Beyond Black Scholes: Fourier Transform Methods for European Option Pricing

Marco Paroli

Supervised by Professor Marco Pirra

Co-supervised by Professor Nicola Borri

Department of Economics and Management

Academic Year 2021/22

Chair of Asset Pricing

LUISS Guido Carli

October 2022

A thesis submitted in partial fulfilment of the requirements for the degree of Corporate Finance.

Contents

Figures

Abbreviations

Notation

Probability Theory

$\mathbb{V}$	Variance
$\mathbb{C}$	Covariance
$\mathbb{E}$	Expectation
$\mathfrak{F}$	Fourier transform
$\mathbb{P}$	Probability measure
$\mathbb{E}_{\mathbb{P}}$	Expectation under $\mathbb{P}$
μ_X	Distribution measure of X
$\xrightarrow{p}$	Convergence in probability
Φ_X	Characteristic function of X
$\xrightarrow{d}$	Convergence in distribution
F_X	Cumulative distribution function of X
φ	Standard Normal probability density function
Φ	Standard Normal cumulative distribution function

Calculus

inf	Infimum
sup	Supremum
$\mathfrak{R}[z]$	Real part of z
$\mathfrak{I}[z]$	Complex part of z
log	Natural logarithm
$\mathbb{1}_A$	Indicator function
$:=$	Equal by definition
exp	Exponential function
$[X]^+$	Maximum between X and 0

Set Theory

$\cup$	Union
$\mathcal{F}$	σ-Field
$\mathcal{F}_t$	Filtration
2^{Ω}	Power set
$\emptyset$	Empty set
$\cap$	Intersection
Ω	Sample space
(a, b)	Open interval
$(X, \mathcal{A}, \mu)$	Measure space
$[a, b]$	Closed interval
$\mathcal{B}(\mathbb{R}^n)$	Borel sets in $\mathbb{R}^n$
$\mathcal{O}^n$	Open sets in $\mathbb{R}^n$
$(\Omega, \mathcal{F}, \mathbb{P})$	Probability space
$\mathbb{R}$	Set of real numbers
$\mathbb{N}$	Set of natural numbers
$\mathbb{C}$	Set of complex numbers
$\mathbb{R}_+$	Set of positive real numbers
$\mathbb{N}_+$	Set of positive natural numbers
$\mathbb{R}_0^+$	Set of nonnegative real numbers
$\mathbb{N}_0^+$	Set of nonnegative natural numbers

Introduction

In quantitative finance, a *derivative* represents a contract whose value is *derived* from other underlying variables (Hull, 2021). Such instruments date back to ancient times: the first derivative was likely dealt with by the Greek philosopher Thales of Miletus around 600 BC. Generally engaged in philosophising, philosophers certainly did not shine for their wealth. An anecdote in Aristotle's *Politics* recounts how Thales, to prove the fallacy of this commonplace, acquired the seasonal use of every nearby olive press at a discount. The subsequent harvest, which proved exceptionally fruitful, allowed him to dispose of the olive presses on his own terms, accumulating significant wealth in the process. Hence, Thales was the first to set up a contingent contract depending upon the realisation of an underlying state variable (here, the abundance of the harvest).

According to Aristotle, "*from his knowledge of astronomy he had observed while it was still winter that there was going to be a large crop of olives...*". Thales would therefore have used the derivative contract for speculative purposes, which is predominant in the use of such financial instruments today (Calderone, 2020). Another attractive perspective, provided by Taleb (2012), is instrumental in introducing the other primary reason for derivative contracts: *hedging*. From this perspective, Thales would have positioned himself to take advantage of his lack of knowledge rather than exploit his superior astronomy understanding: "*he did not need to understand too much the messages from the stars.*"

The following chapters are concerned with a particular type of derivative contract known as an option, giving its holder the right but not the obligation to execute a transaction on the underlying asset within a specified time frame. In particular, this thesis focuses on the valuation of European call options, giving the right to purchase a given stake in a company at a pre-determined price and a specified maturity date. The analysis conducted for the models may be equally carried out considering put options, whose value follows immediately from a fundamental result known as the put-call parity.

