Universidade de Lisboa
Faculdade de Ciências
Departamento de Matemática


# Stochastic evolution of parameters defining probability density functions:

Application to the New York stock market

**Paulo Henrique Contente Rocha**

Dissertação

Mestrado em Matemática

**2014**

Universidade de Lisboa
Faculdade de Ciências
Departamento de Matemática


# Stochastic evolution of parameters defining probability density functions:

Application to the New York stock market

**Paulo Henrique Contente Rocha**

Dissertação

Mestrado em Matemática

Orientação: João Pedro Boto e Pedro G. Lind

**2014**

# Resumo

Nesta dissertação, nós estudamos a evolução de séries temporais não estacionarias com o objetivo de extrair equações diferencias estocásticas apartir dos dados que descrevam a dinâmica do sistema. Centramos o nosso estudo na bolsa de Nova York e testamos quatro modelos bi-parametricos para fitar as distribuições de volume-preço a cada 10 minutos.

Usando os devios relativos e introduzindo uma nova variante da divergência de Kullback-Leibler, argumentamos que o melhor modelo para a distribuição empírica do volume-preço não é sempre o mesmo, e depende de (i) da região do espectro que se pretende modelar e (ii) no período de tempo em que se modela.

Focamo-nos no modelo da Gama inversa, pois apresenta o melhor fit para descrever as caudas da distribuição empírica e estudamos a evolução dos parâmetros que a caracterizam como um processo estocástico. Particularmente, assumimos que a evolução dos parâmetros da distribuição Gama inversa são governados por uma equação de Langevin e derivamos os correspondentes coeficientes de drift e de difusão. Estes fornecem-nos informação que nos permite compreender os mecanismos responsáveis pelo comportamento da bolsa de valores e consequentemente, fazer uma melhor estimativa do risco associado.

O primeiro capítulo expõe o problema que pretendemos tratar nesta dissertação. No segundo capítulo, introduzimos a teoria necessária para entender os conceitos apresentados nos capítulos seguintes. No terceiro capítulo apresentada a metodologia seguida durante o processamento dos dados. No quarto capítulo discutimos qual o melhor modelo para descrever o comportamento da distribuição de volume-preço. No quinto capítulo apresentamos um modelo estocástico para descrever a evolução das caudas da distribuição volume-preço. Por fim, a secção "Discussions and conclusions" encerra a dissertação, onde descrevemos como é que a metodologia aqui seguida pode ser estendida para funções densidade de probabilidade como problema matemático mais geral.

Palavras-chave: Distribuições estocásticas, volatilidade, bolsa de valores, volume-preço.

# Abstract

In this thesis we study the evolution of non-stationary with the aim of extracting the stochastic equations describing it from sets of empirical data. We apply our framework to the New York Stock market (NYSM). We test four different bi-parametric models to fit the correspondent volume-price distributions at each 10-minute lag.

Using the relatives deviations and by introducing a new variant of Kullback-Leibler divergence we present quantitative evidence that the best model for empirical volume-price distributions is not always the same and it strongly depends in (i) the region of the volume-price spectrum that one wants to model and (ii) the period in time that is being modelled.

We then focus in the inverse Gamma distribution which shows to be the best model for describing the tail of the empirical distributions and analyse the evolution of its parameters as a stochastic process. Namely, we assume that the evolution of the inverse Gamma parameters is governed by Langevin equation and derive the corresponding drift and diffusion coefficients. These coefficients provide insight for understanding the mechanisms underlying the evolution of the stock market, and bound the risk associated with such distributions.

The first chapter poses the problem and scope of the thesis. In the second chapter we introduce the theory necessary to understand the concepts addressed in the following chapters. In Chapter 3 we present the methodology used for processing the NYSM data. In the fourth chapter we discuss which model is the best one to describe the volume-price distribution. In the fifth chapter we present a stochastic model to describe the evolution of the distribution tails. Discussions and conclusions closes the thesis, where we describe how the framework proposed in this thesis can be extended to non-stationary probability density functions as a general mathematical problem.

Keywords: Stochastic distributions, volatility, stock market, volume-price.

# Previous Work

The main results of this thesis were already published or submitted for publication in international journals and proceedings:

- **"Stochastic Evolution of Stock Market Volume-Price Distributions"**[17] submitted in June 2014.
- **"Optimal models of extreme volume-prices are time-dependent"**[16] submitted in September 2014.

These were also presented in three intentional conferences as poster or oral presentation:

- Oral presentation at "SMTDA 2014", 3rd Stochastic Modelling Techniques and Data Analysis International Conference (**Lisbon**).
- Poster presentation at "Verhandlungen 2014" (**Dresden**).
- Oral presentation at "IC-MSQUARE 2014", International Conference on Mathematical Modelling in Physical Sciences (**Madrid**)

# Acknowledgements

The author would like to acknowledge to his coordinators, Pedro G. Lind and João Pedro Boto, for offered invaluable assistance, support and guidance. They inspired me greatly to work in this project. A especial thanks goes to Frank Raischel for availability and also for the precious help.

The author also would like to thanks Centro de Física Teórica e Computacional (CFTC) that provided me with the initial grant which was the starting point for this master thesis. Special thanks should also be given to Centro de Matemática e Aplicações Fundamentais da Universidade de Lisboa (CMAF) which in collaboration with Fundação da Faculdade de Ciências gave me financial support to attend the conferences: DPG-Frühjahrstagung 2014, SMTDA 2014 and IC-MSQUARE 2014.

Finally, the author would like to acknowledge all his family and friends for all the support through all this years of academic life.

# Contents

# Part I

# Background

# Chapter 1

# Introduction and Scope

Most of probability theory is devoted to the macroscopic picture emerging from stochastic dynamical systems defined by a host of microscopic random effects. The Brownian motion is the macroscopic picture emerging from the random movement of a particle. On a microscopic level, the particle experiences a random displacement caused, for example, by collisions with other neighbour particles or by external forces. If the initial position at $t_0$ is $x_0$, then at time $t_n$ the new position is given by $x_n = x_0 + \sum_{i=1}^{n} \Delta_x^i$, where the displacements $\{\Delta_x^i\}$ are assumed to be independent, identically distributed random variables. The process $\{x_n, n \geqslant 0\}$ is a random walk and the displacements $\{\Delta_x^i\}$ represent the microscopic increments.

The discovery of Brownian motion is credited to the botanist Robert Brown in 1827. In 1905, Albert Einstein initiated the modern study of random processes after publishing his famous paper "Investigations on the theory of the Brownian movement" [5]. Starting from reasonable hypotheses, Einstein derived and solved a differential equation governing the time evolution of the probability density of a Brownian particle. He was able to write a formula that predicts the mean square displacement of a spherical particle in a fluid.

Three years later (1908), the French physicist Paul Langevin suggested a different approach describing Brownian motion, in his own words, "infinitely more simple" [11]. Langevin applied the Newton's second law of movement to a Brownian particle, deriving what is now called the Langevin equation. Langevin described the velocity as one stationary, Gaussian, and Markovian process so-called Ornstein-Uhlenbeck[19] process while Einstein described it's position as a driftless Wiener process.

Both these descriptions have been generalized into mathematically distinct tools for studying an important class of continuous random processes so-called Markov processes, namely the Langevin equation and the Fokker-Planck equation.

A very important application of Langevin's equation can be found in finance, the so-called Black-Scholes model, this model is a mathematical framework that allows one to describe the evolution of options prices [2], published by Fischer Black and Myron Scholes in their paper "The Pricing of Options and Corporate Liabilities" in 1973.

From this model we can derive a Langevin-type equation, called the Black-Scholes equation, which estimates European options prices. The formula led to a boom in options trading and scientifically legitimised the activities of options markets around the world. Robert C. Merton[14] was the first to publish a paper expanding the mathematical understanding of the options pricing model. In recognition of their work, Robert Merton and

Myron Scholes received the 1997 Nobel Prize in Economics. Fischer Black was already dead at the time.

Despite its initial success and acceptance, the Black-Scholes model has drawbacks. The first drawback is the assumption of a continuous trading, which is matched most closely by foreign exchange markets (or FOREX, is a worldwide decentralized over-the-counter financial market), but which is not fulfilled by other exchange platforms.

The second drawback is the assumption of a continuous price path, which is unrealistic. This is corroborated by the existence of opening gaps, which is the difference between opening day trade price and closing day trade price which is a direct consequence of a non-continuous trading. At the beginning of an exchange trading day, the price of a stock does not necessarily start being traded at the same price that it has at the end of the last day trading session. For example, if there is an important news released in the period of time in which the market is closed, larger opening gaps will occur. It is also important to note that an important input variable of the model, the so-called volatility of future returns [4], is not known in advance.

Third, the Black Scholes model is based on normally distributed asset returns. This aspect contradicts earlier findings by Mandelbrot[13] in 1963 and the fact that the Gaussian approximation in finance became more and more questionable.

Finally, the statistical description of the stock market is often based on stationary random processes. Black-Scholes model implicitly assume a geometric Brownian motion with constant drift and standard deviation or "volatility". It is well known that empirical returns are not normal distributed and that their drift and volatility cannot be assumed constant[18].

The assumptions in the model have been relaxed and generalized in a variety of directions, through the years, leading to different variants of the model that are used today in finance, which aim to overcome these drawbacks. One of them is the scope of the present thesis.

In this thesis we study non-stationary probability density functions and apply our findings to the specific case of volume-price distribution in the NYSM.

In order to find a good fit to the empirical cumulative density function we will consider four well-known bi-parametric distributions, namely the $\Gamma$-distribution, inverse $\Gamma$-distribution, log-normal and the Weibull-distribution [4].

We fit at each 10 minutes time-lag the empirical distribution with each one of these four models and record the respective parameters values, yielding time series for the distribution parameters, which can then be analysed.

Finally, taking the time series of these model parameters, we propose a framework for describing their stochastic evolution. Namely, our approach retrieves the functions, called drift and diffusion coefficients, governing the Fokker-Planck equation for the probability density function of parameters values, as we will see. The physical interpretation of these functions will shed new light to understand the dynamics of the empirical distributions and to provide additional insight concerning the non-stationary evolution of probability density functions in several contexts.

# Chapter 2

# State of the Art

In the following section we will consider $X_n$, where $n \geq 1$ , denoting $\mathbb{R}^d$-valued random variables defined on a probability density space $(\Omega, \mathfrak{F}, P)$. $\Omega$ denotes a set of the possible outcomes where a typical element is $\omega \in \Omega$, $\mathfrak{F}$ denotes a sigma algebra generated by $X(\omega)$ in $\Omega$ and $P$ is called a measure. Some times we may need to refer to the sigma algebra generated by the Borel sets in $\mathbb{R}^d$ which we represent by $\mathfrak{B}^d$.

## 2.1 Convergence Concepts

Here we present some important notions on convergence. Consider $X$ and $X_n$, with $n \geq 1$, random variables $\mathbb{R}^d$-valued defined upon a probability space $(\Omega, \mathfrak{F}, P)$. First we shall define the **almost certainly (ac)** convergence and then we define the **stochastic convergence (st)**.

- Consider a set $N \in \mathfrak{F}$ with zero measure, such that, for all $\omega \notin N$, the sequence of $X_n(\omega) \in \mathbb{R}^d$ converges in the usual sense to $X(\omega) \in \mathbb{R}^d$, then $\{X_n\}$ is said to converge almost certainly (or with probability one) to $X$. It can be written as

$$\lim_{n\to\infty}^{ac} X_n(\omega) = X(\omega)\,. \tag{2.1}$$

- We have stochastic convergence (or in probability) of $\{X_n(\omega)\}$ to $X(\omega)$, if for every $\epsilon > 0$

$$p_n(\epsilon) = P\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\} \to 0 (n \to \infty)\,. \tag{2.2}$$

  We write

$$\lim_{n\to\infty}^{st} X_n(\omega) = X(\omega)\,. \tag{2.3}$$

Almost certainly convergence implies stochastic convergence.

## 2.2 Markov Process

In 1906 A. A. Markov laid the groundwork for theory of Markov stochastic processes. He formulated the principle that the future state of a system is independent of the past when

we have information about the present. One can see this as the causality principle of classical physics carried over to stochastic dynamic systems. It specifies that knowing the state of a system at a given point on time is sufficient to determine its state at any given time in the future. For example in the theory of ordinary differential equations, given the differential equation:

$$\frac{dx}{dt} = f(x(t), t) \tag{2.4}$$

the change taking place in $x(t)$ at a time $t$ depends only on $x(t)$ and $t$ and not on the values of $x(s)$ with $s < t$. A direct consequence of this is that, under certain conditions on $f$, the solution curve for $x(t)$ is uniquely determined by an initial point $(x_0, t_0)$. We say that the system has no memory.

Carrying over this idea to stochastic dynamic systems, we get the Markov property. It states that if the state of a system at a particular initial time $t_0$ is known, the behaviour of the system at any given time $s < t$ has no effect on the knowledge about the system beyond.

The mathematical definition of the Markov property is.

**Definition 2.2.1 (Markov process).** *A stochastic process $\{X_t, t \in [t_0, T]\}$ defined on the probability space $(\Omega, \mathfrak{F}, P)$ with index set $[t_0, T] \subset [0, \infty)$ and with state space $\mathbb{R}^d$ is called a Markov process if the following so-called Markov property is satisfied: For $t_0 \leq s \leq t \leq T$ and all $B \in \mathfrak{B}^d$, the equation*

$$P(X_t \in B | \mathfrak{F}([t_0, s])) = P(X_t \in B | X_s) \tag{2.5}$$

*holds with probability 1.*

Let $X_t$ be a Markov process for $t \in [t_0, T]$. There will be a conditional distribution [1] $P(s, X_s, t, B)$ corresponding to the conditional probability $P(X_t \in B | X_s)$. The function $P(s, X_s, t, B)$, $s$ and $t \in [t_0, T]$ with $s \leq t$, $x \in \mathbb{R}^d$ and $B \in \mathfrak{B}^d$ has the following properties:

P1. For fixed $s \leq t$ and $B \in \mathfrak{B}^d$, we have with probability 1

$$P(s, X_s, t, B) = P(X_t \in B | X_s)\,. \tag{2.6}$$

P2. $P(s, X_s, t, .)$ is a probability on $\mathfrak{B}^d$ for fixed $s \leq t$ and $x \in \mathbb{R}^d$.

P3. $P(s, ., t, B)$ is $\mathfrak{B}^d$-measurable for fixed $s \leq t$ and $x \in \mathbb{R}^d$.

P4. For $t_0 \leq s \leq u \leq t \leq T$ and $B \in \mathfrak{B}^d$ and for almost all $x \in \mathbb{R}^d$ the Chapman-Kolmogorov equation

$$P(s, x, t, B) = \int_{\mathbb{R}^d} P(u, y, t, B) P(s, x, u, dy) \tag{2.7}$$

holds.

