" the Gibbs sampler with observations

| | $\mathcal{P}^* = 1\%$ | | | $\mathcal{P}^* = 5\%$ | | |
|---|---|---|---|---|---|---|
| LTO.MI | $n$ | $LR_{UC}$ | $LR_{CC}$ | $n$ | $LR_{UC}$ | $LR_{CC}$ |
| $\boldsymbol{\mu}$-PPM | 5 | 1.857 | 2.057 | 9 | 1.288 | 1.947 |
| $\boldsymbol{\sigma^2}$-PPM | 1 | 1.237 | 1.245 | 5 | 13.873 | 14.073 |

Table 5.6: Backtesting results: the model is rejected at 5% significance level if $LR_{UC} > 3.84$ (unconditional coverage test), or $LR_{CC} > 5.99$ (conditional coverage test).

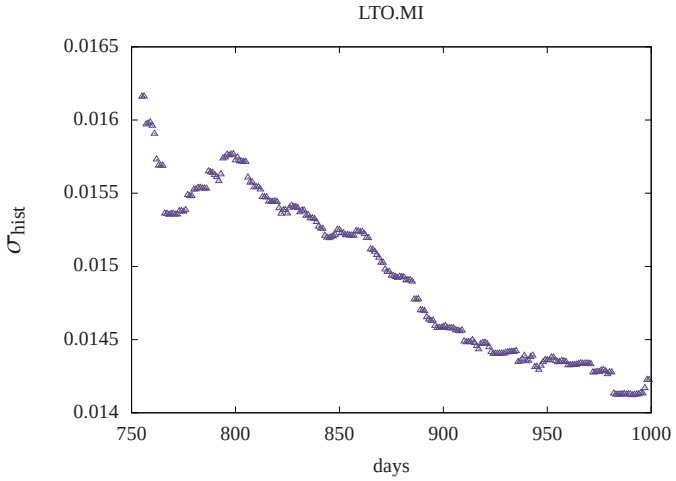



Figure 5.8: Historical volatility of the LTO.MI daily return series computed over a rolling window of 744 returns from the original 1000 data.

corresponding to higher volatility regimes were then compared with ex-post realized returns drawn from lower volatility periods.

## 5.4 Portfolio VaR

As the last application, we will briefly describe some results about portfolio risk estimation, which problem we addressed in [10] in the context of the DGN approximation. Indeed, since the CF for the benchmark DGN model is known, Eq. (5.36), the problem of estimating VaR and ES can be elegantly solved applying our formulae (5.20)-(5.22). The model parameters with respect to which the sensitivities are defined are the one appearing in (5.35), namely $\beta = \{\theta, \delta_{i=1,\dots,N}, \lambda_{i=1,\dots,N}\}$, $N$ being the number of risk factors. The derivatives

of the CF are readily computed

$$\begin{aligned}
\frac{\partial f(\omega)}{\partial \theta} &= i\omega f(\omega) \\
\frac{\partial f(\omega)}{\partial \delta_i} &= -\frac{\delta_i \omega^2}{1 - i\lambda_i \omega}\, f(\omega) \\
\frac{\partial f(\omega)}{\partial \lambda_i} &= \frac{i\omega}{2(1 - i\lambda_i z)} \left[1 - \frac{\delta_i^2 \omega^2}{1 - i\lambda_i z}\right] f(\omega)\,,
\end{aligned}$$

and the VaR and ES sensitivities can be obtained through formulas (5.23) and (5.24). The involved Fourier integrals can be computed both by means of standard trapezoidal integration routines or FFT algorithms. However, in the first case, the integration procedure has to be nested in a root finding one to solve (5.21) for $\Delta^*$ which can be computationally demanding. On the other hand, the advantage of FFT teqnique is that, given an input grid of $\omega$ values, the output of the algorithm is a grid of $\Delta^*$ values which are in biunivocal correspondence with as many values of $\mathcal{P}^*$. This way we obtain an entire strip of VaR and ES, covering any $\mathcal{P}^*$ range of interest, depending on the tuning of the input grid.

The effectiveness of this approach has to be tested with VaR and ES measures computed by the historical approach described in Section 5.1.6 for a synthetic sample of portfolio variations generated from the model (5.34). To this end, we simulated $T = 10^7$ portfolio scenarios for $N = 15$ risk factors, obtained by as many draws of the Gaussian variables $Y \sim N(0, \mathbf{1})$, and the values $V_t$ were constructed from Eq. (5.35) after arbitrarily fixing $\theta$, $\delta$ and $\lambda$. Attention has to be paid here to freeze the values of $Y_t$, so that re-evaluation of the portfolio with shocked parameters is always carried out based on the same sample. More details about efficient algorithms to generate the scenarios under a quadratic approximation of the portfolio losses can be found in [107]. We tested the formulas of Section 5.1.3 for choices of the parameters $\beta$ corresponding to all the three possible cases depending on the sign of the smallest eigenvalue $\lambda^*$. Here we present only the plots corresponding to the case $\lambda^* < 0$, fixing $\theta = 0$, $\delta_{i=1,\ldots,15} = 1$, $\lambda_{i=1,\ldots,5} = \lambda^* = -2$, $\lambda_{i=6,\ldots,9} = 1$ and $\lambda_{i=10,\ldots,15} = 2$. The correspondences $\mathcal{P}^*$-VaR and $\mathcal{P}^* - ES$ are reported in the left panel of Fig. 5.9. The semi-analytical curves are superimposed to the estimates from historical simulation for the relevant values $\mathcal{P}^* = 0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05$ with corresponding 98% CL errors computed as in Section 5.1.6. In Fig. (5.10) the VaR and ES sensitivities are reported. They exhibit larger error bars, due to the propagation of statistical uncertainty of the risk estimates which is magnified by a factor of order $O(1/\Delta\beta)$ through the finite difference formula (5.30). Moreover, it is clear that these errors increase for lower $\mathcal{P}^*$ due to the decreasing of the sample size on the left tail of the Monte Carlo generated values $V_t$. The complete agreement of the Fourier semi-analytical estimates with the MC ones confirm the soundness of our approach. As stated earlier, this set of parameters corresponds to an exponentially damped power law scaling, which