The contingent nature of options contracts leads to an asymmetric payoff structure, prompting the need for (more or less) sophisticated mathematical techniques for their valuation. Pricing a derivative contract is, intuitively, subject to the specification of the underlying asset's dynamics. In this context, the erratic behaviour of prices observed across different asset classes makes stochastic processes the fundamental tools to carry out the analysis. Hence, one of the first contributions to the field is believed to be the application of the Brownian motion by Bachelier (1900) to model stocks trading at the *Paris Bourse*.

Unfortunately, standard Brownian motion is unsuitable for modelling stock prices as it comes with undesirable properties (such as a support of $\mathbb{R}$). However, the same process leads to a much more realistic representation when used to model the dynamics of the log-price. The latter is the approach followed by Black and Scholes (1973) and Merton (1973), who assumed a geometric Brownian motion for the underlying asset's price and relied on the no-arbitrage argument to derive a fundamental partial differential equation. The solution to this equation still represents one of the most famous option pricing formulas, and earned Scholes and Merton (Black had died by the time) a Nobel prize in 1997. It is worth mentioning that the most significant contribution of the authors was not the PDE itself but rather the replication and no-arbitrage arguments used to derive it. The pricing equation worked fine until the *Black Monday* of October 1987, when all the major world markets dropped by at least 20%. Inherent to the geometric Brownian motion assumption is a Gaussian distribution for the returns on the underlying; after the market crash of 1987, investors started to realise *black swans* existed, and rare events are more frequent than under the Normal scenario. Evidence of such a change of mindset is acquired by looking at the volatility parameter in the BSM formula — assuming constant variance for the spot price — implicit in the market prices of options. It turns out that the value implied by market prices is higher for deep out-of-the-money options than for their at-the-money counterparts, revealing that investors think black swans are not so unlikely. Furthermore, empirical distributions of returns are indeed characterised by an excess kurtosis and a left asymmetry. In light of these findings, several models were proposed over the years to relax one or more of the assumptions outlined in 1973. Overall, the models presented fall either in the class of Lévy processes or that of frameworks based on stochastic volatility. The increasing complexity of the dynamics for the underlying price does, however, require more sophisticated mathematical tools to derive an option pricing formula. Luckily, in the early 2000s, authors such as Bakshi and Madan (2000) and Lewis (2001) provided a general solution leveraging the powerful mathematical concept of Fourier transform. This thesis is organised into four chapters, whose content is briefly described below:

- Chapter 1 concerns the mathematical tools needed to explore the pricing models discussed, including a brief overview of measure theory and its particular case known as probability theory. With such tools, the last part of the chapter explores the functional Central Limit Theorem by deriving the well-known Brownian motion starting from a symmetric random walk. Such a process provides the foundations for the BSM framework, whose replication argument is presented along with the general principle of risk-neutral valuation and the two fundamental theorems of asset pricing by Harrison and Pliska (1981, 1983).

- Chapter 2 focuses on the class of Lévy processes, where Brownian motion belongs as the only one with continuous paths. These models are designed to allow for discontinuities in the price path, resembling the dynamics observed on the announcement of a M&A transaction, the burst of a war (or a pandemic), and similar events. As a result, the distribution of returns induced by this class of processes will inevitably be characterised by an excess kurtosis. The chapter discusses the fundamentals of Lévy processes, such as the Poisson process and an elegant decomposition by Lévy and Itô, to conclude with the jump-diffusion model proposed by Merton (1976) and the asymmetric variance-gamma framework due to Madan et al. (1998).

- Chapter 3 regards models based on stochastic volatility. Here, an essential assumption for Lévy processes — the independence of increments — is relaxed. In addition to providing a more accurate fit of the price surface for contracts expiring further out in time, this class of models can explain two stylised facts of empirical distributions of returns — volatility clusters and the leverage effect — while retaining a leptokurtic density. The chapter focuses on deriving the pricing PDE under the influential stochastic volatility framework of Heston (1993) and its extension by Bates (1996), combining the benefits of stochastic volatility and jump-diffusion processes.

- Chapter 4 briefly discusses the concept of the Fourier transform and its fundamental application to probability theory, where each density is entirely characterised by its Fourier transform and which indeed goes under the name of the characteristic function. Given the latter, the original density can then be recovered by inverting the transform through a procedure presented by Gil-Pelaez (1951). The characteristic function of the price process serves a fundamental role in option pricing, thanks to the early results of Stein and Stein (1991) and Heston (1993) and the generalisations provided by Bakshi and Madan (2000) and Lewis (2001). The chapter then concentrates on the discrete Fourier transform — needed to deal with samples of data points belonging to a function — and an influential and efficient algorithm to compute it, due to Cooley and Tukey (1965). The fast Fourier transform is then applied to the pricing equation of Lewis (2001) and used to calibrate the models' parameters on a data set of call options written on the *Apple* stock. The chapter concludes with a discussion of the parameters implied by the market prices.