P5. It is always possible to choose $P(s, x, t, B)$ in such a way that for all $s \in [t_0, T]$ and $B \in \mathfrak{B}^d$ we have:

$$P(s, x, s, B) = I_B(x) = \begin{cases} 1 & \text{for } x \in B \\ 0 & \text{for } x \notin B\,. \end{cases} \tag{2.8}$$

**Definition 2.2.2** (**Transition probability**). *A function $P(x, s, B, t)$ obeying (P2-P5) is called a transition probability. If $X_t$ is a Markov process and $P(s, x, t, B)$ is a transition probability, so that (P5) is satisfied, then $P(s, x, t, B)$ is called a transition probability of the Markov process.*

**Definition 2.2.3** (**Homogeneous Markov process**). *A Markov process $X_t$ for $t \in [t_0, T]$ is said to be homogeneous with respect to time, if its transition probability $P(s, x, t, B)$ is stationary, that is, if the condition*

$$P(s + u, x, t + u, B) = P(s, x, t, B) \tag{2.9}$$

*is identically satisfied for $t_0 \leq s \leq t \leq T$ and $t_0 \leq s + u \leq t + u \leq T$.*

In this case, the transition probability is a function only of $x$, $t - s$ and $B$. Hence, we can write it in the form $P(t - s, x, B)$ for $0 \leq t - s \leq T - t_0$. Consequently, function $P(t - s, x, B)$ is the probability of transition from $x$ to $B$ in time $t - s$, regardless time $t$ and $s$. Thus, for homogeneous processes, the Chapman-Kolmogorov equation becomes

$$P(t + s, x, B) = \int_{\mathbb{R}^d} P(s, y, B)P(t, x, dy)\,. \tag{2.10}$$

## 2.3 White Noise and the Wiener process

The Wiener process is a mathematical model of the Brownian motion of a free particle with no friction. This process is a time and space homogeneous diffusion process with zero drift coefficient. The features of the Wiener process compose the fundamental building block for all (smooth) diffusion processes. Some of the most important properties of a Wiener process are:

W1. Since $W_t$ is a Markov process, all the distributions of $W_t$ are defined by the initial condition

$$W_0 = 0\,. \tag{2.11}$$

W2. A $W_t$ is a Gaussian stochastic process with expectation value $\mathbf{E}(W_t) = 0$ and covariance matrix:

$$\mathbf{E}(W_t W_s) = \min(t, s)I\,. \tag{2.12}$$

W3. $W_t$ is invariant under rotations in $\mathbb{R}$ [1].

W4. If $W_t$ is a Wiener process, the process $-W_t$, $cW_{\frac{t}{c^2}}$ (where $c \neq 0$), $tW_{\frac{1}{t}}$ and $W_{t+s} - W_s$ (where $s$ is fixed and $t \geqslant 0$) are also Wiener processes.

The white noise is generally understood as a stationary Gaussian process, $\Gamma(t)$ defined in $-\infty < t < \infty$, with mean $\mathbf{E}(\Gamma(t)) = 0$ and a constant spectral density $f(\lambda)$ in all real axis. If $\mathbf{E}(\Gamma(t)\Gamma(t + s)) = C(t)$ is the covariance function of $\Gamma(t)$, then

$$f(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\lambda t}C(t)dt = \frac{c}{2\pi} \tag{2.13}$$

for all $\lambda \in \mathbb{R}$, where $c > 0$ is a constant.

Such process has a spectrum in which all frequencies participate with the same intensity. Therefore the name "white noise". The last equation is only compatible with $C(t) = \delta(t)$ the Dirac's delta function, which means that in the traditional sense, such process does not exist in the real world.

In order to make the connection between the white noise and the Wiener process, we start with the fact that, in every measurement of the function $f(t)$, the inertia of the measurement allows us to have access to the average value

$$\Phi_f(\phi) = \int_{-\infty}^{\infty} \phi(t) f(t) dt \,, \tag{2.14}$$

where $\phi(t)$ is a function that characterises the measuring instrument. The function $\Phi_f$ is the generalized function corresponding to $f(t)$ and it is linear and continuous on $\phi$. As a result of this smoothness, we obtain a value for the last integral, even if the function $f(t)$ is not continuous. More precisely, we can define the generalized functions as:

**Definition 2.3.1** (**Generalized functions**). *Let $K$ be the space of all $C_0^\infty(\mathbb{R})$ functions. A sequence $\phi_1(t), \phi_2(t), ..., \phi_i(t), ...$ of such functions is said to converge to $\phi(t) \equiv 0$ if all the functions vanish outside of a single bounded region and if all of them and all of derivatives converge uniformly to zero.*

*Continuous linear functional $\Phi$ defined on the space $K$ is called a generalized function (or, as is communally called in functional analysis, a distribution).*

For example, a generalized function defined as

$$\Phi(\phi) = \phi(t_0) \tag{2.15}$$

for all $\phi \in K$, $t_0 \in \mathbb{R}$ fixed, called the Dirac's delta-function.

In contrast with the classical functions, generalized functions always have derivatives of every order, which are again generalized functions. The derivative of $\Phi$ is given by:

$$\frac{d\Phi(\phi)}{dt} = -\Phi\left(\frac{d\phi}{dt}\right) . \tag{2.16}$$

With the help of generalized functions we are now able to extend this concept to stochastic processes.

**Definition 2.3.2** (**Generalized stochastic processes**). *A generalized stochastic process is a random generalized function in the following sense: to every $\phi \in K$ is assigned a random variable $\Phi(\phi)$ such that the following two conditions hold:*

- *The functional $\Phi$ is linear in $K$ with probability 1. So for an arbitrary function $\phi$ and $\psi$ in $K$ and for arbitrary constants $\alpha$ and $\beta$, the following is satisfied with probability 1:*

$$\Phi(\alpha\phi + \beta\psi) = \alpha\Phi(\phi) + \beta\Phi(\psi) \tag{2.17}$$

- *The generalized function $\Phi(\phi)$ is continuous[1].*

A generalized stochastic process is said to be Gaussian if, for arbitrary linearly independent functions $\phi_1, \phi_2, ..., \phi_n \in K$, the random variable $(\Phi(\phi_1), \Phi(\phi_2), ..., \Phi(\phi_n))$ is normally distributed. Like in the classical case, a generalized Gaussian process is uniquely defined by the continuous linear mean-value functional:

$$\mathbf{E}(\Phi(\phi)) = m(\phi) \tag{2.18}$$

and the continuous bilinear positive-defined covariance functional:

$$\mathbf{E}((\Phi(\phi) - m(\phi))(\Phi(\psi) - m(\psi))) = C(\phi, \psi) \tag{2.19}$$

One of the most important features of the generalized stochastic process is the fact that its derivatives always exit and is itself a generalized process. Generalized stochastic processes are important for addressing Wiener processes. Consider the Wiener process and its generalized derivative, from:

$$\Phi(\phi) = \int_0^\infty \phi(t) W_t dt \tag{2.20}$$

with $W_t$ regard as a generalized Gaussian stochastic process. We have

$$\mathbf{E}(\phi) = m(\phi) = 0 \tag{2.21}$$

and

$$C(\phi, \psi) = \int_0^\infty \int_0^\infty \min(t, s)\phi(t)\psi(s) dt ds \,. \tag{2.22}$$

Then, the "derivative" of a Wiener process can be computed as follows. Consider the generalized derivative of the mean value $\frac{dm}{dt}(\phi) = 0$ and of the covariance

$$\frac{dC}{dt}(\phi, \psi) = C\left(\frac{d\phi}{dt}, \frac{d\psi}{dt}\right) = \int_0^\infty \phi(t)\psi(t) dt \,. \tag{2.23}$$

We can rewrite the last formula as:

$$\frac{dC}{dt}(\phi, \psi) = \int_0^\infty \int_0^\infty \delta(t - s)\phi(t)\psi(t) dt ds \,. \tag{2.24}$$

From eq(2.24), the covariance of the "derivative" of the Wiener process is the generalized function

$$\frac{dC}{dt} = \delta(t - s) \,. \tag{2.25}$$

But as we saw earlier, this is the covariance of the white noise. Thus, the white noise $\Gamma(t)$ is the "derivative" of the Wiener process $W_t$ if taken as a generalized stochastic process. This justifies the notation

$$\frac{dW_t}{dt} = \Gamma(t) \tag{2.26}$$

which in the integral form reads

$$W_t = \int_0^t \Gamma(s) ds \,. \tag{2.27}$$

## 2.4 The Langevin Equation

We consider the continuous Markov process $X_t \in \mathbb{R}^d$. A process $X_t$ with $t \in [t_0, T]$ is said to be continuous if almost all sample functions of the process are continuous in $[t_0, T]$. We now focus our attention to an important class of these processes, the class of the so-called diffusion processes [1].

**Definition 2.4.1 (Diffusion processes).** *A Markov process $X_t$, for $t_0 \leq t \leq T$, with values in $\mathbb{R}^d$ and almost certainly continuous sample functions is called a diffusion process if its transition probability given by $P(s, x, t, B)$ satisfies the following conditions for every $s \in [t_0, T)$, $x \in \mathbb{R}^d$ and $\epsilon > 0$.*

*L1.*

$$\lim_{s \to t} \frac{1}{t-s} \int_{|y-x|>\epsilon} P(s, x, t, dy) = 0\,. \tag{2.28}$$

*L2. There exists a function $f(s, x)$ with values in $\mathfrak{R}^d$ such that*

$$\lim_{s \to t} \frac{1}{t-s} \int_{|y-x|\leq\epsilon} (y-x) P(s, x, t, dy) = f(s, x)\,. \tag{2.29}$$

*L3. There exists a $d \times d$ matrix-valued function $G(s, x)$ such that*

$$\lim_{s \to t} \frac{1}{t-s} \int_{|y-x|\leq\epsilon} (y-x)(y-x)' P(s, x, t, dy) = G(s, x)\,. \tag{2.30}$$

The functions $f$ and $G$ are called the coefficients of the diffusion process. In particular, $f$ is called the drift vector and the $G$ is called the diffusion matrix. $G(s, x)$ is symmetric and non-negative-defined.

A common example of a diffusion process is the Brownian motion. Let $X_t$ denote the coordinate of a sufficiently small particle suspended in a liquid at the instant $t$. Neglecting the inertia of the particle, we may assume that the displacement of the particle has two components, the average displacement caused by the macroscopic velocity of the motion of the liquid and the fluctuation of the displacement caused by the chaotic nature of the thermal motion of the molecules.

Suppose that the velocity of the macroscopic motion of the liquid is given at point $x$ in the instant $t$ by $a(t, x)$. Let us assume that the fluctuation component of the displacement is a random variable whose distribution depends only on the position $x$ of the particle, the instant $t$ at which the displacement occurred and the quantity $\Delta t = t - s$ with $s \leq t$ which is the length of the interval of time during which the displacement occurred.

We assume that the average of the displacement is zero, independently of $t$, $x$ and $\Delta t$.Thus, the equation for the displacement of the particle is

$$X_{t+\Delta t} - X_t = a(t, X_t)\Delta t + \gamma(t, X_t, \Delta t)\,, \tag{2.31}$$

where $\langle \gamma(t, X_t, \Delta t) \rangle = 0$.

Now, if we assume that the properties of the medium only change slightly for small changes of $t$ and $x$. The process is then said to be homogeneous. Therefore, we may assume that

$$\gamma(t, X_t, \Delta t) = \sigma(t, X_t)\gamma(\Delta t)\,, \tag{2.32}$$

where $\sigma(t, x)$ characterizes the properties of the medium at the point $x$ and instant $t$ and $\gamma(\Delta t)$ is the value of the increment that is obtained in the homogeneous case under the condition that $\sigma(\Delta t) = 1$. So $\gamma(\Delta t)$ must be distributed like the increment of the Brownian process, $W_{t+\Delta t} - W_t$.

By doing this we can write the approximate formula

$$X_{t+\Delta t} - X_t \approx a(t, X_t)\Delta t + \sigma(t, X_t)(W_{t+\Delta t} - W_t)\,. \tag{2.33}$$

We shall now replace increments with the differentials $dt$ and $dW_t$ and obtain the following equality

$$dX_t = a(t, X_t)dt + \sigma(t, X_t)dW_t\,. \tag{2.34}$$

We call this equation a Langevin equation, which will be the starting point to determine the diffusion process. The solution of the Langevin equation (2.34) exists and is unique:

**Theorem 2.4.2** (**Existence and uniqueness of solution).** *Suppose that we have a stochastic differential equation with the form*

$$dX_t = f(t, X_t)dt + G(t, X_t)dW_t\,, \tag{2.35}$$

*with $X_{t_0} = x_0$, $t \in [t_0, T]$ and $T < \infty$, where $W_t$ is an Wiener process with values in $\mathbb{R}^m$ and $x_0$ is a random variable independent of $W_t - W_{t_0}$ for $t \geq t_0$. Suppose that $\mathbb{R}^d$-value function $f(t, x)$ and the $d \times m$-value matrix $G(t, x)$ are defined and measurable on $[t_0, T] \times \mathbb{R}^d$ and have the following properties:*

*There exists a constant $K > 0$ such that:*

- *For all $t \in [t_0, T]$ and $x, y \in \mathbb{R}^d$ we have the Lipschitz condition*

$$|f(t, x) - f(t, y)| + |G(t, x) - G(t, y)| \leq K|x - y|\,. \tag{2.36}$$

- *For all $t \in [t_0, T]$ and $x \in \mathbb{R}^d$ we have a restriction on growth*

$$|f(t, x)|^2 + |G(t, x)|^2 \leq K^2(1 + |x|^2)\,. \tag{2.37}$$

*Then the Eq. (2.35) has on $[t_0, T]$ a unique $\mathbb{R}^d$-value solution $X_t$, continuous with probability one, that satisfies the initial condition $X_{t_0} = x_0$. The solution is unique in the sense that if $X_t$ and $Y_t$ are continuous solutions of (2.35) with the same initial value $x_0$, then*

$$P(\sup_{t_0 \leq t \leq T} |X_t - Y_t| > 0) = 0\,. \tag{2.38}$$

To relate Langevin equation with diffusion processes we have the following theorem.

**Theorem 2.4.3** (**Langevin Equation and diffusion processes**). *Suppose that the conditions of the existence and uniqueness theorem 2.4.2 are satisfied for the stochastic differential equation*

$$dX_t = f(t, X_t)dt + G(t, X_t)dW_t\,, \tag{2.39}$$

*with $X_{t_0} = x_0$, $t_0 \leq t \leq T$ and $X_t, f(t,x) \in \mathbb{R}^d$, $W_t$ belongs to $\mathbb{R}^m$ and $G(t,x)$ is a $d \times m$ matrix. If in addition, the functions $f$ and $G$ are continuous with respect to $t$, the solution $X_t$ is a $d$-dimensional diffusion process on $[t_0, T]$ with drift vector $f(t,x)$ and diffusion matrix $B(t,x) = G(t,x)G^{\mathfrak{t}}(t,x)$.*

If the drift $f(X_t, t) \equiv f(X_t)$ and diffusion $G(X_t, t) \equiv G(X_t)$, are independent of time, we have an autonomous stochastic differential equation. The existence and uniqueness of it's solution is given by the following corollary:

**Corollary 2.4.4.** *Consider the autonomous stochastic differential equation*

$$dX_t = f(X_t)dt + G(X_t)dW_t\,, \tag{2.40}$$

*where $X_{t_0} = x_0$ and $f$ and $G$ are continuously differentiable functions such that the following condition (Lipschitz condition) is satisfied: There exists a constant $K$ such that for all $x, y \in \mathfrak{R}^d$*

$$|f(x) - f(y)| + |G(x) - G(y)| \leq K|x - y|\,. \tag{2.41}$$

For every initial condition $x_0$ that is independent of the $m$-dimensional Wiener process $W_t - W_{t_0}$ for $t_0 \leq t$, this Eq. (2.40) has a unique continuous solution $X_t$ in the entire interval $[t_0, \infty)$. The following two theorems are also of great importance.