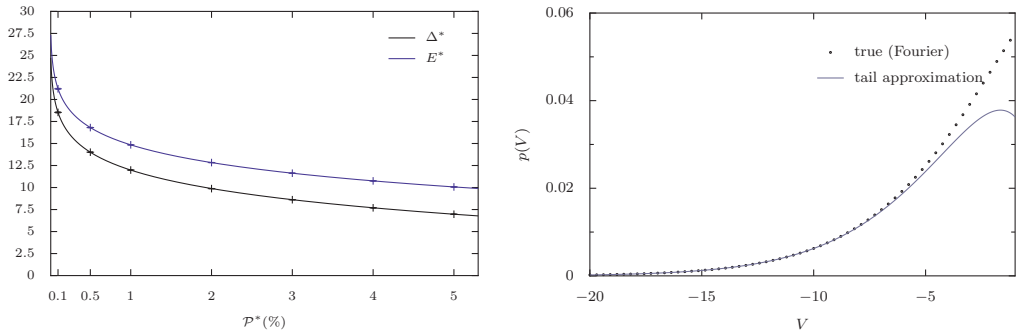



Figure 5.9: Case $\lambda^* < 0$. Left panel: VaR and ES. The semi analytical curves are compared to the points obtained via historical simulation. Right panel: exponentially damped decay of the left tail, corresponding to Eq. (5.37).

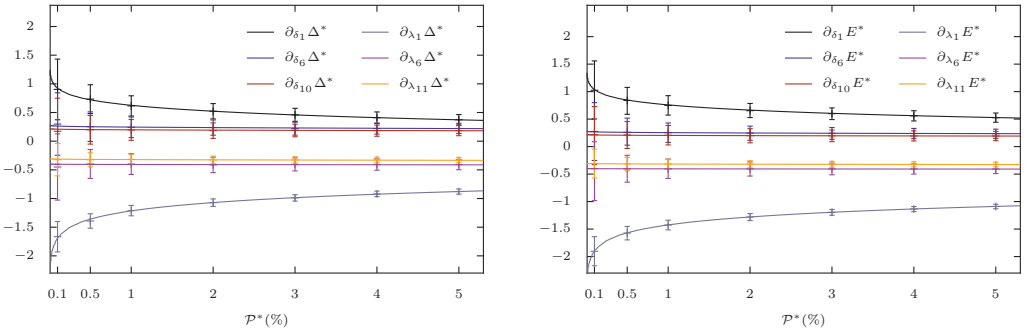



Figure 5.10: Case $\lambda^* < 0$. VaR (left) and ES (right) sensitivities.

is confirmed by the right panel of Fig. 5.9 where the PDF $p(V)$ obtained antitranforming the CF (5.36) is fitted with its analytical approximation (5.37).

# Conclusions and future perspectives

Starting from the work of Bachelier, the Gaussian paradigm has spread widely in the financial world without quantitatively sound motivations. While the Bachelier assumption of a Normal distribution of prices was clearly unrealistic given the non negative nature of prices, the Geometric Brownian Motion (GBM) introduced by Osborne, Samuelson and Merton for price returns is also unrealistic, especially for high frequency data. Observed distributions exhibit a much larger probability of extreme events, especially losses, than expected from the Gaussian probability density function (PDF) fitting the data. That possibly results in a rough underestimation of risk, either when assessing exposures associated to the owning of an asset or portfolio, and when computing the prices of derivative contracts where these estimation errors are prone to be amplified by the nonlinearity of the instrument itself. It is now well understood that high frequency returns are compatible with a Tsallis or Student-$t$ distribution for instance, namely a functional form featuring a much slower, power law decay on the tails. This is not the end of the story: it does not suffices to single out the optimal function fitting the data if it remains a static model. Financial processes exhibit correlations between meaningful quantities at different times, all of which witness a non trivial underlying dynamics. So, the quest is for dynamic models able to reproduce those correlations as well as the evolution over different time scales of the return distributions. The GBM is a dynamic model in which returns are Gaussian at all times and the volatility is constant. So it contradicts also the evidence that the volatility is a time varying, apparently random, process. Stochastic Volatility Models (SVMs) sign a clear step beyond in the modeling, since the coupled dynamics of the returns and the volatility introduces correlations and, usually, an exponential rather than Gaussian decay of the tails.

In this thesis, some interesting cases of SVMs have been studied and developed. The linearized version of the exponential Ornstein-Uhlenbeck model (LinExpOU), discussed in [67, 7, 8] and presented in Chapter 3, looses some interesting properties of the parent model, such as the Log-Normal distribu-

tion of the volatility, but it does not fail to be non Gaussian, as it can be proved by the analysis of the moments, exactly computable starting from the expression of its characteristic function (CF). We derived rigorously the risk neutral dynamics of this model and we found that the expression of the CF is still valid with minimal modifications, thus allowing to implement the efficient calibration strategy depicted in Chapter 5. Our estimates of the risk neutral parameters are dressed with robust errors, resulting in a band for the reconstructed volatility surfaces. A quantitative comparison of the LinExpOU model with the exponential Ornstein-Uhlenbeck and the Stein-Stein models in the light of these uncertainties, highlighted that no statistically significant difference can be traced as far as option prices are concerned. As a future research topic, it would be interesting to study the scaling properties of the risk neutral dynamics, testing to what extent the model calibrated against the present volatility surface is able to "follow" future movements of the surface itself. A related problem would be to asses the sensitivities of the parameters to the movements of the market volatility curves, which is important for traders using these sensitivities to hedge option positions.

The minimal linear model, also developed in Chapter 3 and still under investigation, is a different yet promising and realistic model. The main correlations were characterized analytically and so we did for the moments of the return distribution. These ones may diverge at finite times as a consequence of the power law tails of the stationary distribution of the instantaneous volatility, the latter being an Inverse Gamma in agreement with the empirical analysis of the volatility proxies. This last feature, even though desirable, imposes rigid constraints on the process time scales which prevent the model to capture the long range memory of the volatility autocorrelation. Despite of that, the empirical analysis of this model, carried out for the daily return series of the Standard & Poor 500 index, revealed a high degree of realism despite its linearity. We proposed a systematic methodology for estimating its parameters, and we found values which support the model internal consistency requiring the convergence of the volatility autocorrelation function. Most of all, the analysis of the simulated return PDFs testified the ability of the model to capture the leptokurtic, power law distribution of the data at high frequency and their scaling properties over longer time horizons. Future research activity will be devoted to test the model against ultra high frequency, intraday returns, to asses the parameters sensitivity to the time resolution of the data used for their estimation. Such an analysis will provide the starting point for testing whether the proposed dynamics can be considered realistic at time scales shorter than a trading day, and for evaluating the importance of this intraday dynamics in determining the relevant features of the daily return distribution. A different goal would be exploring possible techniques to solve the model Fokker-Planck equation for the PDF or its CF at least.