Chapter 1

Risk-Neutral Valuation

The insight backing the success of Thales' strategy lies in the asymmetry of his position, characterised by a finite but significant upside potential against a fixed cash outflow. This feature, however, is not found in the "simplest" type of derivative: *forward contracts*. On the contrary to a *spot contract*, here, two parties (or *legs*) agree to buy or sell an asset at a particular future time T and a pre-determined *delivery price* K. Forward contracts are settled at the *delivery date* T, when the underlying asset is "transferred" from the *short leg* to the *long leg*; letting S_T represent the value of the underlying S at maturity, the payoff to the long leg in a forward contract is $S_T - K$. As forwards correspond to *zero-sum games*, the payoff to the short leg is exactly $K - S_T$. Intuitively, since no funds are exchanged at the contract's inception, it costs nothing to open a position on a forward. Analytically, the delivery price is set to avoid *arbitrage* opportunities according to the *no-arbitrage pricing* principle introduced by Merton (1973). An arbitrage opportunity can be defined as a strategy that does not require any outflow at the inception and generates a cash inflow and a loss later, with positive and zero probability, respectively.

Definition 1.1. *A forward contract is an agreement to pay or receive a specified delivery price K at a delivery date T, in exchange for an underlying asset whose price at time $t \in [0, T]$ is S_t. The T-forward price F_t^T at time t corresponds to the delivery price that sets $F_t^T = K$.*

One can derive the no-arbitrage T-forward price fairly quickly by noting that a long position is equivalent to a leveraged investment in the underlying, as summarised in Figure 1.1. Suppose an agent takes a short forward position by agreeing to sell for F_t^T at time T an asset worth S_t at t. Assume the economy features a risk-free asset, such as a *zero-coupon bond* (ZCB), yielding a risk-free rate of return r. An indirect definition of the latter states that an asset with a positive probability of earning a return above r must have a positive probability of earning a return below it. Suppose the agent shorts S_t units of the bond and invest the proceeds in the underlying S. On the delivery date, the agent

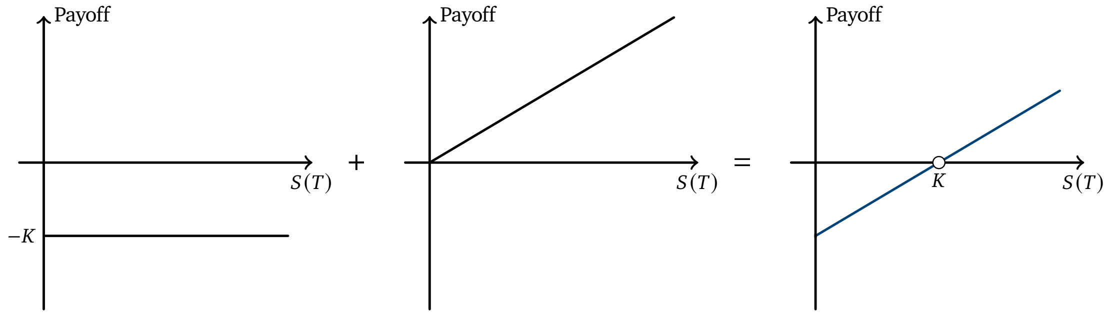


Figure 1.1 A long position on a forward contract is equivalent to — that is, can be *replicated* by — a debt-financed investment in the underlying asset.

delivers the asset and is left with $K - S_t e^{r(T-t)}$: if this amount is positive, the agent found an arbitrage opportunity; in the specular case, taking the long leg of the forward contract would provide a "free lunch". Hence, the no-arbitrage principle implies:

$$K - S_t e^{r(T-t)} = 0 \iff F_t^T = K = S_t e^{r(T-t)} \tag{1.1}$$

As a direct consequence of Definition 1.1, no funds are exchanged at the inception of a forward contract, and the payoffs at the delivery date correspond to the profits and losses (P&Ls) realised by the parties involved in the transaction.