**Theorem 2.4.5** (**Langevin Equation and Markov process**). *If the equation (2.35) satisfies the conditions of the existence and uniqueness of solution then the solution $X_t$ of the equation, for arbitrary initial values, is a Markov process on the interval $[t_0, T]$ whose initial probability distribution at the instant $t_0$ is the distribution of $x_0$ and whose transition probabilities are given by:*

$$P(s, x, t, B) = P(X_t \in B | X_s = x) = P(X_t(s,x) \in B)\,. \tag{2.42}$$

**Theorem 2.4.6** (**Langevin Equation and homogeneous Markov process**). *Suppose that the conditions of the existence and uniqueness theorem 2.4.2 are satisfied for equation (2.35). If the coefficients $f(X_t, t) \equiv f(X_t)$ and $G(X_t, t) \equiv G(X_t)$ are independent of $t$ on the interval $[t_0, T]$, then the solution $X_t$ is, for arbitrary initial values $x_0$, a homogeneous Markov process with the (stationary) transition probabilities*

$$P(X_t \in B | X_{t_0} = x_0) = P(t - t_0, x_0, B) = P(X_t(t_0, x_0) \in B)\,, \tag{2.43}$$

*where $X_t(t_0, x_0)$ is the solution of equation (2.35) with the initial value $X_{t_0} = x_0$. In particular, the autonomous equation*

$$dX_t = f(X_t)dt + G(X_t)dW_t, t \geq t_0 \tag{2.44}$$

*is a homogeneous Markov process defined for all $t \geq t_0$.*

## 2.5 Stochastic integral

Throughout the study of stochastic dynamical systems, we often end up with a stochastic differential equation like:

$$\frac{dX_t}{dt} = f(t, X_t) + G(t, X_t)\xi_t \,, \tag{2.45}$$

where $\xi$ is white. We assumed that $f(t, X_t)$ and $X_t \in \mathbb{R}^d$, $G(t, X_t)$ is a $d \times m$-matrix and $\xi_t$ is a $m$-dimensional white noise. From the deterministic case we have the initial-value problem:

$$\frac{dx}{dt}(t) = w(t, x(t)) \,, \tag{2.46}$$

with $x(0) = x_0$ and $w(t, x(t))$ is a continuous function. The solution of this differential equation is equivalent to the solution of the integral equation:

$$x(t) = x_0 + \int_{t_0}^{t} w(s, x(s))ds \,. \tag{2.47}$$

We have the same for the stochastic differential equations. The equation (2.45) can be rewritten in its integral form

$$X_t = X_0 + \int_{t_0}^{t} f(s, X_s)ds + \int_{t_0}^{t} G(s, X_s)\xi_s ds \,, \tag{2.48}$$

where $X_0$ is an arbitrary random variable. Using the result from the last section, we can eliminate the white noise from this equation. The second integral can be rewritten as

$$\int_{t_0}^{t} G(s, X_s)\xi_s ds = \int_{t_0}^{t} G(s, X_s)dW_s \,. \tag{2.49}$$

Equation (2.48) takes the form

$$X_t = X_0 + \int_{t_0}^{t} f(s, X_s)ds + \int_{t_0}^{t} G(s, X_s)dW_s \,. \tag{2.50}$$

The first integral int this last equation can be understood as the well known Riemann integral. The problem resides in the second integral. Since almost all sample functions of $W_t$ are of unbounded variation [1], we cannot in general interpret this integral as a Riemann-Stieltjes integral. Till the end of this section we will only focus on the second integral.

The purpose is to define the stochastic integral

$$\int_{t_0}^{t} \sigma(s)dW_s \tag{2.51}$$

for arbitrary $t \geqslant t_0$ and all $\sigma \in M_2[t_0, t]$, where $M_2[t_0, t]$ is a linear space (note that in the last integral the $X_t$ dependence is omitted to simplify the notation). In order to do this we shall start by defining this integral to for the step functions in $M_2[t_0, t]$. Then we shall extend this definition to the all set $M_2[t_0, t]$. Now we introduce the notions of filtration and adapted process,

**Definition 2.5.1** (**Filtration).** *A filtration, define over a mensurable space* $(\Omega, \mathfrak{F})$ *is a family of sub-*$\sigma$*-algebras of* $\mathfrak{F}$*, notated as* $\mathfrak{F}_t$*, such that if* $0 \leq s \leq t$*,* $\mathfrak{F}_s \subset \mathfrak{F}_t$*.*

**Definition 2.5.2** (**Adapted process).** *A stochastic process* $X_t$*, with* $t \in I$*, is said to be adapted (or non-anticipating) to the filtration* $\mathfrak{F}_t$*, if for all* $t \in I$ *the random variable* $X_t$ *is* $\mathfrak{F}_t$*-mensurable.*

We define a step function as:

**Definition 2.5.3** (**Step function).** *A function* $\sigma(t)$ *is called a step function if there is a decomposition* $t_0 < t_1 < ... < t_n = t$ *such that* $\sigma(s) = \sigma(t_{i-1})$ *for all* $s \in [t_{i-1}, t_i)$*, where* $i = 1, ..., n$*.*

The step function $\sigma(t)$ must be non-anticipating relative to the sigma-algebra generated by $\{W_s, s \leq t\}$, for all $t$. Now we defined a stochastic integral for this step functions as a $\mathbb{R}^d$-valued random variable

$$\int_{t_0}^{t} \sigma(s) dW_s = \sum_{i=1}^{n} \sigma(t_{i-1}) \left( W_{t_i} - W_{t_{i-1}} \right) . \tag{2.52}$$

This last integral as the following properties:

- $$\int_{t_0}^{t} (a\sigma_1 + b\sigma_2) dW_s = a \int_{t_0}^{t} \sigma_1 dW_s + b \int_{t_0}^{t} \sigma_2 dW_s, \tag{2.53}$$
  where $a, b \in \mathbb{R}$ and $\sigma_1, \sigma_2 \in M_2[t_0, t]$.
- For $\mathbf{E}(|\sigma(s)|) < \infty$ for all $s \in [t_0, t]$ we have:
$$\mathbf{E}\left( \int_{t_0}^{t} \sigma(s) dW_s \right) = 0 . \tag{2.54}$$
- For $\mathbf{E}\left(|\sigma(s)|\right) < \infty$, where $s \in [t_0, t]$, we have the following property
$$\mathbf{E}\left( \left( \int_{t_0}^{t} \sigma(s) dW_s \right) \left( \int_{t_0}^{t} \sigma(s) dW_s \right)^{\mathbf{t}} \right) = \int_{t_0}^{t} \mathbf{E}\left( \sigma(s) \sigma(s)^{\mathbf{t}} \right) ds . \tag{2.55}$$
  In particular, when $\sigma = \sigma^{\mathbf{t}}$ we have
$$\mathbf{E}\left( \left| \int_{t_0}^{t} \sigma(s) dW_s \right|^2 \right) = \int_{t_0}^{t} \mathbf{E}\left( |\sigma(s)|^2 \right) ds . \tag{2.56}$$

Now we want to extend the definition of stochastic integral for arbitrary functions in $M_2[t_0, t]$. The following lemma states that the set of step functions is dense in $M_2[t_0, t]$.

**Definition 2.5.4** (**Function in** $M_2[t_0, t]$**).** $\sigma(s) \in M_2[t_0, t]$*, if there exists a series of step functions* $\sigma_n(s)$ *such that*

$$\overset{ac}{\lim_{n \to \infty}} \int_{t_0}^{t} |\sigma(s) - \sigma_n(s)|^2 ds = 0 . \tag{2.57}$$

If the last lemma holds for a function $\sigma \in M_2[t_0, t]$ and a sequence $\{\sigma_n\}$ of step functions, the following is also true

$$\overset{st}{\lim_{n\to\infty}} \int_{t_0}^{t} |\sigma(s) - \sigma_n(s)|^2 \, ds = 0 \,. \tag{2.58}$$

One can show that this last assertion implies stochastic convergence of the sequence of integrals

$$\int_{t_0}^{t} \sigma_n(s) dW_s \tag{2.59}$$

to a specific random variable. To this end we shall use the following estimate for the stochastic integral of step functions:

**Lemma 2.5.5.** *Let $\sigma(s) \in M_2[t_0, t]$ be a step function. Then, for all $N > 0$ and $c > 0$*

$$P\left[\left|\int_{t_0}^{t} \sigma(s) dW_s\right| > c\right] \leqq \frac{N}{c^2} + P\left[\int_{t_0}^{t} |\sigma(s)|^2 \, ds > N\right] . \tag{2.60}$$

**Lemma 2.5.6.** *Let $\sigma \in M_2[t_0, t]$ and $\sigma_n(t)$ a sequence of step functions for which*

$$\overset{st}{\lim_{n\to\infty}} \int_{t_0}^{t} |\sigma(s) - \sigma_n(s)|^2 \, ds = 0 \,. \tag{2.61}$$

*If we define*

$$\int_{t_0}^{t} \sigma_n(s) dW_s \tag{2.62}$$

*by equation (2.52), then*

$$\overset{st}{\lim_{n\to\infty}} \int_{t_0}^{t} \sigma_n(s) dW_s = I(\sigma) \,, \tag{2.63}$$

*where $I(\sigma)$ is a random variable that does not depend on the special choice of sequence $\{\sigma_n\}$*

*Proof.* Since

$$\int_{t_0}^{t} |\sigma_n(s) - \sigma_m(s)|^2 \, ds \leqq 2 \int_{t_0}^{t} |\sigma(s) - \sigma_n(s)|^2 \, ds + 2 \int_{t_0}^{t} |\sigma(s) - \sigma_m(s)|^2 \, ds \,, \tag{2.64}$$

it follows from the assumption that

$$\overset{st}{\lim} \int_{t_0}^{t} |\sigma_n(s) - \sigma_m(s)|^2 \, ds = 0 \tag{2.65}$$

as $n, m \to \infty$. This is the same as saying

$$\lim_{n,m\to\infty} P\left[\int_{t_0}^{t} |\sigma_n(s) - \sigma_m(s)|^2 \, ds > \epsilon\right] = 0 \,, \tag{2.66}$$

for all $\epsilon > 0$ By applying the lemma (2.5.5) to $\sigma_n(s) - \sigma_m(s)$ we get

$$\lim_{n,m\to\infty} \sup P\left[\left|\int_{t_0}^{t} \sigma_n(s)dW_s - \int_{t_0}^{t} \sigma_m(s)dW_s\right| > \delta\right] \leqq \frac{\epsilon}{\delta^2} + \\ + \lim_{n,m\to\infty} \sup P\left[\int_{t_0}^{t} |\sigma_n(s) - \sigma_m(s)|^2 \, ds > \epsilon\right] = \frac{\epsilon}{\delta^2}\,. \quad (2.67)$$

Since $\epsilon$ is an arbitrary positive number, we have

$$\lim_{n,m\to\infty} P\left[\left|\int_{t_0}^{t} \sigma_n(s)dW_s - \int_{t_0}^{t} \sigma_m(s)dW_s\right| > \delta\right] = 0\,. \quad (2.68)$$

Since every stochastic Cauchy sequence also converges stochastically, there exists a random variable $I(\sigma)$ such that

$$\int_{t_0}^{t} \sigma_n(s)dW_s \to^{st} I(\sigma(s))\,. \quad (2.69)$$

□

From lemma 2.5.6 we get the following definition

**Definition 2.5.7 (Stochastic integral).** *For every function $\sigma \in M_2[t_0, t]$ ($d \times m$-matrix valued) the stochastic integral of $\sigma$ with respect to the $m$-dimensional Wiener process $W_t$ defined over the interval $[t_0, t]$ is defined as the random variable $I(\sigma)$, which is almost certainly determined in the accordance with the lemma (2.5.6)*

$$\int_{t_0}^{t} \sigma(s)dW_s = \overset{st}{\lim_{n\to\infty}} \int_{t_0}^{t} \sigma_n(s)dW_s\,, \quad (2.70)$$

*where $\{\sigma_n\}$ is a sequence of step functions in $M_2[t_0, t]$ that approximates $\sigma$ in the sense of*

$$\overset{st}{\lim_{n\to\infty}} \int_{t_0}^{t} |\sigma(s) - \sigma_n(s)|^2 \, ds = 0\,. \quad (2.71)$$

In the extending definition of stochastic integral from step functions to arbitrary functions in $M_2[t_0, t]$, the most important properties are the following

- $$\int_{t_0}^{t} (a\sigma_1 + b\sigma_2)dW_s = a\int_{t_0}^{t} \sigma_1 dW_s + b\int_{t_0}^{t} \sigma_2 dW_s\,, \quad (2.72)$$

  where $a, b \in \mathbb{R}$ and $\sigma_1, \sigma_2 \in M_2[t_0, t]$

- For $N > 0$, $c > 0$ and $\sigma \in M_2[t_0, t]$

$$P\left[\left|\int_{t_0}^{t} \sigma(s)dW_s\right| > c\right] \leqq \frac{N}{c^2} + P\left[\int_{t_0}^{t} |\sigma|^2 \, ds > N\right]\,. \quad (2.73)$$

- The relationship

$$\overset{st}{\lim_{n\to\infty}} \int_{t_0}^{t} |\sigma(s) - \sigma_n(s)|^2 \, ds = 0 \tag{2.74}$$

implies

$$\overset{st}{\lim_{n\to\infty}} \int_{t_0}^{t} \sigma_n(s) dW_s = \int_{t_0}^{t} \sigma(s) dW_s \,, \tag{2.75}$$

where $\{\sigma_n\} \in M_2[t_0, t]$ does not need to be step functions

- If

$$\int_{t_0}^{t} \mathbf{E}(|\sigma(s)|^2) ds < \infty \,, \tag{2.76}$$

we then have

$$\mathbf{E}\left( \int_{t_0}^{t} \sigma dW_s \right) = 0 \tag{2.77}$$

and

$$\mathbf{E}\left( \int_{t_0}^{t} \sigma(s) dW_s \right) \left( \int_{t_0}^{t} \sigma(s) dW_s \right)^{\mathbf{t}} = \int_{t_0}^{t} \mathbf{E}(\sigma(s)\sigma(s)^{\mathbf{t}}) ds \,. \tag{2.78}$$

In particular we have

$$\mathbf{E}\left( \left| \int_{t_0}^{t} \sigma(s) dW_s \right|^2 \right) = \int_{t_0}^{t} \mathbf{E}\left( |\sigma(s)|^2 \right) ds \,. \tag{2.79}$$

The relationship

$$X_t = \int_{t_0}^{t} \sigma(s) dW_s \tag{2.80}$$

can also be written as

$$dX_t = \sigma(t) dW_t \tag{2.81}$$

In fact stochastic differentials are simply a more compact symbolic notation for relationships of the form (2.48).

**Definition 2.5.8 (Stochastic differential).** *A stochastic process $X_t$ defined by equation*

$$X_t = X_s + \int_{s}^{t} f(u) du + \int_{s}^{t} G(u) dW_u \tag{2.82}$$

*possesses the stochastic differential*

$$dX_t = f(t) dt + G(t) dW_t \tag{2.83}$$

Finally, we introduce a very important theorem in stochastic analysis

**Theorem 2.5.9 (Itô formula).** *Let $v(t, X_t)$ denote a continuous function defined as $[t_0, T] \times \mathbb{R}^d \to \mathbb{R}^k$. With continuous partial derivatives, $\frac{\partial}{\partial t} v(t, X_t)$, $\frac{\partial}{\partial x} v(t, X_t)$ and $\frac{\partial^2}{\partial x^2} v(t, X_t)$.*

*If the d-dimensional stochastic process $X_t$ is defined on $[t_0, T]$ by the stochastic equation (2.83) then, for the k-dimensional process*

$$Y_t = v(t, X_t)\,, \tag{2.84}$$

*defined on the interval $[t_0, T]$ with initial value $Y_{t_0} = v(0, X_{t_0})$ we have*

$$v(T, X_T) = v(t_0, X_{t_0}) + \int_{t_0}^{T} \frac{\partial}{\partial t} v(s, X_s) ds \int_{t_0}^{T} \frac{\partial}{\partial x} v(s, X_s) dX_s + \\ + \frac{1}{2} \int_{t_0}^{T} \frac{\partial^2}{\partial x^2} v(s, X_s) G(s) G^{\mathbf{t}}(s) ds \tag{2.85}$$

## 2.6 The Fokker-Planck Equation

The important property of diffusion processes for our proposes is that their transition probability $P(s, x, t, B)$ is, under certain assumptions, uniquely determined by the drift and diffusion coefficients $f(x, t)$ and $G(x, t)$ respectively. If these coefficients are such that the Cauchy problem for the equation

$$-\frac{\partial u}{\partial t} = f(x, t)\frac{\partial u}{\partial x} + \frac{1}{2} G(x, t)\frac{\partial^2 u}{\partial x^2}\,, \tag{2.86}$$

with $u(s, x) = \phi(x)$, have a unique solution, in the region $x \in \mathbb{R}$ with $s \in (0, t)$ for every $t \in [0, T]$, for all $\phi(x)$ belonging to some class of functions that is everywhere-dense with respect to the metric of uniform convergence in the space of all continuous functions.