In all the SVMs considered here, the process driving the volatility belongs to the more general class of Multiplicative Noise Diffusion Processes analyzed in Chapter 2 extending the characterization made in [9]. There, the proposed algorithmic strategy established a simple connection between the time dependent function rescaling the drift and diffusion coefficients of the Îto equation and the relaxation modes of the moments of the PDF. In particular, we were able to modulate this time dependence to reproduce power law scalings or more general stretched exponential ones. The analytical expressions of the moments may help the analysis of time series such as those considered in [44, 45, 49, 54] providing a simple way to fix model parameters out of sample averages and it may also be exploited to study the behavior of Edgewoth-like cumulant expansions of the PDF.

In Chapter 4 we presented two alternative models based on the Bayesian framework of the Product Partition Models (PPMs), which were introduced in [5, 6]. They are not dynamic in the sense the SVMs are, but they take into account that the distribution of returns may change with time. We identified clusters of observations based on their distribution parameters and we defined and singled out outlying points in the series. In this respect, a possible development would be supporting these models with a macroeconomic analysis to check out whether the atypical returns can be associated with specific economic events triggering a regime switching. In Chapter 5 the same models were successfully applied to VaR estimation for single assets. Minor clusters introduce significant deviations from the crude Gaussian Maximum Likelihood (ML) estimates assuming *i.i.d.* returns; moreover the comparison with a ML approach fitting the empirical returns with a Student-$t$ distribution highlighted that the PPM approach leads to compatible risk estimates, even though it starts from a Gaussian prior assumption. This was an indirect evidence that the relaxation of the identical distribution hypothesis allows to capture the effect of fat tails in the VaR estimates. In principle, these models could be extended to the multivariate case of portfolios as well, considering a hierarchical approach to model the variance-covariance matrix of the problem. From a computational point of view, this extension could be highly demanding even for small portfolios; however the dimensionality of the problem could be reduced applying filtering techniques, such as those developed in [108, 109, 110, 111], which allow to retain only the statistically significant information present in the correlation matrix.

Finally, at the end of Chapter 5, we presented a strictly financial application to non linear portfolios, discussing some of the results published in [10]. At this aim we developed an efficient computational methodology for risk assessment based on the generalized Fourier transform formalism. The extreme generality of this method, requiring the knowledge of the CF only, placed the computation of risk measures and their sensitivities with respect to the portfolio composition into a unified perspective, the latter aspect being of major importance for portfolio managers. We considered the specific case of the Delta-Gamma-

Normal model, widely exploited in practice, testing our results with the aid of MC simulation. As a future research, it would be worth to explore possibilities to build a more realistic portfolio model in which the risk factors feature a stochastic volatility dynamics, and for which the portfolio CF can be drawn thus making our methodology still feasible.

As a last remark, the author hopes that, even when addressing applications which are of interest in financial practice, the "physical" focus on the actual statistical features of the underlying fundamental quantities and on their dynamics was manifest. In the author's opinion, this focus is what makes a substantial difference between Econophysics and financial engineering or industry.

*" . . . it is still common practice in the world of quantitative finance and in the derivative industry to use blatantly irrelevant models that can always be brute force calibrated on market data to spit out meaningless numbers. In my view, financial engineering is at the stage of Ptolemea's epicycles before Kepler's ellipses. After so much twisting and tweaking (calibration is the politically correct word for it), epicycles were more precise that ellipses . . . but of course, this was no theory. "*
J. P. Bouchaud

Appendix A

# LinExpOU model: cumulants

In the following we report the analytical expressions of the cumulants of the Linear model, obtained by differentiation of the CF (3.12).

$$k_{1,\tau} = \frac{\bar{m}^2}{\alpha}(Z_0 - 1)(e^{-\alpha\tau} - 1) - \frac{\bar{m}^2}{2\alpha}\alpha\tau + X_0,$$

$$\begin{aligned} k_{2,\tau} =& \frac{1}{4}\frac{\bar{m}^2}{\alpha}\Bigg\{ -2\left(\frac{k\bar{m}}{\alpha}\right)^2 \left[e^{-2\alpha\tau} - 4e^{-\alpha\tau} - 2\alpha\tau + 3\right] + \frac{k^2}{\alpha}\left[e^{-2\alpha\tau} + 2\alpha\tau - 1\right] \\ & - 2(Z_0 - 1)^2\left[e^{-2\alpha\tau} - 1\right] - 8(Z_0 - 1)\left[e^{-\alpha\tau} - 1\right] + 4\alpha\tau\Bigg\} \\ & + 2\,\frac{k\bar{m}^3}{\alpha^2}\rho\left\{(Z_0 - 1)\left[e^{-\alpha\tau} + \alpha\tau e^{-\alpha\tau} - 1\right] - \left[e^{-\alpha\tau} + \alpha\tau - 1\right]\right\}, \end{aligned}$$

$$\begin{aligned} k_{3,\tau} =& \frac{3}{2}\frac{k^2\bar{m}^3}{\alpha^3}\Bigg\{(Z_0 - 1)\left[e^{-3\alpha\tau} - 2e^{-2\alpha\tau} + e^{-\alpha\tau}(3 + 2\alpha\tau) - 2\right] \\ & \qquad + 2\left[e^{-2\alpha\tau} - 4e^{-\alpha\tau} - 2\alpha\tau + 3\right]\Bigg\} \\ & + \frac{3}{2}\frac{k\bar{m}^3}{\alpha^2}\rho\Bigg\{\left(\frac{k\bar{m}}{\alpha}\right)^2\left[-e^{-2\alpha\tau}(3 + 2\alpha\tau) + 4e^{-\alpha\tau}(3 + \alpha\tau) + 4\alpha\tau - 9\right] \\ & \qquad + \frac{k^2}{\alpha}\left[e^{-2\alpha\tau}(1 + \alpha\tau) + \alpha\tau - 1\right] \\ & \qquad - (Z_0 - 1)^2\left[e^{-2\alpha\tau}(1 + 2\alpha\tau) - 1\right] \\ & \qquad + 2(Z_0 - 1)\left[e^{-2\alpha\tau} - 2e^{-\alpha\tau}(2 + \alpha\tau) + 3\right] + 4\left[e^{-\alpha\tau} + \alpha\tau - 1\right]\Bigg\} \\ & + 3\,\frac{k^2\bar{m}^4}{\alpha^3}\rho^2\Bigg\{(Z_0 - 1)\left[e^{-\alpha\tau}(2 + 2\alpha\tau + \alpha^2\tau^2) - 2\right] \\ & \qquad - 2\left[e^{-\alpha\tau}(2 + \alpha\tau) + \alpha\tau - 2\right]\Bigg\}, \end{aligned}$$