Financial options are related to forward contracts in that they share the same upside potential. However, the most significant contribution provided by options is undoubtedly the introduction of asymmetry in the payoff, bounded at zero from below; clearly, the intrinsically contingent nature of options contracts comes at a cost.

Definition 1.2. *A call option gives its buyer the right, but not the obligation, to buy an underlying security for a pre-determined exercise (or strike) price within an expiration date T. Similarly, a put option gives its buyer the right, but not the obligation, to sell an underlying security for a pre-determined strike price within an expiration date T.*

A further distinction, due to Samuelson (1965), concerns the tenor in which the holder of an options contract can exercise his rights: an *American option* can be exercised at any time before or at maturity T; a *European option*, on the other hand, can only be exercised at maturity.* Nevertheless, American and European options share the same payoff at the exercise date $t \in [0, T]$, namely $[S_T - K]^+$ and $[K - S_T]^+$ for call and put options, respectively.

The existence of options may be rationalised through the two primary scopes such contracts are adopted in — speculation and hedging. The former is quite evident with

*Samuelson was introduced to options with early exercise as sophisticated financial instruments, understandable only to a European mind. He then decided to reverse the meanings in his definition.

options contracts, generally "controlling" for one hundred shares of the underlying — for instance, the holder of a (say) call will buy 100 shares of asset S for a price K, as long as $S_T - K > 0$. Such built-in leverage makes options contracts especially suitable for speculative purposes. Risk management represents another field featuring widespread use of options, which may be employed to hedge a position and comply with regulatory requirements set forth by banking authorities (e.g., *EBA*), such as *Value at Risk* (VaR).

The nonlinearity of options' payoffs is evident in a graphical representation of the latter (Figure 1.2). Determining the value provided by asymmetry, known as the option's *premium*, requires an accurate description of asset prices. In turn, such modelling assumptions are based on a compendium of formal definitions from the theory initiated by Kolmogorov in 1933 for probability fields. Such framework is thoroughly based on the contributions of Borel and Lebesgue, discussed in Appendix A.

1.1 | Probability Theory

The concepts presented in Appendix A are sufficient to introduce the theoretical underpinnings of this thesis. Probability theory is closely related to measure theory; the former is nothing but the latter restricted to a particular type of measure spaces $(X, \mathcal{A}, \mu)$ where $\mu(X) = 1$, as clarified by the following definition.

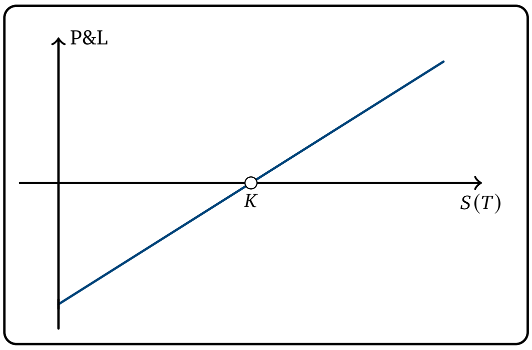
The measure-theoretic approach to probability theory unifies *discrete* and *continuous probability distributions* (see Bertrand paradox) while providing sensible definitions of probabilities outside of $\mathbb{R}^n$ (e.g., the space of continuous functions for *Brownian motion*).

Definition 1.3. *Let Ω be a nonempty set, and let $\mathcal{F}$ be a σ-algebra on subsets of Ω. Following Kolmogorov (1933), a probability measure $\mathbb{P}$ is a function that, to every set $A \in \mathcal{F}$, assigns a value in $[0, 1]$ called the probability of A and denoted $\mathbb{P}(A)$. We require $\mathbb{P}(\Omega) = 1$ and σ-additivity — that is, when $\{A_j\}_{j \in \mathbb{N}_+}$ is a sequence of pairwise disjoint events in $\mathcal{F}$ we have:*

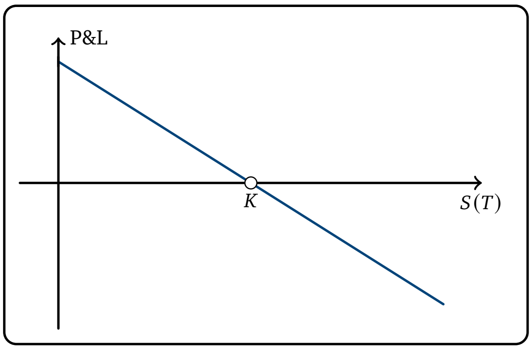
$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}\left(A_j\right) \tag{1.2}$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.