**Theorem 2.6.1** (**Backward Kolmogorov Equation**). *Let $X_t$, for $t_0 \leq t \leq T$, denote a d-dimensional diffusion with continuous coefficients $f(x, t)$ and $G(x, t)$. The limit relations in definition 2.2.2 hold uniformly. Let $\phi(x)$ denote a continuous bounded function such that the function $u(x, t) = \int \phi(y) P(t, x, s, dy)$ has bounded continuous first and second derivative with respect to $x$. Then u(x,t) has a derivative $\frac{\partial u}{\partial t}$ in the region $t \in (0, s)$, $x \in \mathbb{R}$, which satisfies the equation*

$$-\frac{\partial u}{\partial t} = f(x, t)\frac{\partial u}{\partial x} + \frac{1}{2} G(x, t)\frac{\partial^2 u}{\partial x^2} \tag{2.87}$$

*and the boundary condition $\lim_{s \to t} u(x, t) = \phi(x)$.*

*The Eq. (2.87) is called the backward Kolmogorov equation (or the first Kolmogorov equation).*

**Theorem 2.6.2** (**Density and backward Kolmogorov Equation**). *Suppose that the assumptions of theorem 2.6.1 regarding $X_t$ hold. If $P(s, x, t, .)$ has a density $p(s, x, t, y)$ that is continuous with respect to $s$ and the derivatives $\frac{\partial p}{\partial x}$ and $\frac{\partial^2 p}{\partial x^2}$ exist and are continuous with respect to $s$. Then $p$ is the fundamental solution of the equation of the backward equation*

$$-\frac{\partial p}{\partial s} = f(x, t)\frac{\partial p}{\partial x} + \frac{1}{2} G(x, t)\frac{\partial^2 p}{\partial x^2}\,, \tag{2.88}$$

*which satisfies the end condition*

$$\lim_{s \to t} p(s, x, t, y) = \delta(x - y)\,, \tag{2.89}$$

*where $\delta$ is Dirac's delta function.*

For example, the transition probability of the $\mathbb{R}$-valued Wiener process which has $f(t,x) = 0$ and $G(t,x) = 1$, is given by

$$p(s,x,t,y) = (2\pi(t-s))^{-\frac{d}{2}} \exp\left(-|y-x|^2/2(t-s)\right) . \tag{2.90}$$

This transition probability is, for fixed $t$ and $y$, a fundamental solution of the backward equation

$$\frac{\partial p}{\partial s} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2} . \tag{2.91}$$

If $X_t$ is a homogeneous process, then the coefficients $f(s,x) \equiv f(x)$ and $B(s,x) \equiv B(x)$ are independent of $s$. Since $P(s,x,t,B) = P(t-s,x,B)$, the sign of $\frac{\partial p}{\partial s}$ changes in the backward equation. For example, for the density $p(s,x,y)$ we have:

$$\frac{\partial p}{\partial s} = f(x)\frac{\partial p}{\partial x} + \frac{1}{2}G(x)\frac{\partial^2 p}{\partial x^2} . \tag{2.92}$$

**Theorem 2.6.3** (**Forward Kolmogorov Equation**). *Let $X_t$, for $t_0 \leq t \leq T$, denote a $d$-dimensional diffusion process for which the limit relationships of the definition of transition probability (2.2.2) hold uniformly in $s$ and $x$ and which possesses a transition density $p(s,x,t,y)$. If the derivatives $\frac{\partial p}{\partial t}$, $\frac{\partial f(t,y)p}{\partial y}$ and $\frac{\partial^2 G(t,y)p}{\partial y^2}$ exit and are continuous functions, then, for fixed $s$ and $x$ such that $s \leq t$, the transition density $p(s,x,t,y)$ is a fundamental solution of Kolmogorov's forward equation (or Komogorov's second equation) or also known as the Fokker-Planck equation.*

$$-\frac{\partial p}{\partial s} = f(x,t)\frac{\partial p}{\partial x} - \frac{1}{2}G(x,t)\frac{\partial^2 p}{\partial x^2} . \tag{2.93}$$

In the next section we shall focus our attention in a special case of stochastic differential equations, namely the ones in the form

$$\frac{dX_t}{dt} = f\left(t, X_t\right) + G\left(t, X_t\right)\Gamma(t) , \tag{2.94}$$

where $\Gamma(t)$ is a Gaussian white noise and the functions $f(t,X_t)$ and $G(t,X_t)$ are in general non-linear foundation of the state $X_t$ of the system.

## 2.7 Linear Stochastic Differential Equations

A much more complete theory can be developed when the coefficients functions $f(t,x)$ and $G(t,x)$ are linear on $x$, especially when $G(t,x)$ is independent of $x$. We start with the definition of linear stochastic differential equation.

**Definition 2.7.1** (**Linear stochastic differential equation**). *A stochastic differential equation for the $d$-dimensional process $X_t$ on the interval $[t_0,T]$*

$$\frac{dX_t}{dt} = f\left(t, X_t\right) + G\left(t, X_t\right)\Gamma(t) \tag{2.95}$$

*is said to be linear if the functions $f(t,X_t)$ and $G(t,X_t)$ are linear functions of $x \in \mathbb{R}^d$ on $[t_0,T] \times \mathbb{R}^d$.*

Explicitly, drift and diffusion coefficients have the form

$$f(t,x) = A(t)x + a(t)\,, \tag{2.96}$$

where $A(t)$ is a $d \times d$-matrix and $a(t)$ is a $\mathbb{R}^d$ and

$$G(t, X_t) = (B_1(t)x + b_1(t), ..., B_m(t)x + b_m(t))\,, \tag{2.97}$$

where $B_k(t)$ is an $d \times d$-matrix and $b_k(t) \in \mathbb{R}^d$. Thus, a linear differential stochastic equation has the form:

$$dX_t = (A(t)X_t + a(t))dt + \sum_{i=1}^{m}(B_i(t)X_t + b_i(t))dW_t{}^i\,, \tag{2.98}$$

where $W_t = (W_t^1, ..., W_t^m)$.

Equation (2.98) is said to be homogeneous if $a(t) = b_1(t) = ... = b_m(t) = 0$ and is said to be linear in the narrow sense if $B_1(t) = ... = B_m(t) \equiv 0$. The unique continuous solution is guaranteed through:

**Theorem 2.7.2** (**Existence and uniqueness of solution of linear stochastic differential equation).** *The linear stochastic equation (2.98), for every initial value $X_{t_0} = x_0$ that is independent of $W_t - W_{t_0}$ (with $t \geq t_0$), has an unique continuous solution throughout the interval $[t_0, T]$ provided only the functions $A(t), a(t), B_i(t)$ and $b_i(t)$ are measurable and bounded. If the assumption holds in every subinterval of $[t_0, \infty)$, there exist a unique global solution.*

**Corollary 2.7.3.** *A global solution always exit for the autonomous linear differential equation*

$$dX_t = (AX_t + a)dt + \sum_{i=1}^{m}(B_iX_t + b_i)dW_t{}^i\,, \tag{2.99}$$

*with $X_0 = x_0$ and $A, a, B_i$ and $b_i$ independent of $t$.*

Let us now consider the stochastic linear equations in the narrow sense, i.e. with $B_1(t) = ... = B_m(t) \equiv 0$.

**Theorem 2.7.4** (**Solution of SDE in narrow sense).** *The linear stochastic differential equation*

$$dX_t = (A(t)X_t + a(t))dt + b(t)dW_t\,, \tag{2.100}$$

*with $X_0 = x_0$, has the solution*

$$X_t = \Phi(t)\left(x_0 + \int_{t_0}^{t}\Phi(s)^{-1}a(s)ds + \int_{t_0}^{t}\Phi(s)^{-1}B(s)dW_s\right) \tag{2.101}$$

*on $[t_0, T]$, where $\Phi(t)$ is the fundamental matrix of the deterministic equation*

$$\frac{dX_t}{dt} = A(t)X_t\,. \tag{2.102}$$

**Corollary 2.7.5.** *If the matrix $A(t) \equiv A$ in equation (2.100) is independent of $t$, then*

$$X_t = x_0 e^{A(t-t_0)} + \int_{t_0}^{t} \int_{t_0}^{t} e^{A(t-s)}(a(s)ds + b(s)dW_t) \tag{2.103}$$

*is a solution.*

**Theorem 2.7.6 (Stationary Gaussian process).** *The solution of the equation (2.100) with $X_{t_0} = x_0$ is a stationary Gaussian process if $A(t) \equiv A, a(t) \equiv 0, b(t) \equiv b$, the eigenvalues of $A$ have negative real parts and $x_0$ is $\mathfrak{N}(0, K)$-distributed, where $K$ is the solution*

$$K = \int_0^\infty e^{At} B B^{\mathbf{t}} e^{A^{\mathbf{t}} t} dt \tag{2.104}$$

*of the equation*

$$AK + KA^{\mathbf{t}} = -BB^{\mathbf{t}}. \tag{2.105}$$

*In that case, for the process $X_t$*

$$\mathbf{E}(X_t) = 0 \tag{2.106}$$

*and*

$$\mathbf{E}(X_s X_t) = \begin{cases} e^{A(s-t)}K & \text{for } s \geqslant t \geqslant t_0 \\ K e^{A^{\mathbf{t}}(s-t)} & \text{for } t \geqslant s \geqslant t_0\,. \end{cases} \tag{2.107}$$

As an example, we take the stochastic equation for the Brownian motion of a particle under the influence of friction but no other force field which yields the Langevin equation

$$\frac{dX_t}{dt} = -\alpha X_t + \beta \Gamma(t)\,, \tag{2.108}$$

where $\alpha > 0$ and $\beta$ are constants. In the context of the Brownian motion of a particle, $X_t$ is one of the three scalar velocity components of the particle and $\Gamma(t)$ is scalar white noise. The correspondent stochastic differential equation is given by:

$$dX_t = -\alpha X_t dt + \sigma dW_t\,, \tag{2.109}$$

with $X_0 = 0$. This is a linear and autonomous stochastic differential equation and therefore, according to the Corollary (2.7.5) the solution reads

$$X_t = e^{-\alpha t} c + \sigma \int_0^t e^{-\alpha(t-s)} dW_s\,. \tag{2.110}$$

## 2.8 Langevin approach

The study of the behaviour of complex systems, such the ones described by stochastic time series, must be based on the assessment of the non-linear interactions and the strength of fluctuating forces, which leads to the problem of retrieving a stochastic dynamical system from the data. We address the problem of how to reconstruct stochastic evolution equations from the data in terms of the Langevin equation or the corresponding Fokker-Planck equation.

For Markovian stochastic processes, the time evolution of the associated probability density function is given by the Kramer-Moyal expansion:

$$\frac{\partial P}{\partial t}(x,t) = \sum_{k\geq 0} \left(-\frac{\partial}{\partial x}\right)^k \left[D_k(x)P(x,t)\right] , \tag{2.111}$$

with coefficients $D_k(x)$ given by:

$$D_k(x) = \lim_{\tau\to 0} \frac{1}{\tau} M_k(x,\tau) \tag{2.112}$$

and

$$M_k(x,\tau) = \frac{1}{k!} \left\langle (X_{t+\tau} - X_t)^2 \right\rangle |_{X_t=x} . \tag{2.113}$$

For diffusion processes Eq. (2.111) reduces to the Fokker-Planck equation

$$\frac{\partial}{\partial t} P(x,t) = -\frac{\partial}{\partial x} \left[D_1(x)P(x,t)\right] + \frac{\partial^2}{\partial x^2} \left[D_2(x)P(x,t)\ \right] . \tag{2.114}$$

Therefore the processes governed by the Itô-Langevin equation Eq. (5.1) must have $D_k(x) = 0$ for $k \geq 3$. One way to guarantee that all these coefficients are null for $k \geq 3$ is through the Pawula theorem:

**Theorem 2.8.1 (Pawula theorem).** *Let be $P(s,x,t,B)$ a positive transition probability with $0 \leq s \leq t \leq T$, $B \in \mathfrak{B}^d$ and $x \in \mathbb{R}^d$.*

*Then, if in the Kramers-Moyal expansion any coefficient $D_{2r}(x,t) = 0$ with $r \geq 1$ then all coefficients $D_n$ with $n \geq 3$ must vanish.*

*Proof.* In order to derive the Pawula theorem we will need the generalized Schwartz inequality [15]:

$$\left(\int f(x')g(x')P(x')dx'\right)^2 \leq \int f^2(x')P(x')dx' \int g^2(x')P(x')dx' , \tag{2.115}$$

in which $P$ is a non-negative function and $f$ and $g$ are arbitrary functions. Now we make the following choices for the functions $P$, $f$ and $g$:

$$f(x') = (x'-x)^n , \tag{2.116}$$

$$g(x') = (x'-x)^{n+m} \tag{2.117}$$

and

$$P(x') = P(x', t+\tau|x,t) , \tag{2.118}$$

for $n$ and $m \geq 0$. Now using [15]

$$M_k(x,\tau) = \frac{1}{k!} \left\langle (X_{t+\tau} - X_t)^2 \right\rangle |_{X_t=x} = \int (x'-x)^k P(x', t+\tau|x,t)dx' \tag{2.119}$$

in Eq. (2.115) we obtain the following inequality:

$$M_{2n+m}^2 \leq M_{2n}.M_{2n+2m} . \tag{2.120}$$

When $m = 0$ we obtain the relation:

$$M_{2n}^2 \leq M_{2n} \,. \tag{2.121}$$

which is fulfilled for every $n$. If we consider $n \geq 1$ and $m \geq 1$ and using Eq. (2.112) and Eq. (2.113) we obtain from Eq. (2.120):

$$((2n+m)! D_{2n+m})^2 \leq (2n)!(2n+m)! D_{2n} D_{2n+2m} \,. \tag{2.122}$$

If $D_{2n}$ is zero, then $D_{2n+m}$ must be also zero, i.e.,

$$D_{2n} = 0 \Rightarrow D_{2n+1} = D_{2n+2} = ... = D_{2n+m} = 0 \,. \tag{2.123}$$

Furthermore, if $D_{2n+2m} = 0$ then $D_{2n+m}$, must bee zero too, i.e.,

$$D_{2r} = 0 \Rightarrow D_{r+n} = 0 \tag{2.124}$$

for $(n = 1, ..., r-1)$, i.e.,

$$D_{2r-1} = ... = D_{r+1} = 0 \tag{2.125}$$

for $r \geq 2$. Form Eq. (2.123) and the repeated use of Eq. (2.125), we conclude that if any $D_{2r} = 0$ for $r \geq 1$ all coefficients $D_n$ for $n \geq 3$ must vanish. □

The coefficient $D_4$ is then the key coefficient to be investigated in order to establish the validity of the modelling of the data series by the Itô-Langevin and the associated Fokker-Planck equations. This method can bee viewed as an extension of the multifractal description of stochastic processes[8]. Details on this method can be found in Refs. [9, 6, 7, 12, 3].

# Part II

# Contributions

# Chapter 3

# Data processing

In this Chapter we describe how we extracted and process the empirical data and motivate the study of volume-price distributions.