$$
\begin{aligned}
k_{4,\tau} = 3\,\frac{k^2\bar{m}^4}{\alpha^3}\Bigg\{ & \frac{1}{2}\left(\frac{k\bar{m}}{\alpha}\right)^2 \Big[-e^{-4\alpha\tau} + 4e^{-3\alpha\tau} - 4e^{-2\alpha\tau}(3+\alpha\tau) \\
& \qquad +4e^{-\alpha\tau}(7+2\alpha\tau) + 8\alpha\tau - 19\Big] \\
& + \frac{k^2}{8\alpha}\left[e^{-4\alpha\tau} + 4e^{-2\alpha\tau}(1+2\alpha\tau) + 4\alpha\tau - 5\right] \\
& - \frac{1}{2}(Z_0-1)^2\left[e^{-4\alpha\tau} + 4\alpha\tau e^{-2\alpha\tau} - 1\right] \\
& + 2(Z_0-1)\left[-e^{-3\alpha\tau} + 2e^{-2\alpha\tau} - e^{-\alpha\tau}(3+2\alpha\tau) + 2\right] \\
& \qquad +2\left[-e^{-2\alpha\tau} + 4e^{-\alpha\tau} + 2\alpha\tau - 3\right]\Bigg\} \\
+6\,\frac{k^3\bar{m}^5}{\alpha^4}\rho\Bigg\{ & (Z_0-1)\left[3e^{-3\alpha\tau}(1+\alpha\tau) - 2e^{-2\alpha\tau}(3+2\alpha\tau)\right. \\
& \qquad \left. +e^{-\alpha\tau}(9+7\alpha\tau+2\alpha^2\tau^2) - 6\right] \\
& + \left[-e^{-3\alpha\tau} + 2e^{-2\alpha\tau}(5+2\alpha\tau) - e^{-\alpha\tau}(35+10\alpha\tau)\right. \\
& \qquad \left. -12\alpha\tau + 26\right]\Bigg\} \\
+3\,\frac{k^2\bar{m}^4}{\alpha^3}\rho^2\Bigg\{ & 4\left(\frac{k\bar{m}}{\alpha}\right)^2 \Big[-e^{-2\alpha\tau}(3+3\alpha\tau+\alpha^2\tau^2) \\
& \qquad +e^{-\alpha\tau}(12+6\alpha\tau+\alpha^2\tau^2) + 3\alpha\tau - 9\Big] \\
& + \frac{k^2}{\alpha}\left[e^{-2\alpha\tau}(3+4\alpha\tau+2\alpha^2\tau^2) + 2\alpha\tau - 3\right] \\
& - 2(Z_0-1)^2\left[e^{-2\alpha\tau}(1+2\alpha\tau+2\alpha^2\tau^2) - 1\right] \\
& + 4(Z_0-1)\left[2e^{-2\alpha\tau}(1+\alpha\tau) - e^{-\alpha\tau}(6+4\alpha\tau+\alpha^2\tau^2) + 4\right] \\
& \qquad -2\left[e^{-2\alpha\tau} - 4e^{-\alpha\tau}(3+\alpha\tau) - 6\alpha\tau + 11\right]\Bigg\} \\
4\,\frac{k^3\bar{m}^5}{\alpha^4}\rho^3\Bigg\{ & (Z_0-1)\left[e^{-\alpha\tau}(6+6\alpha\tau+3\alpha^2\tau^2+\alpha^3\tau^3) - 6\right] \\
& \qquad -3\left[e^{-\alpha\tau}(6+4\alpha\tau+\alpha^2\tau^2) + 2\alpha\tau - 6\right]\Bigg\}\,.
\end{aligned}
$$

Appendix B

# SVMs: computational details

**Derivation of Eq.** (3.28)**.** After expressing $Y_{t+\tau}$ in terms of its integral solution form $t$ to $t+\tau$, the function $f(Y;t,\tau) \doteq \left\langle \left[ Y_t^2 Y_{t+\tau} e^{\sqrt{c}\Delta_t W_2(\tau)} \right] \right\rangle$ can be rewritten in the form

$$\begin{aligned} f(Y;t,\tau) &= \left\langle \left[ Y_t^2 \left( Y_t + \int_t^{t+\tau} (aY_s + b)ds + \sqrt{c} \int_t^{t+\tau} Y_s dW_{2,s} \right) e^{\sqrt{c}\Delta_t W_2(\tau)} \right] \right\rangle \\ &= \left\langle \left[ e^{\sqrt{c}\Delta_t W_2(\tau)} \right] \right\rangle (\mu_3(t) + b\tau\mu_2(t)) + a \int_t^{t+\tau} \left\langle \left[ Y_t^2 Y_s e^{\sqrt{c}\Delta_t W_2(\tau)} ds \right] \right\rangle \\ &+ \sqrt{c} \int_t^{t+\tau} \left\langle \left[ Y_t^2 Y_s e^{\sqrt{c}\Delta_t W_2(\tau)} dW_{2,s} \right] \right\rangle \ . \end{aligned}$$