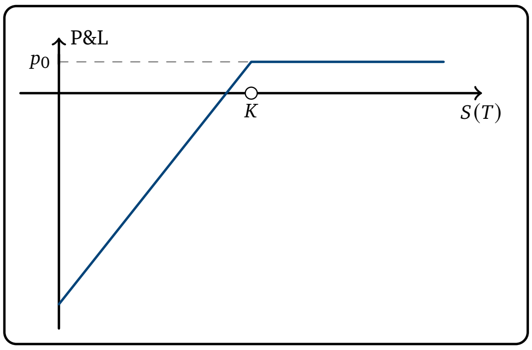
In terms of Definition A.6, a probability measure is just a map $\mathbb{P} \colon \mathcal{F} \to [0, 1]$ such that the Kolmogorov axioms are satisfied. When dealing with an *uncountably infinite* sample space Ω, each sequence $\omega = \omega_1 \omega_2 \ldots \in \Omega$ is assigned probability zero. For instance, if one were to toss a coin infinitely many times, the sequence $HHHH \ldots \in \Omega_\infty$ would be exactly



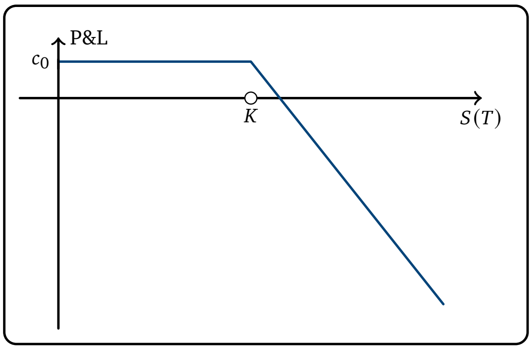
Long position on a forward contract



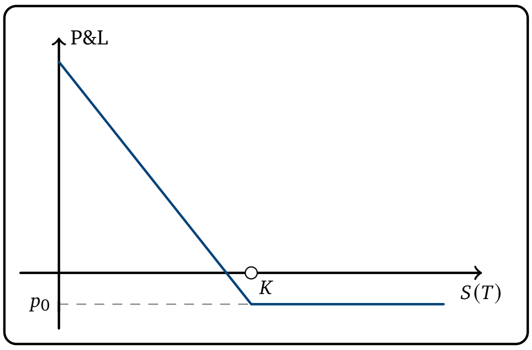
Short position on a forward contract



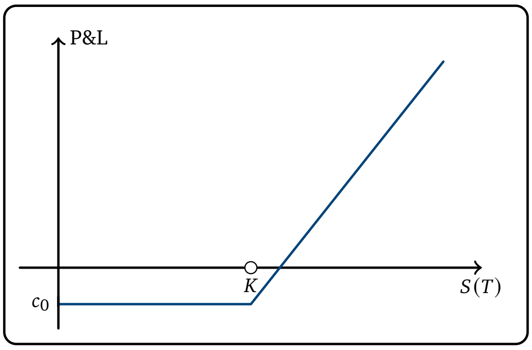
Short position on a put option



Short position on a call option



Long position on a put option



Long position on a call option

Figure 1.2 P&L functions for the most known *plain vanilla* derivative contracts. Forward prices are set so that the value of a forward contract is zero at the inception: payoff at maturity and P&L coincide. The main contribution of options contracts is due to their intrinsically nonlinear payoffs.

as likely as $HTHT\ldots \in \Omega_\infty$: both have zero probability. Formally, if a set A in the σ-algebra $\mathcal{F}$ on Ω is such that $\mathbb{P}(A) = 1$, then event A is said to occur *almost surely* (or *a.s.* for short).

At any time $t \geq 0$, the price of a given asset is modelled by means of a *random variable* (RV), prompting the need to provide a formal definition of the latter.

Definition 1.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a measurable space. A random variable X is a measurable map $X\colon \Omega \to \mathbb{R}$. Equivalently, X is a random variable if for every Borel-measurable set $B \in \mathcal{B}(\mathbb{R})$ the pre-image of X is $\mathcal{F}$-measurable:*

$$\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F} \tag{1.3}$$

Note that Definition 1.4 is restricted to the case of $\mathbb{R}$-valued RVs, where the line of real numbers $\mathbb{R}$ is equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{R})$.