All the data were collected from the website `http://finance.yahoo.com/` with a sample rate of 0.1 mins starting in January 27th 2011 and ending in April 6th 2014, a total of 976 days ($\sim 10^5$ data points). However, after filtering out all the weekends, holidays, after-hours and nights we end up with $\sim 25500$ data points. Spurious events are also removed, such as occur due to the inevitable recording errors. Each register refers to one specific enterprise and is composed by the following fields: company name, last trade price, volume, day's highest price, day's lowest price, last trade date, 200-moving average and average daily volume. A total of $N_e \sim 2000$ companies are listed for each time-span of 10 minutes. Since we don't have access to the instantaneous trading price of each transaction made, we have to consider the last trading price as the best representative of the price change on each set of ten-minutes trading volume.

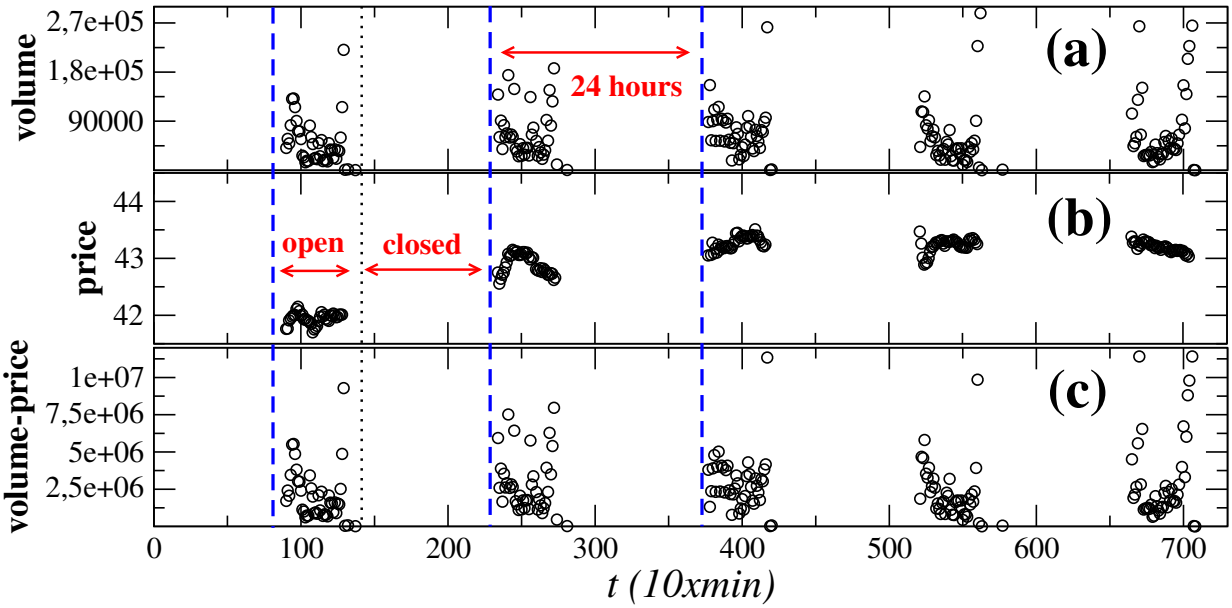


Figure 3.1: Illustration of the volume and price evolution for one company during four days: **(a)** volume $V$, **(b)** price $p$ and **(c)** volume-price $pV$ time-series.

Two important variables of stock market data are the trading volume and the stocks prices. Figure 3.1a and 3.1b show the evolution of the trading volume $V$ and the last trade price $p$ respectively for one single company, as well as their product 3.1c, so-called volume-price $s = pV$ , during approximately five working days (one week). In Fig. 3.1b one observes price changes that during a period of $\sim 6.5$ hours, the period corresponding to the open time of NYSM, generally from 9:30 a.m. to 4:00 p.m. ET. Furthermore, during two hours after 4:00 p.m. the trading volume fluctuates abruptly, which reflects the so-called after-hours trading, illustrated Fig. 3.2(**I**). Typically, after-hours trading occurs from 4:00 to 8:00 p.m. ET. In these periods the dynamics seems to be very different from the normal trading period. See Fig. 3.2(II). Still, changes in capitalization during these after-hours periods can be neglected. In the following, we will only consider the study of the normal trading period.

The volume and the price interact with each other. For instance volume can play an important role on the stocks prices [20], as illustrated in Fig. 3.3a. It appears that high values of volume triggers high prices and small volumes triggers low prices. Large volumes indicate high liquidity of the market. It is important to investigate the relationship between the two variables since both are the products of the same market mechanisms. The discussion of one of these variables cannot be complete without incorporating the other one. Since prices and volumes are recorded simultaneously and are the result of the same trading activities instead of considering the volume $V$ and the price $p$, we consider the volume-price solely.

While the price and volume distribution are useful for portfolio purposes, the distribution of volume-prices provides information about the entire capital traded in the market.

Figure 3.3b shows the autocorrelation function (ACF) of the price $p$, volume $V$ and volume-price $s$. The ACF is given by the following expression:

$$C_\tau = \frac{1}{N\sigma^2} \sum_{t=1}^{N-\tau} (x_t - \langle x \rangle)(x_{t+\tau} - \langle x \rangle)\,, \tag{3.1}$$

where $\sigma^2$ is the variance of $x$, $\langle x \rangle$ is the average of $x$, $N$ is the total number of data points $x_i$ and $\tau$ is the time-lag.

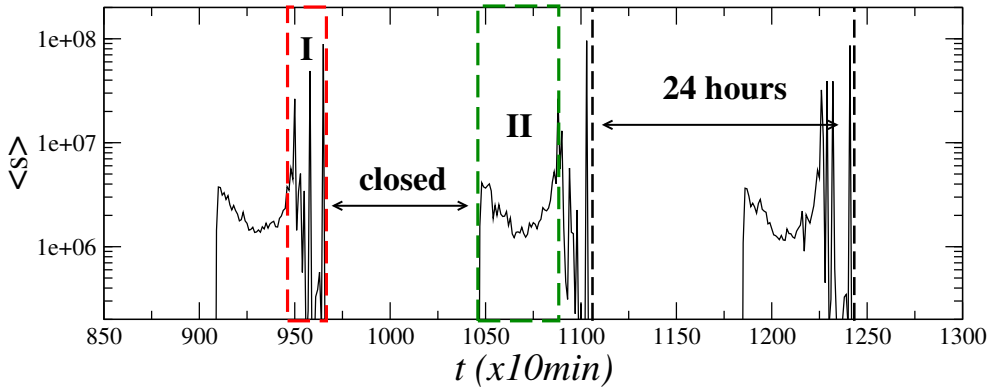


Figure 3.2: Here are represented three days of trading in the NYSM. On the y-axis we have the mean $\langle s \rangle$ at each ten minutes window and and on the x-axis we have the time $t$ in units of ten minutes. The red region **I** corresponds to the after-hours trading period. The green region **II** corresponds to the normal period of trading (see text).

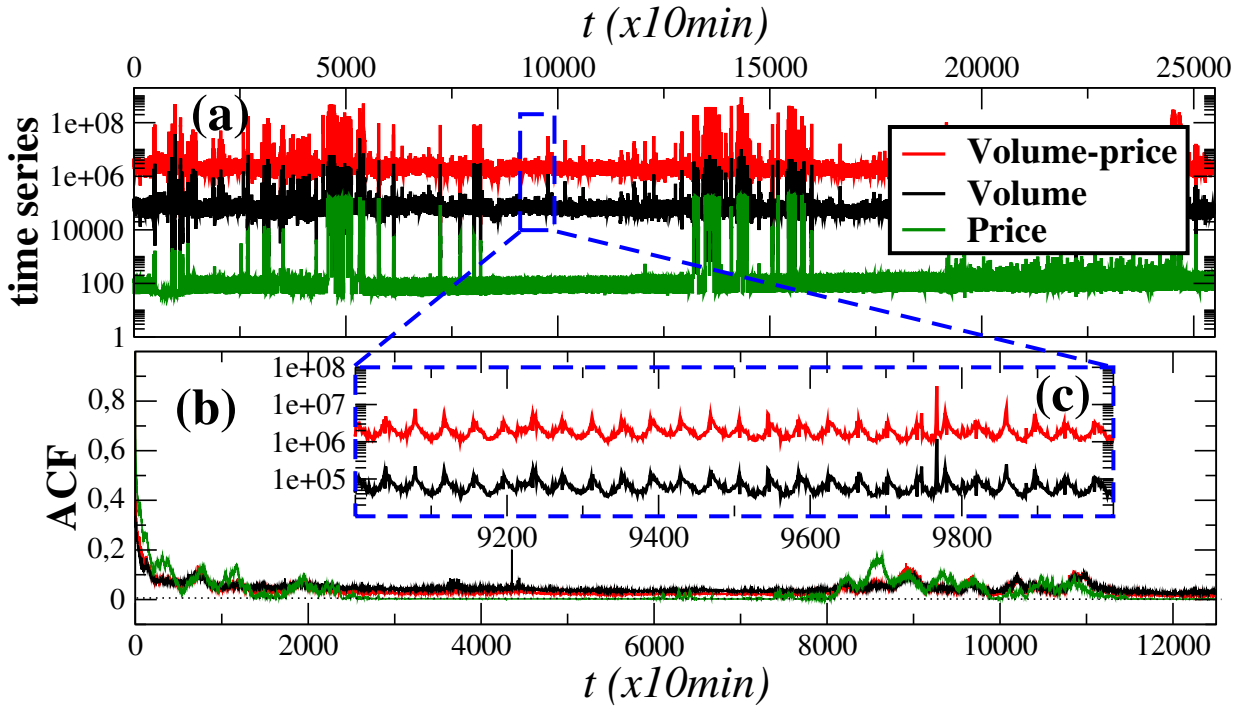


Figure 3.3: **(a)** Time series of ten minutes average of: price, volume and volume-price. Here each point corresponds to ten minutes window and all the weekends, holidays, after-hours and nights were filtered out from the data. **(b)** Autocorrelation function (ACF) of price (in green), volume (in black), volume-price (in red). **(c)** Zoom in on volume and volume-price time series.

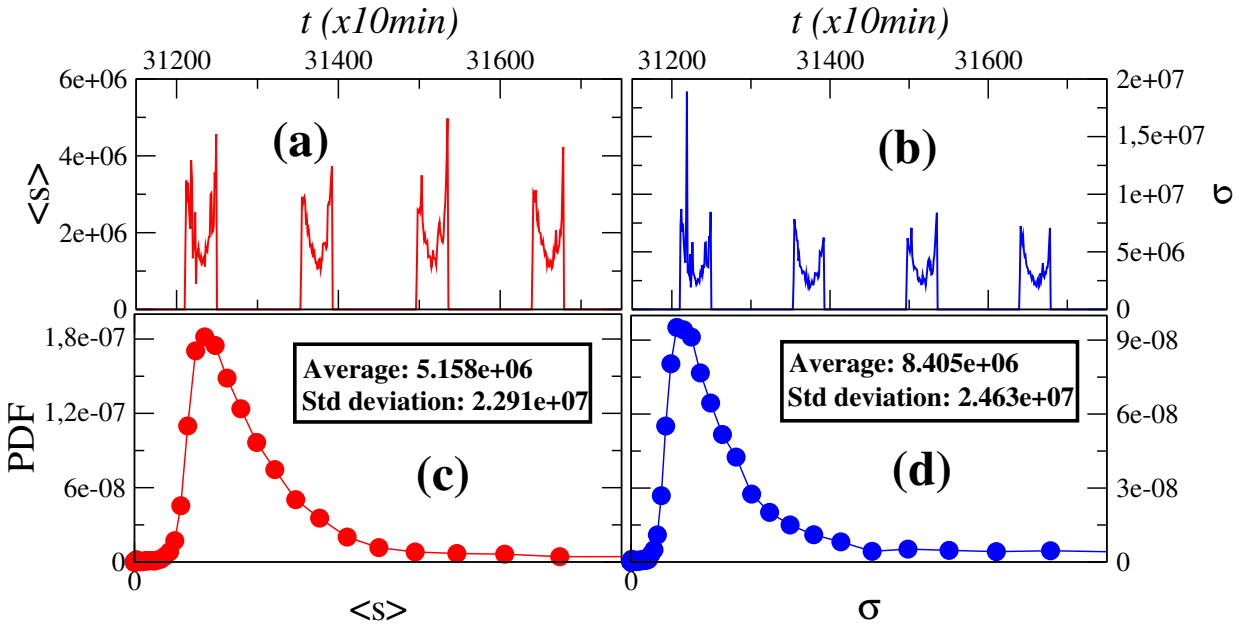


Figure 3.4: To characterize the evolution of the density functions one first considers the time series of **(a)** the empirical volume-price average $\langle s \rangle$ and of **(b)** the corresponding standard deviation $\sigma$.

We see that the ACF for the volume and the volume-price does not decay monotonically to zero, suggesting some kind of periodicity. A zoom in of a section of these two series, as in Fig. 3.3c, clearly shows oscillations with a period of one day, comprehending 39 data points.

Apparently, the series of volume-price inherit the oscillation-like structure from the series of trading volumes. Both have a pick at the beginning of each day and another pick at the closing time. The price time series does not show these oscillations. The correlation between the volume and volume-price is $\sim 0.8$.

For each 10-minute interval we compute the cumulative density distribution (CDF) of the volume-price and record its respective average $\langle s \rangle$ over the listed companies, and standard deviation $\sigma$. In Fig. 3.4 we plot the time series of the means of volume-price $\langle s \rangle$, the standard deviation series of volume-price $\sigma$ and the respective PDF.

# Chapter 4

# Fitting and error analysis

In this chapter we study the evolution of volume-price distributions on NYSM. We fit the empirical data with each one the four models described in Eqs. (4.1)-(4.4), yielding one time series for each parameter $\phi$ and $\theta$ defining each model, which will be analysed posteriorly. Also, in this chapter, we make an error analysis to determined which of the four models is the best to describe the dynamics.

## 4.1 Four models for volume-price distributions

For each 10-minute interval we compute the cumulative density function (CDF) $F_{emp}(x)$ of the volume-price. In order to find a good fit to the empirical CDF we will consider four well-known biparametric distributions, namely the Gamma distribution, inverse Gamma distribution, log-normal and the Weibull distribution. We fit the empirical CDF data (bullets in Fig. 3.4b) with these four different models, which are used for finance data analysis[4]. For each of those the correspondent probability density functions are: $\Gamma-$distribution,

$$p_{\Gamma}(s) = \frac{s^{\phi_{\Gamma}-1}}{\theta_{\gamma}^{\phi_{\Gamma}}\Gamma[\phi_{\Gamma}]} exp\left[-\frac{s}{\theta_{\Gamma}}\right] , \tag{4.1}$$

the inverse Gamma

$$p_{1/\Gamma}(s) = \frac{\theta_{1/\Gamma}^{\phi_{1/\Gamma}}}{\Gamma[\phi_{1/\Gamma}]} s^{-\phi_{1/\Gamma}-1} exp\left[-\frac{\theta_{1/\Gamma}}{s}\right] , \tag{4.2}$$

the log-normal

$$p_{\ln}(s) = \frac{1}{\sqrt{2\pi}\theta_{\ln} s} exp\left[-\frac{(\log s - \phi_{\ln})^2}{2\theta_{\ln}^2}\right] \tag{4.3}$$

and the Weibull

$$p_{W}(s) = \frac{\phi_{W}}{\theta_{W}^{\phi_{W}}} s^{\phi_{W}-1} exp\left[-\left(\frac{s}{\theta_{W}}\right)^{\phi_{W}}\right] . \tag{4.4}$$

The error of each parameter value, $\Delta\phi$ and $\Delta\theta$ when making the fit, using a least square scheme, is also taken into account. In the next two sections, we evaluate how accurate is a fitting by analysing the relatives error and the Kullback-Leibler divergence. We also introduce a new variate of this last one.