Taking into account that for $t \le s \le t+\tau$ we can always split $\Delta_t W_2(\tau)$ as

$$\begin{aligned} \Delta_t W_2(\tau) &= W_{2,t+\tau} - W_{2,t} = W_{2,t+\tau} - W_{2,s} + W_{2,s} - W_{2,t} \\ &= \Delta_s W_2(t+\tau-s) + \Delta_t W_2(s-t)\,, \end{aligned}$$

and the function $f(Y;t,\tau)$ becomes

$$\begin{aligned} f(Y;t,\tau) &= \left\langle \left[ e^{\sqrt{c}\Delta_t W_2(\tau)} \right] \right\rangle (\mu_3(t) + b\tau\mu_2(t)) \\ &+ a \int_0^{\tau} \left\langle \left[ Y_t^2 Y_{t+\tau'} e^{\sqrt{c}\Delta_t W_2(\tau')} \right] \right\rangle \left\langle \left[ e^{\sqrt{c}\Delta_{t+\tau'} W_2(\tau-\tau')} \right] \right\rangle d\tau' \\ &+ \sqrt{c} \int_0^{\tau} \left\langle \left[ Y_t^2 Y_{t+\tau'} e^{\sqrt{c}\Delta_t W_2(\tau')} \right] \right\rangle \left\langle \left[ e^{\sqrt{c}\Delta_{t+\tau'} W_2(\tau-\tau')} dW_{2,t+\tau'} \right] \right\rangle \ , \end{aligned} \tag{B.1}$$

where we changed the variable of integrations to $\tau' = s - t$. Since the process $\sqrt{c}\Delta_{t+\tau'} W_2(\tau-\tau')$ is Normally distributed with zero mean and variance $c(\tau - \tau')$, and recalling the expression of the Gaussian characteristic function, $\phi^G$, we can write

$$\left\langle \left[ \exp\left\{ \sqrt{c}\Delta_{t+\tau'} W_2(\tau-\tau') \right\} \right] \right\rangle = \phi^G(\omega)\big|_{\omega=-i} = \exp\left\{ \frac{c}{2}(\tau-\tau') \right\} \ .$$

Application of the Novikov theorem also gives

$$\left\langle \left[ e^{\sqrt{c}\Delta_{t+\tau'} W_2(\tau-\tau')} dW_{2,t+\tau'} \right] \right\rangle = \left\langle \left[ \frac{\delta}{\delta \zeta_{W_2}(t+\tau')} \exp\left\{ \sqrt{c} \int_{t+\tau'}^{t+\tau} \zeta_{2,s} ds \right\} \right] \right\rangle d\tau'$$
$$= \sqrt{c} \exp\left\{ \frac{c}{2}(\tau - \tau') \right\} ,$$

where we expressed the Wiener variation in terms of a Gaussian white noise $\zeta_{2,t}$ as $dW_{2,t} = \zeta_{2,t}\, dt$. Replacing the previous expressions in Eq. (B.1) we conclude that $f(Y;t,\tau)$ has to satisfy

$$f(Y;t,\tau) - (a+c) \int_0^\tau f(Y;t,\tau') e^{\frac{c}{2}(\tau-\tau')} d\tau' = e^{\frac{c}{2}\tau} \left[ \mu_3(t) + b\tau \mu_2(t) \right] \ . \quad \text{(B.2)}$$

Eq. (B.2) is a Volterra equation of the second kind, whose solution is precisely the expression (3.28).

**Computation of $\left\langle Y_t^2 Y_{t+\tau}^2 \right\rangle$ entering Eq. (3.33).** With reference to the SVM (3.16), the cross correlation $\left\langle Y_t^m Y_{t+\tau}^n \right\rangle$ can be computed exactly. Provided to express $Y_{t+\tau}^n$ as integral solution from $t$ to $t+\tau$

$$Y_{t+\tau}^n = Y_t^n + \int_t^{t+\tau} \left[ F_n Y_s^n + A_n Y_s^{n-1} \right] ds + \int_t^{t+\tau} \dots dW_{2,s}$$

it is straightforward to check that $\left\langle Y_t^m Y_{t+\tau}^n \right\rangle$ satisfies the following equation

$$\frac{d}{d\tau} \left\langle Y_t^m Y_{t+\tau}^n \right\rangle = F_n \left\langle Y_t^m Y_{t+\tau}^n \right\rangle + A_n \left\langle Y_t^m Y_{t+\tau}^{n-1} \right\rangle , \quad \text{(B.3)}$$

which is an ODE provided the correlation $\left\langle Y_t^m Y_{t+\tau}^{n-1} \right\rangle$ was computed at the step $n-1$. Starting from $n=0$, we readily compute the following correlation

$$\left\langle Y_t^2 Y_{t+\tau} \right\rangle = e^{a\tau} \mu_3(t) - \frac{b}{a} \left[ 1 - e^{a\tau} \right] \mu_2(t) \ ;$$

substitution in Eq. (B.3) for the case $m = n = 2$ provides the ODE satisfied by $\left\langle Y_t^m Y_{t+\tau}^{n-1} \right\rangle$, whose solution reads:

$$\left\langle Y_t^2 Y_{\tau(s)}^2 \right\rangle = e^{F_2\tau} \mu_4(t) + \frac{A_2}{a - F_2} \left[ e^{a\tau} - e^{F_2\tau} \right] \mu_3(t)$$
$$- \left\{ \frac{b}{a} \frac{A_2}{F_2} \left[ e^{F_2\tau} - 1 \right] - \frac{b}{a} \frac{A_2}{a - F_2} \left[ e^{a\tau} - e^{F_2\tau} \right] \right\} \mu_{2(t)} \, .$$

Appendix C

# Notations

| | |
|---|---|
| ABM | Arithmetic Brownian Motion |
| B&S | Black & Scholes |
| CC | Conditional Coverage |
| CF | Characteristic Function |
| CL | Confidence Level |
| CLT | Central Limit Theorem |
| DGN | Delta-Gamma-Normal |
| ES | Expected Shortfall |
| ExpOU | Exponential Ornstein-Uhlenbeck |
| FFT | Fast Fourier Transform |
| GBM | Geometric Brownian Motion |
| GFT | Generalized Fourier Transform |
| GST | Generalized Student-$t$ |
| $H$ | The Heaviside step function |
| *i.i.d.* | independent, identically distributed |
| LinExpOU | Linearized Exponential Ornstein-Uhlenbeck |
| log | The logarithm to the base $e$ |
| ML | Maximum Likelihood |
| ODE | Ordinary Differential Equation |
| PD | Product Distribution |
| PDF | Probability Density Function |
| PPM | Product Partition Model |
| *r.h.s.* | right hand side |
| SDE | Stochastic Differential Equation |
| S&P500 | Standard & Poor 500 |
| SVM | Stochastic Volatility Model |
| S2 | Stein-Stein |
| UC | Unconditional Coverage |
| VaR | Value-at-Risk |

# Bibliography


[1] E. Majorana, "Il valore delle leggi statistiche nella fisica e nelle scienze sociali", Scientia **36**, 58 (1942).