If two RVs X and Y are such that $\mathbb{P}\{\omega \in \Omega : X(\omega) = Y(\omega)\} = 1$, X and Y are said equal *almost surely under* $\mathbb{P}$. If $\mu_X = \mu_Y$, then X and Y are said identical in distribution, a fact denoted by $X \overset{d}{=} Y$. Clearly, one has $X = Y\ a.s. \implies \mu_X = \mu_Y$.

It should be noted that knowledge of μ_X allows characterising a random variable X fully if and only if X represents the only source of randomness in the probabilistic model considered. In particular, when several RVs are involved, one can restrict the analysis to operations on the (marginal) distributions only if the RVs are mutually independent.

A random variable X *induces* a distribution measure μ_X on the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Intuitively, the distribution of a random variable is itself a probability measure assigned to the Borel sets $B \subset \mathcal{B}(\mathbb{R})$ rather than $F \subset \mathcal{F}$. Formally, μ_X is defined as follows.

Definition 1.5. *Let $X\colon \Omega \to \mathbb{R}$ be a real-valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution measure of X is defined as the map $\mu_X\colon \mathcal{B}(\mathbb{R}) \to [0,1]$ which assigns to each Borel set $B \subset \mathcal{B}(\mathbb{R})$ a mass $\mu_X(B)$, where*

$$\mu_X(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B) \tag{1.4}$$

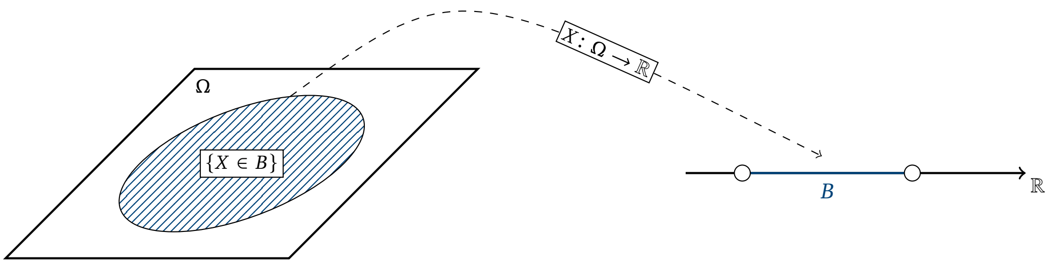


Figure 1.3 The distribution measure $\mu_X\colon \mathcal{B}(\mathbb{R}) \to [0,1]$ of a random variable X defines itself a probability measure on the members of the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ rather than on the events in $\mathcal{F}$.

One can easily show that μ_X satisfies the conditions mentioned in Definition 1.3 and is indeed a proper probability measure. First, by definition X is a measurable map $X : \Omega \to \mathbb{R}$; hence, $X^{-1}(\mathbb{R}) = \Omega \implies \mu_X(\mathbb{R}) = \mathbb{P}(\Omega) = 1$. Similarly, $X^{-1}(\emptyset) = \emptyset \implies \mu_X(\emptyset) = \mathbb{P}(\emptyset) = 0$. To show that μ_X satisfies σ-additivity, consider a sequence $\{B_j\}_{j\in\mathbb{N}_+}$ of pairwise disjoint Borel sets in $\mathcal{B}(\mathbb{R})$; since the pre-image of X is *stable under intersections and unions*,

$$i \neq j \implies X^{-1}(B_i) \cap X^{-1}(B_j) = X^{-1}(B_i \cap B_j) = X^{-1}(\emptyset) = \emptyset \tag{1.5}$$

Thus, $\{X^{-1}(B_j)\}_{j\in\mathbb{N}_+} \in \mathcal{F}$ is a sequence of pairwise disjoint measurable sets. Finally,

$$\mu_X\left(\bigcup_{j=1}^{\infty} B_j\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{j=1}^{\infty} B_j\right)\right) = \mathbb{P}\left(\bigcup_{j=1}^{\infty} X^{-1}(B_j)\right) = \sum_{j=1}^{\infty} \mathbb{P}\left(X^{-1}(B_j)\right) = \sum_{j=1}^{\infty} \mu_X(B_j) \tag{1.6}$$

so that μ_X satisfies σ-additivity and defines a proper probability measure. □

An essential notion is that of a random variable's *characteristic function* (CF), defined as the *Fourier transform* (FT) of the distribution measure induced by the RV. The CF of a random variable fully characterises its distribution, and many probabilistic properties of RVs are strictly related to the analytical properties of their CFs.