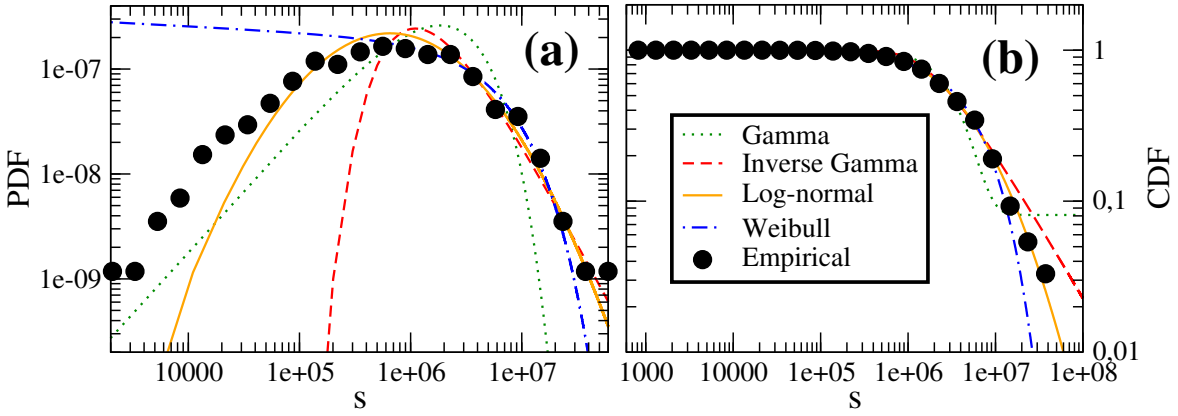


Figure 4.1: **(a)** Numerical probability density function fitted by the four different distributions: log-normal distribution $\Gamma-$distribution, inverse $\Gamma-$distribution, Weibull-distribution. **(b)** Numerical cumulative density function also fitted by the four different distributions.

Figure 4.1 shows the an example of a ten-minutes window of probability and cumulative density functions of volume-price. In Fig. 4.1 we have numerical PDF (bullets) being fitted by the four models using the same fitting parameters from the CDF. In this figure we can see that the Gamma distributions seems to deliver the worst fit. The Weibull seems to deliver a very poor fit for low values of $s$, however for the high values the fit is good. The inverse Gamma also shows a very poor fit for low values of $s$, but for high values we have a very good match. Finally, the log-normal distribution delivers a very good fit for mid-low and high values of $s$.

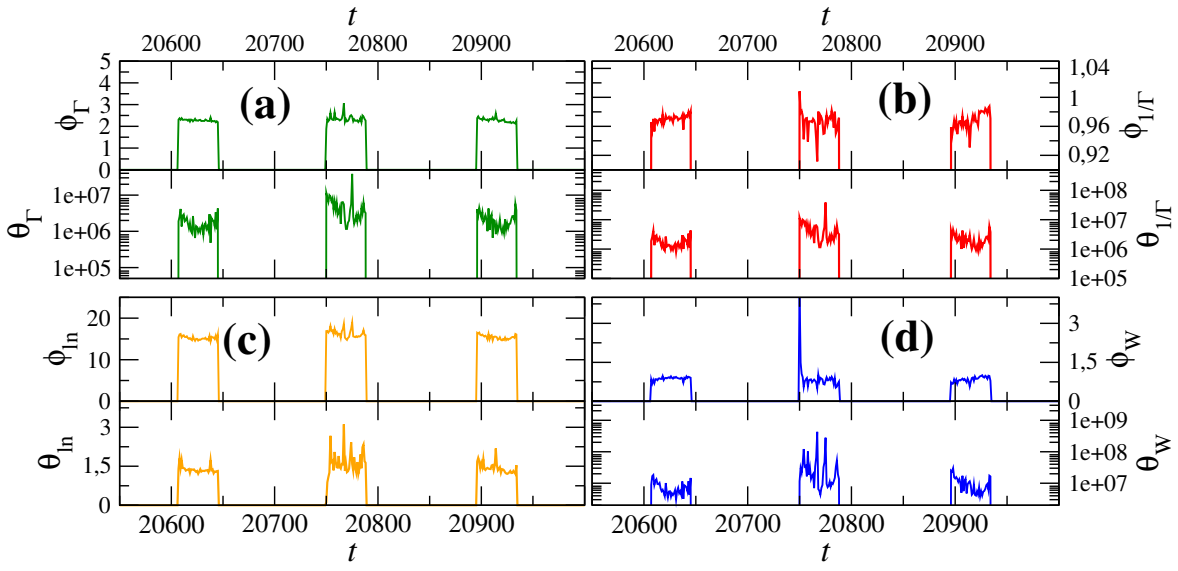


Figure 4.2: Time series of the two parameters characterizing the evolution of the cumulative density function (CDF) of the volume-price $s$: **(a)** $\Gamma$-distribution **(b)** inverse $\Gamma$-distribution, **(c)** log-normal distribution and **(d)** Weibull distribution. Each point in these time series correspond to 10-minute intervals. Periods with no activity correspond to the period where market is closed, and therefore will not be considered in our approach. In all plots, different colours correspond to different distributions.

## 4.2 Relative deviations

In order to evaluate how accurate a model is, we first consider the relative error, $\Delta\phi$ and $\Delta\theta$, of each parameter value, $\phi$ and $\theta$ respectively. Figure 4.3a and 4.3b show the PDF for the observed relative errors of $\phi$ and $\theta$ respectively. From these two plots it seems that each distribution fits quite well the empirical CDF data, since relatives errors are mostly under five percent, except the one for parameter $\theta$ in the Gamma distribution. See Tab. 4.1.

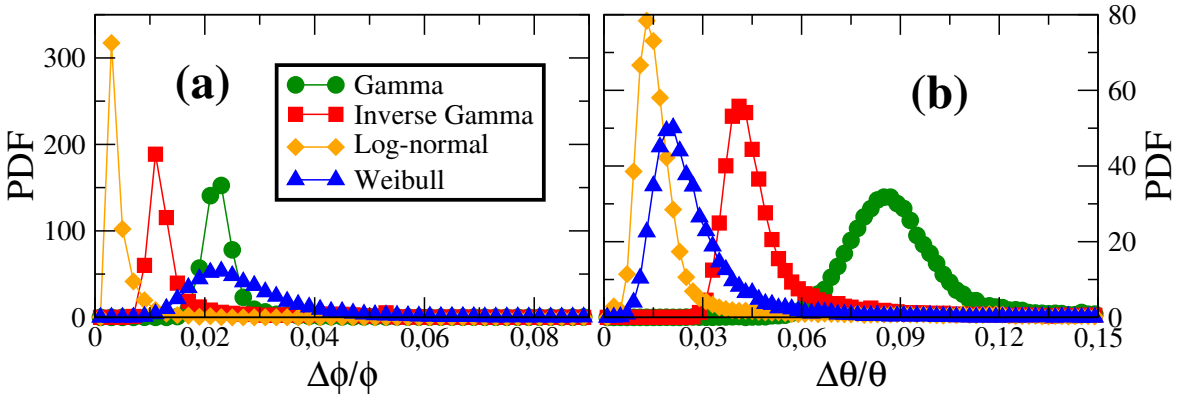


Figure 4.3: Probability density function of the resulting relative errors **(a)** $\Delta\phi$ and **(b)** and $\Delta\theta$, corresponding to the fitting parameters $\phi$ and $\theta$ respectively.

The log-normal has the smallest error average, around 0.12% for parameter $\phi$ and around 2.25% for $\theta$. The inverse Γ-distribution shows also acceptable deviations. Specially for the parameter $\phi$, with an average error of about 1.52%. Noting that parameter $\phi$ in inverse Γ-distribution controls the tail of the distribution for large values of volume-price. The other two models are not as good as the log-normal and inverse Γ-distributions.

| | Param. err. $\Delta\phi/\phi$ | | Param. err. $\Delta\theta/\theta$ | |
|---|---|---|---|---|
| | Average | Std Dev. | Average | Std Dev. |
| $\Gamma-$distribution | 2.75e-02 | 3.44e-02 | 1.01e-01 | 7.87e-02 |
| Inverse $\Gamma-$distribution | 1.52e-02 | 9.76e-03 | 6.34e-02 | 6.08e-02 |
| **Log-normal** | **1.25e-03** | **1.33e-03** | **2.25e-02** | **2.88e-02** |
| Weibull | 3.89e-02 | 6.89e-02 | 3.23e-02 | 3.95e-02 |

Table 4.1: The average and standard deviations of the value distributions for each parameter error, $\Delta\phi$ and $\Delta\theta$, in Fig. 4.1e-f. The best fits are obtained for the log-normal distribution and inverse Γ-distribution.

In a previous work [17], where the evolution of the mean volume-price $\langle s\rangle$ was considered separately and the models were used to fit the distribution of the normalized volume-price, $s/\langle s\rangle$, the optimal model according to relative deviations was only the inverse Γ-distribution.

## 4.3 A variant of the Kullback-Leibler divergence for tail distributions

The relative deviations do no take into account the observation frequency of each value of the volume-price. For that one needs to consider a weight given by the probability density function or another density function. To weight each value in the volume-price spectrum according to some density function we introduce here the *generalized* Kullback-Leibler divergence:

$$D^{(F)}(P||Q) = \sum_i \ln\left(\left|\frac{P(i)}{Q(i)}\right|\right) F(i)\Delta x\,, \tag{4.5}$$

where $Q(i)$ is the empirical distribution, $P(i)$ is the modelled PDF and $F(i)$ is a weighting function. For $F(i) = P(i)$ one obtains the standard Kullback-Leibler divergence[10], where the logarithmic deviations are heavier weighted in the central region of the distribution. Figure 4.4a shows the distribution of $D^{(P)}$ values obtained when considering each one of the four models. Once again one observes that the log-normal distribution is the one yielding smaller deviations.

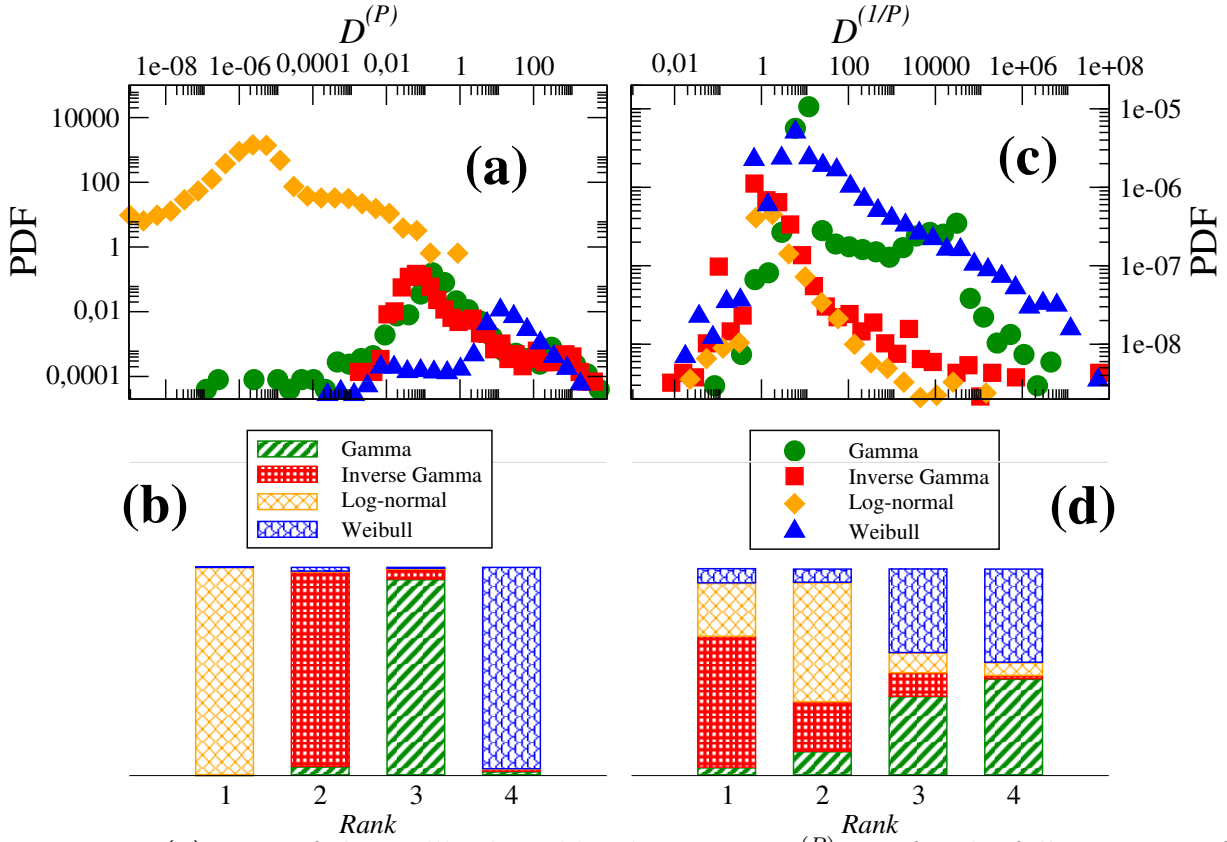


Figure 4.4: **(a)** PDF of the Kullback-Leibler divergence $D^{(P)}$ test for the full spectrum of the volume-price. **(b)** Percentage of accuracy rankings for each model, using the Kullback-Leibler divergence $D^{(P)}$. A model with rank 1 is more accurate then a model with rank 2. **(c)** PDF of the tail Kullback-Leibler divergence $D^{(1/P)}$ (see text) using only the values of $s$ larger the median of the distribution. **(d)** Percentage of accuracy rankings for each model, using the tail Kullback-Leibler divergence $D^{(1/P)}$.

Since the $D^{(P)}$ distributions overlap, one may argue that the log-normal distribution might not be *always* the best model during the three years concerned by our data. To address this question we plot in 4.4b the ranking ordering all four models in their accuracy for each time step. Almost always the log-normal is the best model, followed by the inverse $\Gamma$-distribution, meaning that it is the best model for the central region of the volume-price spectrum.

Volume-price values are not equally important, but the most important region of its spectrum is not the central region. It is the region of larger volume-price changes. For instance, a deviation from the observed distribution in the region of small volume-prices result in a smaller fluctuation of the amount of transactions than in the region of largest values, where the risk is the highest and therefore should be more accurately fitted. A different functions $F$ should therefore be taken in Eq. (4.5).

To weight the largest volume-prices we consider only the region of the distribution for $s$ larger than the median and then take $F(i) = 1/P(i)$, in Eq. (4.5), whenever $P(i) \neq 0$(taken $F(i)$ otherwise). In this way, the largest values of the volume-price, i.e. those for which $P(i)$ is the smallest will be weighted heavier than the others.

$$D^{(1/P)}(P||Q) = \sum_i \ln\left(\left|\frac{P(i)}{Q(i)}\right|\right)\frac{1}{P(i)}\Delta x\,. \tag{4.6}$$

Figures 4.4c and 4.4d show respectively the distance distributions of $D^{(1/P)}$ values and the corresponding rankings respectively.

| | $D^{(P)}$ | | $D^{(1/P)}$ | |
|---|---|---|---|---|
| | Average | Std Dev. | Average | Std Dev. |
| $\Gamma-$distribution | 3.90 | 5.53e+04 | 7.34e+05 | 1.90e+13 |
| Inverse $\Gamma-$distribution | 3.86 | 5.63e+04 | 4.34e+06 | 3.25e+14 |
| Log-normal | 2.27e-04 | 3.68e-05 | 1.12e+06 | 4.34e+15 |
| Weibull | 41.71 | 2.69e+05 | 6.45+05 | 6.10e+12 |

Table 4.2: Average and standard deviation for Kullback-Leibler divergence $D^{(P)}$ and for the new variant $D^{(1/P)}$.

Interestingly, not only the best model is the inverse $\Gamma$-distribution but the dominance of one single model in each rank is not strong as when considering the full distributions. This indicates that in NYSM the best model of the volume-price tail distribution is most probably the inverse $\Gamma$-distribution but the probability that another model, is the best one is significative.