[2] A. Einstein, "On the movement of small particles suspended in a stationary demanded by the molecular-kinetic theory of heat", Ann. Phys. **17**, 549 (1905).

[3] M. Smoluchowski, "Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen", Ann. Phys. **326**, 756 (1906).

[4] V. Pareto, *Cours d'economie politique* (Librairie Droz, 1964).

[5] G. Bormetti, M. E. De Giuli, D. Delpini, and C. Tarantola, "Bayesian Value-at-Risk with product partition models", Quant. Finance **doi:10.1080/14697680903512786** (2010).

[6] D. Delpini, G. Bormetti, M. E. De Giuli, and C. Tarantola, in *S. Co. 2009 - Complex data modeling and computationally intensive statistical methods for estimation and prediction. Politecnico di Milano, September 14-16, 2009 Proceedings* (2009).

[7] G. Bormetti, V. Cazzola, D. Delpini, G. Montagna, and O. Nicrosini, "The low volatility fluctuations regime of the exponential Ornstein-Uhlenbeck model", J. Phys. Conf. Ser. **221**, 012014 (2010).

[8] G. Bormetti, V. Cazzola, and D. Delpini, "Option pricing under Ornstein-Uhlenbeck stochastic volatility: a linear model", Int. J. Theoretical Appl. Finance **13**, 1047 (2010).

[9] G. Bormetti and D. Delpini, "Exact moment scaling from multiplicative noise", Phys. Rev. E **81**, 032102 (2010).

[10] G. Bormetti, V. Cazzola, D. Delpini, and G. Livan, "Accounting for risk of non linear portfolios", Eur. Phys. J. B **76**, 157 (2010).

[11] L. Bachélier, Ph.D. thesis (1900).

[12] W. Paul and J. Baschnagel, *Stochastic Processes, From Physics to Finance* (Springer-Verlag, 2000).

[13] M. F. M. Osborne, "Brownian Motion in the stock market", Oper. Res. **7**, 145 (1959).

[14] P. A. Samuelson, "Rational theory of warrant pricing", Ind. Manag. Rev. **6**, 13 (1965).

[15] P. Samuelson and R. Merton, "A complete model of warrant pricing that maximizes utility", Ind. Manag. Rev. **10**, 17 (1969).

[16] J.-P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing: from Statistical Physics to Risk Management* (Cambridge University Press, 2003).

[17] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 2000).

[18] B. B. Mandelbrot, "The variation of certain speculative prices", J. Bus. **36**, 394 (1963).

[19] E. F. Fama, "The behavior of stock market prices", J. Bus. **38**, 34 (1965).

[20] R. N. Mantegna and H. E. Stanley, "Stochastic processes with ultraslow convergence to a Gaussian: the truncated Lévy flight", Phys. Rev. Lett. **73**, 2946 (1994).

[21] I. Koponen, "Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process", Phys. Rev. E **52**, 1197 (1995).

[22] V. Plerou, P. Gopikrishnan, L. A. N. Amaral, M. Meyer, and H. E. Stanley, "Scaling of the distribution of price fluctuations of individual companies", Phys. Rev. E **60**, 6519 (1999).

[23] M. Gell-Mann and C. Tsallis, *Nonextensive entropy: interdisciplinary applications* (Oxford University Press, USA, 2004).

[24] C. Tsallis, C. Anteneodo, L. Borland, and R. Osorio, "Nonextensive statistical mechanics and economics", Physica A **324**, 89 (2003).

[25] G. Bormetti, E. Cisana, G. Montagna, and O. Nicrosini, "A non-Gaussian approach to risk measures", Physica A **376**, 532 (2007).

[26] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C. Peng, and H. E. Stanley, "Statistical properties of the volatility of price fluctuations", Phys. Rev. E **60**, 1390 (1999).

[27] M. Pasquini and M. Serva, "Multiscale behaviour of volatility autocorrelations in a financial market", Econ. Lett. **65**, 275 (1999).

[28] J. Perelló and J. Masoliver, "A correlated stochastic volatility model measuring leverage and other stylized facts", Int. J. Theoretical Appl. Finance **5**, 541 (2002).

[29] S. Micciché, G. Bonanno, F. Lillo, and R. N. Mantegna, "Volatility in financial markets: stochastic models and empirical results", Physica A **314**, 756 (2002).

[30] P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer, and H. E. Stanley, "Scaling of the distributions of fluctuations of financial market indices", Phys. Rev. E **60**, 5305 (1999).

[31] M. M. Dacorogna, U. A. Müller, R. J. Nagler, R. B. Olsen, and O. V. Pictet, "A geographical model for the daily and weekly seasonal volatility in the foreign exchange market", J. Int. Money Finance **12**, 413 (1993).

[32] Z. Ding, C. W. J. Granger, and R. F. Engle, "A long memory property of stock market returns and a new model", J. Empirical Finance **1**, 83 (1993).

[33] J. Perelló, J. Masoliver, and J.-P. Bouchaud, "Multiple time scales in volatility and leverage correlations: a stochastic volatility model", Appl. Math. Finance **11**, 27 (2004).

[34] J. Masoliver and J. Perelló, "Multiple time scales and the exponential Ornstein-Uhlenbeck stochastic volatility model", Quant. Finance **6**, 423 (2006).

[35] Y. Liu, P. Cizeau, M. Meyer, C. K. Peng, and H. E. Stanley, "Correlations in economic time series", Physica A **245**, 437 (1997).

[36] E. F. Fama, "Efficient capital markets: A review of theory and empirical work", J. Finance **25**, 383 (1970).

[37] M. Raberto, E. Scalas, G. Cuniberti, and M. Riani, "Volatility in the Italian stock market: an empirical study", Physica A **269**, 148 (1999).

[38] M. M. Dacorogna, U. A. Müller, R. J. Nagler, R. B. Olsen, and O. V. Pictet, "A geographical model for the daily and weekly seasonal volatility in the foreign exchange market", J. Int. Money Finance **12**, 413 (1993).

[39] H. Risken and H. Haken, *The Fokker-Planck Equation: Methods of Solution and Applications Second Edition* (Springer, 1989).

[40] P. Langevin, "The theory of Brownian movement", CR Acad. Sci **146**, 530 (1908).

[41] T. S. Biró and A. Jakovác, "Power-law tails from multiplicative noise", Phys. Rev. Lett. **94**, 132302 (2005).