Definition 1.6. *The characteristic function $\Phi_X : \mathbb{R} \to \mathbb{C}$ of the $\mathbb{R}$-valued RV X is defined as*

$$\Phi_X(z) = \mathbb{E}\left[\exp(izX)\right] = \int_{\mathbb{R}} e^{izx}\, d\mu_X(x) \tag{1.7}$$

for all $z \in \mathbb{R}$.

The n-th *moment* of a RV X is defined as $m_n := \mathbb{E}[X^n]$. Similarly, the n-th *centered moment* μ_n is defined as the n-th *moment* of $(X - \mathbb{E}[X])^n$.

$$\mu_n := \mathbb{E}[(X - \mathbb{E}[X])^n] \tag{1.8}$$

A RV does not necessarily admit (finite) moments of all orders; for instance, it is well-known that the *Student's t-distribution* with n degrees of freedom has moments up to the n-th order. The moments of a RV are connected to the derivatives of its CF when evaluated at $z = 0$. In particular, the moments of X can be obtained through Φ_X as follows:

$$m_n := \mathbb{E}[X^n] = i^{-n} \frac{\partial^n}{\partial z^n} \Phi_X(z)\bigg|_{z=0} \tag{1.9}$$

It should be noted that a RV X admits finite moments of all orders if and only if the map $z \to \Phi_X(z)$ belongs to the class of twice differentiable continuous functions C^2.

Just as the value of an asset at a given point in time $t \geq 0$ is represented by a RV, the dynamics of such value are proxied by a *stochastic process* — that is, a collection $\{X_t\}_{t\geq 0}$ of random variables indexed by time t. Most importantly, a stochastic process may be seen as a function $X\colon [0,T]\times\Omega \to \mathbb{R}$ of both time $t \in [0,T]$ and the randomness $\omega \in \Omega$. Such observation prompts the need to define measures on *function spaces*. A trivial choice for a real-valued stochastic process would be the set of all maps $f\colon [0,T] \to \mathbb{R}$; such space, however, happens to be "too large". For instance, stochastic processes with continuous paths can be seen as RVs defined on the *space of continuous functions*, $C([0,T]\times\mathbb{R})$. A well-known (Gaussian) measure on $C([0,T]\times\mathbb{R})$ is given by the *Wiener measure*, describing the *Wiener process* explored in the next section. As the title of this thesis suggests, however, discontinuous paths will characterise most of the processes discussed.

Definition 1.7. *A function $f\colon [0,T] \to \mathbb{R}$ is said to be càdlàg ("continue à droite, limite à gauche") if it is right-continuous with left limits. For each $t \in [0,T]$ the limits*

$$f_{t-} \coloneqq \lim_{s\uparrow t} f_s \qquad f_{t+} \coloneqq \lim_{s\downarrow t} f_s \tag{1.10}$$

exist and $f_t = f_{t+}$

Clearly, all continuous functions are càdlàg, but the latter are also allowed to exhibit discontinuities. If t is a point of discontinuity for a càdlàg function f, $\Delta f(t) \coloneqq f(t) - f(t-)$ will denote the *jump* of f at time t. It can be shown that any càdlàg function is characterised by a finite number of jumps larger than ε for all $\varepsilon \in \mathbb{R}_+$ and (possibly) a countably infinite amount of "small" jumps (Fristedt and Gray, 1997).

One final remark shall be made on the choice of the càdlàg class as opposed to *càglàd* (*"continue à gauche, limite à droite"*) functions, where $f(t) = f(t-)$. In a time dimension, left stands for "before" and right for "after"; hence, if the price process were left-continuous, an investor could foresee the value at t by simply approaching time t along the process path. As jumps should not be predictable events, the class of càdlàg functions is appropriate.

One can imagine that, as time goes on and t increases, more and more information is revealed to the investor — that is, values proxied by random variables at time t may become deterministic at $\tau > t$ if the information revealed in τ allows to do so. Such notion is modelled through collections of σ-algebras indexed by time, known as *filtrations*.