# Chapter 5

# The stochastic evolution of non-stationary distributions

In this section we describe in some detail how to quantitatively describe the evolution of the parameter $\phi$ in the inverse-$\Gamma$ distribution. We will show a specific example that, while the volume-price $s$ may evolve according to a non-markovian process, the parameters characterizing the corresponding distribution of volume-prices are themselves Markovian. Acording to [8], we shall consider that the parameter $\phi$ characterizing the tail evolution in the inverse-$\Gamma$, is assumed to be governed by a deterministic part to which a Gaussian $\delta$-correlated white noise is added:

$$dX_t = D_1(X_t)dt + \sqrt{2D_2(X_t)}dW_t\,, \tag{5.1}$$

where $W_t$ is a Wiener process (see Chapter 2).

A complete analysis of experimental data, which is generated by the interplay of deterministic dynamics and dynamical noise, has to address the following issues:

- Identification of the order parameters.
- Extracting the deterministic dynamics.
- Evaluating the properties of the fluctuations.

## 5.1 The stochastic evolution of inverse-$\Gamma$ parameters

To explore the inverse-$\Gamma$ distribution model, we first consider the meaning of its two parameters. A closer look at Eq. (4.2) leads to the conclusion that while $\theta$ characterizes the shape of the distribution for the lowest range of volume-prices, the parameter $\phi$ characterizes the power law tail $\sim s^{-\phi-1}$. Since it is this tail that incorporates the largest fluctuations of volume-prices, in this section we focus on the evolution of parameter $\phi$ solely.

Figure 5.1a shows that the time series of parameter $\phi$ might present some kind of periodicity. In (b) we see that the time series of $\theta$ have the same kind of periodicity presented in volume and the volume-price time series. The correlation between $\theta$ and

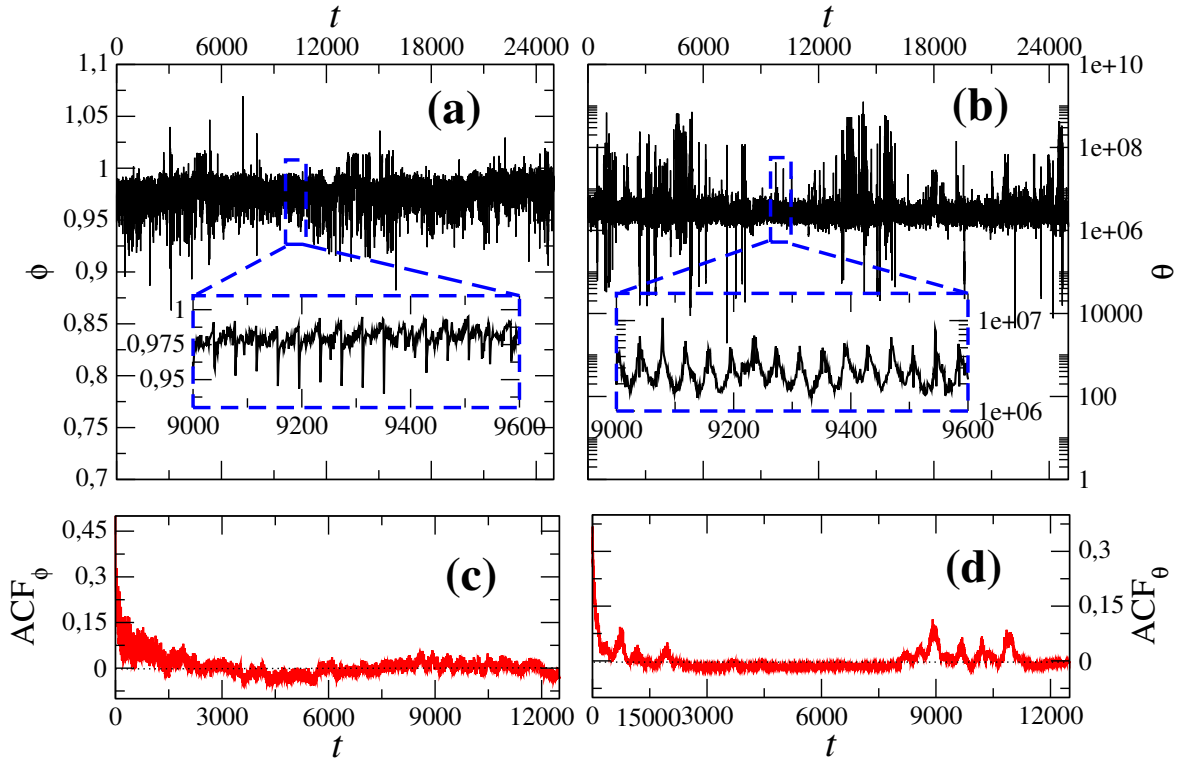


Figure 5.1: **(a)** Time series of parameter $\phi$ and **(b)** parameter $\theta$ of inverse-$\Gamma$ distribution. Zoom in of each time-series are shown inside the dashed boxes. Autocorrelation function of **(c)** parameter $\phi$ and of **(d)** parameter $\theta$.

volume-price is 0.999. While the correlation between $\phi$ and volume-price is 0.205, suggesting a weak relationship between them.

The plot (c) starts with a high autocorrelation at lag 1 that slowly declines. It continues decreasing until it becomes reach zero, starting to fluctuate around it. These suggest the presence of periodicity. However, the pattern found in volume-price and $\theta$ time series is not seen here.

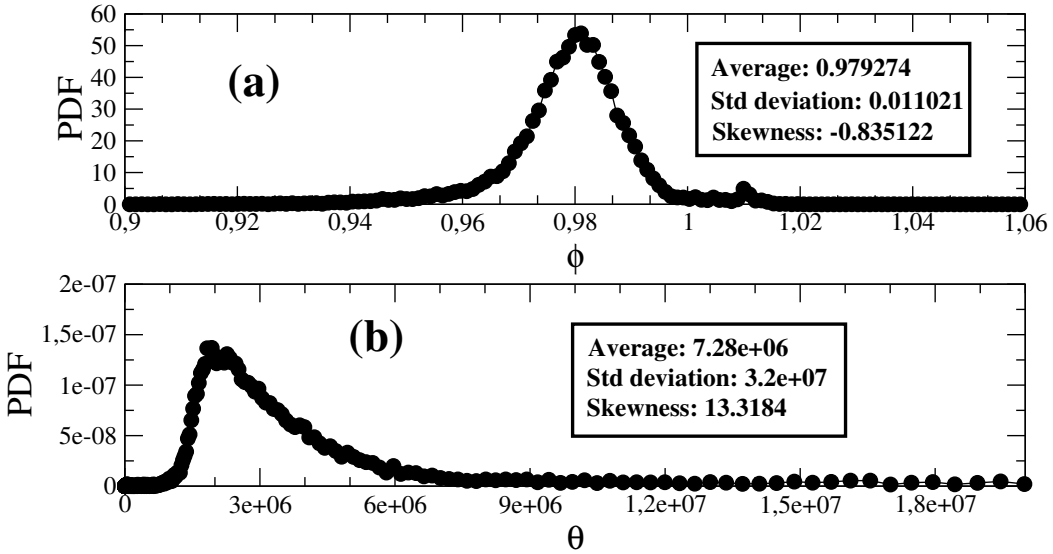


Figure 5.2: **(a)** PDF of parameter $\phi$ and **(b)** PDF of parameter $\theta$ from inverse-$\Gamma$ distribution.

Figure 5.2a and 5.2b shows the probability density function of each parameter $\phi$ and $\theta$. We can see that the values of parameter $\phi$ are spread in the interval $[0.96; 1]$.

## 5.2 Testing the Markov property

In this section, we present evidence that the Markov property holds. The data considered here is composed of approx. 25000 points. To test the Markov property of the $\phi$-series, we compute separately $p(x_1, \tau_1|x_2, \tau_2)$ and $p(x_1, \tau_1|x_2, \tau_2; x_3 = 0, \tau_3)$, and compare them.

Figure 5.3 shows the contour plot of $p(x_1, \tau_1|x_2, \tau_2)$ and $p(x_1, \tau_1|x_2, \tau_2; x_3 = 0, \tau_3)$, for the scale $\tau_1 = \tau_{\min} = 10\,\text{min}$, $\tau_2 = 2\tau_{\min}$ and $\tau_3 = 3\tau_{\min}$. The proximity of corresponding contour lines indicates that the equality

$$p(x_1, \tau_1|x_2, \tau_2) = p(x_1, \tau_1|x_2, \tau_2; x_3 = 0, \tau_3)\,, \tag{5.2}$$

holds for the chosen set of scales. Additionally, two cuts through the conditional probability densities are provided for fixed values of $x_1$, namely at $\langle x_1\rangle \pm \frac{1}{2}\sigma$.

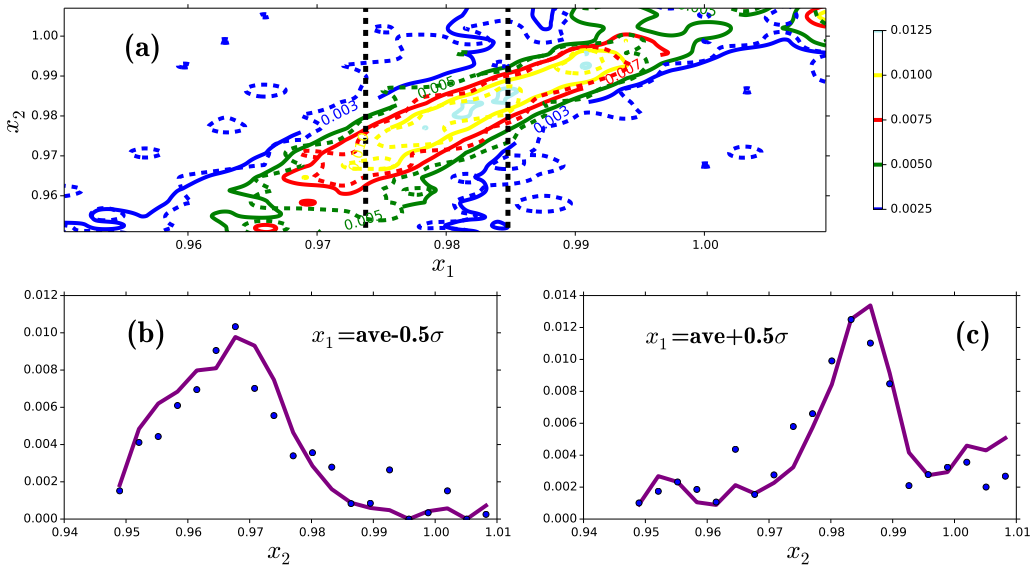


Figure 5.3: **(a)** Contour plots of the conditional PDF's $p(x_1, \tau_1|x_2, \tau_2)$ (solid lines) and $p(x_1, \tau_1|x_2, \tau_2; x_3 = 0, \tau_3)$ (dashed lines) for $\tau_1 = \tau_{\min}$, $\tau_2 = 2\tau_{\min}$ and $\tau_3 = 3\tau_{\min}$, with $\tau_{\min} = 10$ min. The dashed vertical lines at $x_1 = \langle x_1\rangle \pm \frac{\sigma}{2}$ indicates the cut shown in **(b)** and **(c)** respectively.

## 5.3 Drift and Diffusion coefficients

Having shown that within minimal accuracy we can assume that $\phi$ evolves according to a Markov process, we now describe mathematical framework introduce in[8], which allow us to estimate the values of the drift vector and diffusion matrix directly from data. The procedure is as follows.

We consider our data represented in by a discrete process from step $t_j$ to $t_{j+1} = t_j + \tau$ as

$$\phi(t_{j+1}) = \phi(t_j) + D_1(\phi(t_j), t_j)\tau + \sqrt{\tau}g(\phi(t_j), t_j)\Gamma(t_j)\,. \tag{5.3}$$

Equation (5.3) is a discrete version of the Langevin equation (2.34). Both functions $D_1$ and $D_2$ can be numerically determined for a set of bins $\phi_1, ..., \phi_n$, covering the full range of $\phi$-values. More precisely the drift vector assigned to bin $\phi_i$ can be determined as the limit when $\tau \to 0$ of

$$D_1(\phi_i) = \lim_{\tau \to 0} \frac{1}{\tau} M_1(\phi_i, \tau)\,, \tag{5.4}$$

where $M_1(\phi_i, \tau)$ is the conditional moment given by

$$M_1(\phi_i, \tau) = \frac{1}{N_i} \sum_{\phi(t_j) \in i} (\phi(t_j + \tau) - \phi(t_j))\,, \tag{5.5}$$

and the sum is over all the $N_i$ points, contained in bin $i$. Similarly, the diffusion matrix can be estimated as

$$D_2(\phi_i) = \lim_{\tau \to 0} \frac{1}{\tau} M_2(\phi_i, \tau)\,, \tag{5.6}$$

where the $M_2(\phi_i, t_j, \tau)$ is the second conditional moment, given by:

$$M_2(\phi_i, \tau) = \frac{1}{N_i} \sum_{j} [(\phi(t_j + \tau) - \phi(t_j))]^2\ . \tag{5.7}$$

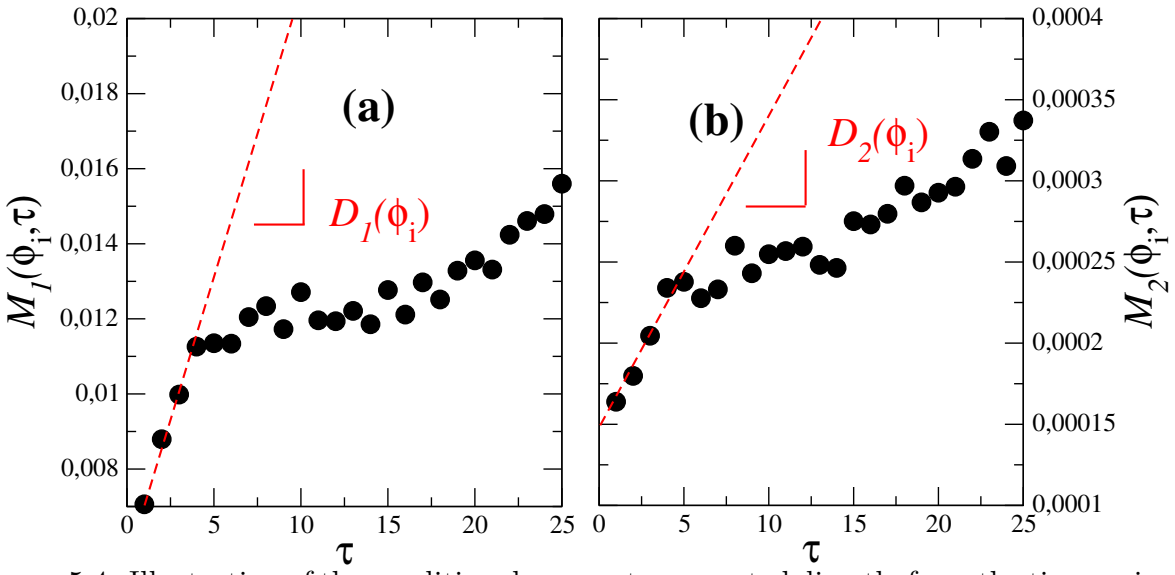


Figure 5.4: Illustration of the conditional moments computed directly from the time series of the $\phi$ time-series for the inverse-$\Gamma$: **(a)** first conditional moment $M_1$ and **(b)** first conditional moment $M_2$, from which one can conclude about the possible existence of measurement noise sources (see text). Here $x_i$ is the bin including the average value $\langle\phi\rangle$.