[42] A. Schenzle and H. Brand, "Multiplicative stochastic processes in statistical physics", Phys. Rev. A **20**, 1628 (1979).

[43] C. Renner, J. Peinke, and R. Friedrich, "Experimental indications for Markov properties of small-scale turbulence", J. Fluid Mech. **433**, 383 (2001).

[44] R. Friedrich, J. Peinke, and C. Renner, "How to quantify deterministic and random influences on the statistics of the foreign exchange market", Phys. Rev. Lett. **84**, 5224 (2000).

[45] R. Friedrich and J. Peinke, "Description of a turbulent cascade by a Fokker-Planck equation", Phys. Rev. Lett. **78**, 863 (1997).

[46] G. Wilk and Z. Włodarczyk, "Interpretation of the nonextensivity parameter $q$ in some applications of Tsallis statistics and Lévy distributions", Phys. Rev. Lett. **84**, 2770 (2000).

[47] L. Borland, "Microscopic dynamics of the nonlinear Fokker-Planck equation: A phenomenological model", Phys. Rev. E **57**, 6634 (1998).

[48] J. L. McCauley, G. H. Gunaratne, and K. E. Bassler, "Hurst exponents, Markov processes, and fractional Brownian motion", Physica A **379**, 1 (2007).

[49] F. Ghasemi, M. Sahimi, J. Peinke, and M. R. R. Tabar, "Analysis of non-stationary data for heart-rate fluctuations", J. Biol. Phys. **32**, 117 (2006).

[50] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge, "Turbulent cascades in foreign exchange markets", Nature **381**, 767 (1996).

[51] L. Borland, "Option pricing formulas based on a non-Gaussian stock price model", Phys. Rev. Lett. **89**, 098701 (2002).

[52] L. Borland, "A theory of non-Gaussian option pricing", Quant. Finance **2**, 415 (2002).

[53] F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, P. G. Lind, and M. Haase, "Reconstruction of complex dynamical systems affected by strong measurement noise", Phys. Rev. Lett. **97**, 090603 (2006).

[54] F. Ghasemi, M. Sahimi, J. Peinke, R. Friedrich, G. R. Jafari, and M. R. R. Tabar, "Markov analysis and Kramers-Moyal expansion of nonstationary stochastic processes with application to the fluctuations in the oil price", Phys. Rev. E **75**, 060102 (2007).

[55] G. E. Uhlenbeck and L. S. Ornstein, "On the theory of the Brownian motion", Phys. Rev. **36**, 823 (1930).

[56] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics", J. Stat. Phys. **52**, 479 (1988).

[57] J. P. Fouque, G. Papanicolaou, and K. R. Sircar, *Derivatives in Financial Markets with Stochastic Volatility* (Cambridge University Press, 2000).

[58] S. L. Heston, "A closed-form solution for options with stochastic volatility with applications to bond and currency options", Rev. Financ. Stud. **6**, 327 (1993).

[59] J. Cox, J. Ingersoll, and S. Ross, "A theory of the term structure of interest rates", Econometrica **53**, 385 (1985).

[60] W. Feller, "Two singular diffusion problems", Ann. Math. **54**, 173 (1951).

[61] A. A. Dragulescu and V. M. Yakovenko, "Probability distribution of returns in the Heston model with stochastic volatility", Quant. Finance **2**, 443 (2002).

[62] E. M. Stein and J. C. Stein, "Stock price distributions with stochastic volatility: An analytic approach", Rev. Financ. Stud. **4**, 727 (1991).

[63] R. Schoebel and J. Zhu, "Stochastic volatility with an Ornstein-Uhlenbeck process: An extension", SSRN eLibrary (1998).

[64] L. O. Scott, "Option pricing when the variance changes randomly: Theory, estimation, and an application", J. Financ. Quant. Anal. **22**, 419 (1987).

[65] J. Perelló, "Market memory and fat tail consequences in option pricing on the expOU stochastic volatility model", Physica A **382**, 213 (2007).

[66] J. Perelló, R. Sircar, and J. Masoliver, "Option pricing under stochastic volatility: the exponential Ornstein–Uhlenbeck model", J. Stat. Mech. **2008**, P06010 (2008).

[67] G. Bormetti, V. Cazzola, G. Montagna, and O. Nicrosini, "The probability distribution of returns in the exponential Ornstein–Uhlenbeck model", J. Stat. Mech. **2008**, P11013 (2008).

[68] J. Perelló, J. Masoliver, and N. Anento, "A comparison between several correlated stochastic volatility models", Physica A **344**, 134 (2004).

[69] G. 't Hooft and M. Veltman, "Scalar one-loop integrals", Nucl. Phys. B **153**, 365 (1979).

[70] E. A. Novikov, "Functionals and the random-force method in turbulence theory", Sov. Phys. JETP-USSR **20**, 1290 (1965).

[71] J. Bouchaud, Y. Gefen, M. Potters, and M. Wyart, "Fluctuations and response in financial markets: the subtle nature ofrandom'price changes", Quant. Finance **4**, 176 (2004).

[72] J. Bouchaud, J. Kockelkoren, and M. Potters, "Random walks, liquidity molasses and critical response in financial markets", Quant. Finance **6**, 115 (2006).

[73] J. A. Hartigan, "Partition models", Commun. Stat. A **19**, 2745 (1990).

[74] D. Barry and J. A. Hartigan, "Product partition models for change point problems", Ann. Stat. **20**, 260 (1992).

[75] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", IEEE 6 **6**, 721 (1984).

[76] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities", J. Am. Stat. Assoc. **85**, 398 (1990).

[77] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines", J. Chem. Phys. **21**, 1087 (1953).

[78] W. Hastings, "Monte carlo sampling methods using Markov chains and their applications", Biometrika **57**, 97 (1970).

[79] P. G. Hoel, S. C. Port, and S. C. J., *Introduction to stochastic processes*, The Houghton Mifflin series in statistics (Houghton Mifflin, Boston [u.a.], 1972).

[80] D. Dey, P. Müller, and D. Sinha, *Practical nonparametric and semiparametric Bayesian statistics* (Springer New York, 1998).

[81] C. E. Antoniak, "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems", Ann. Stat. **2**, 1152 (1974).

[82] C. A. Bush and S. N. MacEachern, "A semiparametric Bayesian model for randomised block designs", Biometrika **83**, 275 (1996).

[83] F. A. Quintana and P. L. Iglesias, "Bayesian clustering and product partition models", J. R. Stati. Soc. B **65**, 557 (2003).

[84] S. Petrone, G. Roberts, and J. Rosenthal, "A note on convergence rates of Gibbs sampling for nonparametric mixtures", Unpublished manuscript (1998).

[85] F. Black and M. Scholes, "The pricing of options and corporate liabilities", J. Polit. Econ. **81**, 637 (1973).

[86] R. Merton, "Theory of rational option pricing", Bell J. Econ. Manag. Sci. **4**, 141 (1973).

[87] P. Samuelson, "Proof that properly anticipated prices fluctuate randomly", Manag. Rev. **6**, 41 (1965).

[88] J. M. Harrison and S. R. Pliska, "Martingales and stochastic integrals in the theory of continuous trading", Stoc. Proc. Appl. **11**, 215 (1981).

[89] "Basel II: International convergence of capital measurement and capital standards: a revised framework", (2004), Basel Committee.

[90] D. Backus, S. Foresi, K. Li, and W. Wu, "Accounting for biases in Black-Scholes", Working paper series-New York University Salomon Center S (1997).

[91] S. Ciliberti, J. Bouchaud, and M. Potters, "Smile dynamics: a theory of the implied leverage effect", Wilmott Journal **1**, 87 (2009).

[92] D. Lamberton and B. Lapeyre, *Introduction to stochastic calculus applied to finance* (Chapman and Hall, London, 1996).

[93] A. L. Lewis, in *Other Exponential Lévy Processes," Environ Financial Systems and OptionCity.net* (2001).

[94] E. Lukacs, *Characteristic functions* (Hafner Publ. Co., 1970), 2nd ed.

[95] C. Rouvinez, "Going greek with var", Risk **10**, 57 (1997).

[96] R. Martin, "Shortfall: who contributes and how much?", Risk Magazine **22**, 84 (2009).

[97] G. Bormetti, V. Cazzola, G. Livan, G. Montagna, and O. Nicrosini, "A generalized Fourier transform approach to risk measures", J. Stat. Mech. **2010**, P01005 (2010).

[98] H. David and H. Nagaraja, *Order statistics* (Wiley-Interscience, 2003).

[99] P. Kupiec, "Techniques for verifying the accuracy of risk measurement models", J. Derivatives **3**, 73 (1995).

[100] P. Christoffersen, "Evaluating interval forecasts", Int. Econ. Rev. **39**, 841 (1998).

[101] P. Jorion, *Value at Risk* (McGraw-Hill, New York, 2007), 3rd ed.

[102] S. Jaschke, "The Cornish-Fisher expansion in the context of delta-gamma-normal approximations", J. Risk **4**, 33 (2002).

[103] S. Jaschke, C. Klüppelberg, and A. Lindner, "Asymptotic behavior of tails and quantiles of quadratic forms of Gaussian vectors", J. Multivariate Anal. **88**, 252 (2004).

[104] J. Mina and A. Ulmer, "Delta-Gamma four ways", RiskMetrics Group (1999).

[105] D. Duffie and J. Pan, "An overview of Value at Risk", J. Derivatives **4**, 7 (1997).

[106] R. H. Loschi, F. R. B. Cruz, P. L. Iglesias, and R. B. Arellano-Valle, "A Gibbs sampling scheme to the product partition model: an application to change-point problems", Comput. Oper. Res. **30**, 463 (2003).

[107] P. Glasserman, P. Heidelberger, and P. Shahabuddin, "Portfolio Value-at-Risk with heavy-tailed risk factors", Math. Finance **12**, 239 (2002).

[108] M. Tumminello, F. Lillo, and R. N. Mantegna, "Kullback-Leibler distance as a measure of the information filtered from multivariate data", Phys. Rev. E **76**, 031123 (2007).

[109] M. Tumminello, F. Lillo, and R. Mantegna, "Hierarchically nested factor model from multivariate data", Europhys. Lett. **78**, 30006 (2007).

[110] M. Tumminello, F. Lillo, and R. Mantegna, "Kullback-Leibler distance as a measure of the information filtered from multivariate data", Phys. Rev. E **76**, 31123 (2007).

[111] M. Tumminello, F. Lillo, and R. Mantegna, "Shrinkage and spectral filtering of correlation matrices: a comparison via the kullback-leibler distance", Acta Phy. Pol. B **38**, 4079 (2007).

# Acknowledgments

*Desidero ringraziare il mio supervisore, il Prof. Oreste Nicrosini, per la sua costante disponibilità e per la sua concretezza nell'affrontare i problemi, e il Prof. Guido Montagna per la sua capacità di comunicare l'entusiasmo della ricerca. Un ringraziamento particolare va al Dott. Giacomo Bormetti, collega insuperabile ma soprattutto grande amico, con il quale è stato un piacere e un onore lavorare. Ringrazio il Prof. Fulvio Piccini, esempio per chiunque aspiri a diventare un bravo ricercatore. E grazie anche al Dott. Giovanni Balossini, a Luca Barzè, Christopher Bignamini, Valentina Cazzola e Giacomo Livan per aver reso il nostro corridoio nel Dipartimento di Fisica un luogo al quale affezionarsi.*

*Per la fruttuosa collaborazione ringrazio la Prof.ssa Maria Elena De Giuli e la Dr.ssa Claudia Tarantola, il Dott. Andrea Pallavicini per le utili discussioni e il Prof. Fabrizio Lillo per aver accettato di leggere questo mio lavoro.*

I wish to thank my supervisor, Prof. Oreste Nicrosini, for his tireless helpfulness and his pragmatism in dealing with problems, and Prof. Guido Montagna for his ability in passing on its enthusiasm for the research activity. A special thanks goes to Dr. Giacomo Bormetti, unequaled colleague and most of all a great friend, working with whom was a pleasure and an honor. Thanks to Prof. Fulvio Piccinini, a model for anyone aiming to become a good researcher. And also thanks to Dr. Giovanni Balossini, to Luca Barzè, Christopher Bignamini, Valentina Cazzola and Giacomo Livan for contributing to make our hallway at the Physics Department a place which to become really attached to.

For fruitful collaboration I wish to thank Prof. Maria Elena De Giuli and Dr. Claudia Tarantola, Dr. Andrea Pallavicini for useful discussions and Prof. Fabrizio Lillo for accepting to referee this work.