Figures 5.4a and 5.4b show the first and second conditional moments respectively, as a function of $\tau$, for a given bin value $\phi_i$. For the lowest range of $\tau$ values one sees a linear dependence of the conditional moments, which enables to directly extract the corresponding value of the drift and diffusion in Eq. (2.112). Namely, by computing the slopes of $M_1$ and $M_2$ for each bin in variable $\phi$ yields a complete definition of both the drift $D_1$ and the diffusion $D_2$ coefficients for the full range of observed $\phi$ values.

Figure 5.5a and 5.5b shows the drift and diffusion respectively. While the diffusion term has an almost constant amplitude, $D_2 \sim 10^{-8}$, the drift is linear on $\phi$ with a negative sloped and a fixed point close to one ($\phi_f \sim 0.99$).

Figure 5.5c shows the ratio $\frac{D_4}{D_2}$, revealing that $D_4 \sim 10^{-4} D_2$ which, within numerical accuracy, allow us to use the Pawula theorem 2.8.1.

From the results shown in Fig. 5.5 one can now propose a model for the evolution of $\phi$ which will enable us to derive a risk measure for the tail of volume-price.

$$d\phi = -k(\phi - \phi_f)dt + \beta dW_t \,. \tag{5.8}$$

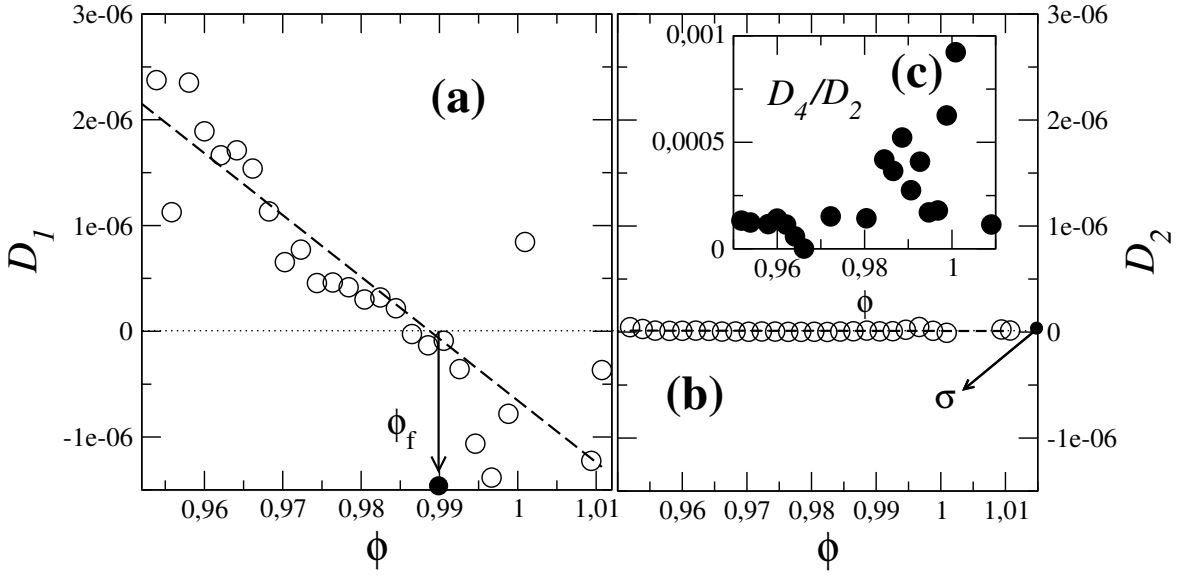


Figure 5.5: **(a)** The drift and **(b)** diffusion coefficients characterizing the stochastic evolution of the parameter $\phi$ that describes the tail of the inverse-$\Gamma$ distribution. **(c)** Quotient $\frac{D_4}{D_2}$, from this we can see that $D_4$ coefficient is negligible when compared with $D_2$ and $D_1$ (see text).

Notice that differential stochastic equation is similar to a stochastic equation for the Brownian motion of a particle under the influence of friction with friction coefficient $k$ and diffusion coefficient $\beta$.

According to corollary 2.7.5, the solution for such stochastic differential equation is given by:

$$\phi = \phi_0 e^{-k(t-t_0)} + \int_{t_0}^{t} e^{-k(t-s)}(k\phi_f ds + \beta dW_s) \,. \tag{5.9}$$

In appendix A we present the calculus of the expected value and the variance of $\phi$, which are given by

$$\mathbf{E}(\phi) = \mathbf{E}(\phi_0)e^{-k(t-t_0)} + \phi_f \left(1 - e^{-k(t-t_0)}\right) \xrightarrow[t\to\infty]{} \phi_f \tag{5.10}$$

and

$$\mathrm{Var}(\phi) = \frac{\beta^2}{2k}\left(1 - e^{-k(t-t_0)}\right) \xrightarrow[t\to\infty]{} \frac{\beta^2}{2k} \,. \tag{5.11}$$

For sufficiently long times, $t \to +\infty$, Eq. (5.8) describes the stochastic evolution with an

average $\phi_f$ and a variance of $\frac{\beta^2}{2k}$. In other words $\phi$ take, typically values in the range $\left[\phi_f - \frac{\beta}{2k}, \phi_f + \frac{\beta}{2k}\right]$.

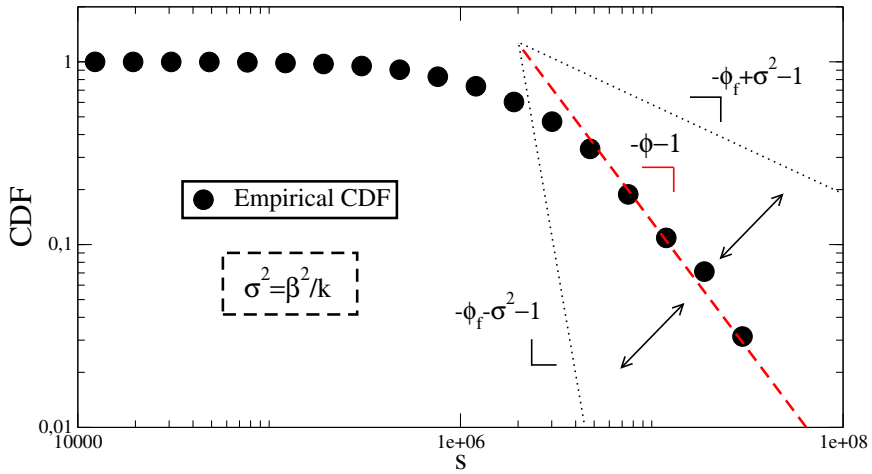


Figure 5.6: Due to the linear drift coefficient, namely, $D_1(\phi) = -k(\phi - \phi_f)$, we have an harmonic restoring force mechanism on volume-price tails.

As schematically represented in Fig. 5.6, the volume-price tails fluctuate around an inverse square law $\sim s^{-2}$ driven by a restoring force which can be modelled through Hooke's law. Furthermore, the fluctuations around the inverse square law are quantified by the diffusion amplitude $\sqrt{2D_2}$ (volatility) of the tail parameter. Since $\phi$ parameterise the exponent of the power-law that describes the tail of volume-price, by modelling it one could get more insight about the phenomenon of extreme events.

# Part III

# Discussions and conclusions

In this thesis we analysed the stochastic evolution of volume-price distributions present in the New York stock market during the last three years sampled with a lag of ten minutes. We tested four different candidate distributions typically used in finance and found that the best model depends (i) in the region of the spectrum that one wants to fit and (ii) in the time period during which the fit is made. Such finding is corroborated by the non-stationary character of such stochastic variables in finance.

Further, we were concerned with the study of extreme events present in the New York stock market. We investigate which is the best fit for the tail of volume-price distributions and found it to be the inverse-$\Gamma$ distribution, which has a power law tail.

Focusing on the parameter $\phi$ that controls the tail of the inverse-$\Gamma$ distribution we extracted a Langevin equation governing its stochastic evolution directly from the parameter's time series. While the deterministic contribution (drift) depends linearly on the parameter, with a restoring force around unity, the stochastic contribution (diffusion) is almost constant. Considering both contributions together, our findings show that the tail of the volume-price distributions evolve stochastically around an inverse square law with a constant parameter volatility.

This parameter volatility can be proposed as a risk measure for the expected tail of New York assets. It must be noticed that the above approach is only valid for Markovian processes, which seem to be the case of the parameter here considered. We tested it comparing two-point and three-point conditional probabilities.

Since volume-price distribution evolve stochastically, we also addressed the possibility of having not always the same model as optimal model. We found that the best model to fit the empirical data is not always the same. The log-normal distribution delivered the best fit in the center of the distribution when we consider all spectrum of empirical values of volume-price. Differently, when we look at the tail of the distribution covering the extreme events solely, the best fit, for the majority of our data was delivered by the inverse-$\Gamma$. Possible explanation for this lays in the fact that not always we are presented with extreme events, so the log-normal distribution would fit better the empirical data then the inverse-$\Gamma$.

For future work, one could try to make a combination of these two models. Perhaps through the using of a control parameter that would allow to continuous sweep from the log-normal to the inverse-$\Gamma$ model. Such a new hybrid model would perhaps describe better the dynamics present in the stock market.

The Langevin analysis here proposed can also be extended to both parameters $\phi$ and $\theta$ characterizing the inverse Gamma model. Next step would be to consider the evolution of these two parameters defining the inverse-$\Gamma$ distribution and extract a system of coupled stochastic differential equations.

Finally, whether the inverse Gamma distribution is in general the best to model volume-price tail distributions is up to our knowledge an open question. Still, the analysis here propose can be extended to other markets or even in other contexts where non-stationary processes are observed, in fields from physics, to biology, meteorology and medicine[8].

# Part IV

# Appendices

# Appendix A

# The stochastic evolution of the inverse-gamma tail

In Chap. 5 we showed that the parameter $\phi$ in the inverse-gamma is given by the following stochastic equation:

$$d\phi = -k(\phi - \phi_f)dt + \beta dW_t \,. \tag{A.1}$$

From 2.7.5 this differential has the solution

$$\phi = \phi_0 e^{-k(t-t_0)} + \int_{t_0}^{t} e^{-k(t-s)}(k\phi_f ds + \beta dW_s) \tag{A.2}$$

$$= \phi_0 e^{-k(t-t_0)} + \phi_f e^{-k(t-s)}|_{t_0}^{t} + \int_{t_0}^{t} \beta e^{-k(t-s)} dW_s \tag{A.3}$$

$$= \phi_0 e^{-k(t-t_0)} + \phi_f \left(1 - e^{-k(t-t_0)}\right) + \beta \int_{t_0}^{t} e^{-k(t-s)} dW_s \,. \tag{A.4}$$

The expected value follows

$$\mathbf{E}(\phi) = \mathbf{E}(\phi_0) e^{-k(t-t_0)} + \phi_f \left(1 - e^{-k(t-t_0)}\right) \,. \tag{A.5}$$

When $t \to +\infty$

$$\mathbf{E}(\phi) \to \phi_f \,. \tag{A.6}$$

The variance is given by

$$\mathrm{Var}(\phi) = \mathbf{E}(\phi^2) - \mathbf{E}(\phi)^2 \,, \tag{A.7}$$

with

$$\mathbf{E}(\phi^2) = \mathbf{E}\left(\left(\phi_0 e^{-k(t-t_0)} + \phi_f\left(1 - e^{-k(t-t_0)}\right) + \beta\int_{t_0}^{t} e^{-k(t-s)}dW_s\right)^2\right) \tag{A.8}$$

$$= \mathbf{E}\Bigg(\left(\phi_0 e^{-k(t-t_0)} + \phi_f(1 - e^{-k(t-t_0)})\right)^2$$

$$+ 2\left(\phi_0 e^{-k(t-t_0)} + \phi_f(1 - e^{-k(t-t_0)})\right)\left(\beta\int_{t_0}^{t} e^{-k(t-s)}dW_s\right)$$

$$+ \left(\beta\int_{t_0}^{t} e^{-k(t-s)}dW_s\right)^2\Bigg) \tag{A.9}$$

$$= \mathbf{E}\left(\left(\phi_0 e^{-k(t-t_0)} + \phi_f(1 - e^{-k(t-t_0)})\right)^2 + \left(\beta\int_{t_0}^{t} e^{-k(t-s)}dW_s\right)^2\right) \tag{A.10}$$

$$= \left(\mathbf{E}(\phi_0)e^{-k(t-t_0)}\right)^2 + 2\mathbf{E}(\phi_0)\phi_f(e^{-k(t-t_0)} + \left(e^{-k(t-t_0)}\right)^2)$$

$$+ \phi_f^2(1 - e^{-k(t-t_0)})^2 + \frac{\beta^2}{2k}e^{-2kt}\left(e^{2kt} - e^{2kt_0}\right), \tag{A.11}$$

yielding,

$$\mathrm{Var}(\phi) = \mathbf{E}(\phi^2) - \Big(\left(e^{-k(t-t_0)}\mathbf{E}(\phi_0)\right)^2$$

$$+ 2\mathbf{E}(\phi_0)\phi_f e^{-k(t-t_0)}(1 - e^{-k(t-t_0)}) + \phi_f^2(1 - e^{-k(t-t_0)})^2\Big) \tag{A.12}$$

$$= \frac{\beta^2}{2k}\left(1 - e^{-k(t-t_0)}\right). \tag{A.13}$$

# Part V

# References

# Bibliography


[1] Ludwig Arnold. *Stochastic Differential Equations: Theory and Applications*. 1992.

[2] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81:37–654, 1973.

[3] F. Boettcher, 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.

[4] S. Camargo, S.M.D. Queirós, and C. Anteneodo. Bridging stylized fatcs in finance and data non-stationarities. *Eur. Phys. J. B*, 86:159, 2013.

[5] A. Einstein. Investigation on the theory of, the brownian movement. *Ann. d. Physik*, 17:549, 1905.

[6] R. Friedrich and J. Peinke. Description of a turbulent cascade by a fokker-planck equation. *Phys. Rev. Lett.* , 78:863, 1997.

[7] R. Friedrich, J. Peinke, and Ch. Renner. How to quantify deterministic and random influences on the statistics of the foreign exchange market. *Phys. Rev. Lett.* , 84:5224, 2000.

[8] R. Friedrich, J. Peinke, M. Sahimi, and M.R.R. Tabar. Approaching complexity by stochastic methods: From biological systems to turbulence. *Phys. Rep.* , 506:87, 2011.

[9] R. Friedrich, J. Peinke, and M.R.R. Tabar. *Complexity in the view of stochastic processes*. 2008.

[10] S. Kullback and R.A. Leibler. On information and sufficiency. *An. Math. Stat.*, 22:7986, 1951.

[11] P. Langevin. On the theory of brownian motion. *C.R. Acad. Sci. (Paris)*, 146:530–533, 1908.

[12] P.G. Lind, A. Mora, J.A.C. Gallas, and M. Haase. Reducing stochasticity in the north atlantic oscillation with coupled langevin equations. *Phys. Rev. E*, 72:056706, 2005.

[13] B. Mandelbrot. The variation of certain speculative prices. *J. Business*, 36:94–419, 1963.

[14] R.C. Merton. Theory of rational option pricing. *Bell J. Econom. Manag. Sci.*, 4:141–183, 1973.

[15] H. Risken. *The Fokker-Planck Equation*. Springer, 1984.

[16] P. Rocha, F. Raischel, J.P. Boto, and P.G. Lind. Optimal models of extreme volume-prices are time-dependent.

[17] P. Rocha, F. Raischel, J. Cruz, and P.G. Lind. Stochastic evolution of volume-price distributions.

[18] R. Schfer and T. Guhr. Local normalization: Uncovering correlations in non-stationary financial time series. *Physica A*, 389:38563865, 2010.

[19] G.E. Uhlenbeck and L.S. Ornstein. On the theory ofthe brownian motion. *Phys.Rev.*, 36:823841, 1930.

[20] Y. Yuan, X. Zhuang, and Z. Liu. Price-volume multifractal analysis and application in chinese stock markets. *Physica A*, 391:3484–3495, 2012.