Definition 1.8. *Let Ω be a non-empty set. Let T be a fixed positive number, and assume that for each $t \in [0,T]$ there exists a σ-algebra $\mathcal{F}_t$. Assume that if $s < t$, then every set in $\mathcal{F}(s)$ is also found in $\mathcal{F}_t$ — that is, $s < t \implies \mathcal{F}(s) \subseteq \mathcal{F}_t$. Then, for $t \in [0,T]$, we call the collection $\{\mathcal{F}_t\}_{t\in[0,T]}$ of non-decreasing σ-algebras a filtration.*

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration is called a *filtered probability space*. Once the information revealed over time is described by a filtration $\mathcal{F}_t$, one can distinguish between known quantities given $\mathcal{F}_t$ from those still random. A random variable is said to be $\mathcal{F}_t$-measurable if its value will be revealed at time t. A stochastic process whose realisation at time t is *resolved* by $\mathcal{F}_t$ is said to be *nonanticipating*.

Definition 1.9. *A stochastic process $\{X_t\}_{t\in[0,T]}$ is said to be nonanticipating (or adapted) with respect to the filtration $\{F_t\}_{t\in[0,T]}$ if the RV X_t is $\mathcal{F}_t$-measurable $\forall t \in [0,T]$.*

Consider now a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$.

Definition 1.10. *A càdlàg process $\{X_t\}_{t\in[0,T]}$ is said to be a martingale if*

1. *The process $\{X_t\}_{t\in[0,T]}$ is $\mathcal{F}_t$-adapted*
2. *The expectation $\mathbb{E}[|X_t|]$ is finite for all $t \in [0,T]$*
3. *For all $s < t$, $\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$*

Definition 1.10 makes sense only with respect to a filtration and once the probability measure $\mathbb{P}$ has been specified. When more than one mapping $\mathcal{F} \to [0,1]$ is considered, the term $\mathbb{P}$-martingale clarifies the measure under which the process is a martingale.

1.2 | Brownian Motion

Brownian motion (BM) is named after the Scottish botanist Robert Brown, who observed the jittery behaviour of minute pollen particles suspended in water. Thereafter, the well-known stochastic process has played a crucial role in financial engineering since its introduction in Bachelier's doctoral thesis.* Preceding the work of Einstein (1905) on the movement of a particle suspended in a liquid, Bachelier believed that stock returns should be independent and normally distributed. Following this intuition, he tried to model the price of an asset traded at the *Paris Bourse* by relying on a Brownian motion.

Around thirty years later, Kolmogorov (1931) referred to the construction of BM in Bachelier (1900). In the same paper, Kolmogorov proved that continuous *Markov processes*, also known as *diffusions*, are entirely determined by a parameter accounting for the *drift* and another for the *diffusive* part (i.e., the size of the random shock). A few decades later, the multiplicative version of Bachelier's BM would have laid the foundations for the famous model by Black and Scholes (1973) and Merton (1973).

A Brownian motion may be considered the continuous counterpart of a *random walk*. The purpose of this section is to explore such a proposition, which is just another way of phrasing *Donsker's invariance principle* (Donsker, 1951). Such a principle is nothing but a functional extension of the well-known *Central Limit Theorem*, reported below.

*It is worth mentioning that another appearance of Brownian motion can be traced back to the work on time series conducted in 1880 by Thiele of Copenhagen.

Theorem 1.1. *Let $\{X_i\}_{i=1}^n$ be a sequence of independent and identically distributed RVs with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2 < \infty$. Let $\bar{X}_n$ denote the sample average of the sequence — that is, $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Then, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ as n approaches infinity.*

Consider now a sequence of i.i.d. random variables $\{X_i\}_{i\in\mathbb{N}_+}$ such that each realisation of each RV is equally likely to be either 1 or −1. Letting S_n define the sum of the first n RVs, one has derived a random walk. The probabilistic properties of such a discrete process can be easily derived by noting that each RV is normally distributed with zero mean and unit variance — that is, $\mathbb{E}[X_i] = 0.5(-1+1) = 0$ and $\mathbb{V}[X_i] = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = 1$.

If the domain of the random walk is restricted to the time interval $[0, T]$, one may wonder about the value of S at a point in time t once an infinite number of steps is allowed between 0 and T. Clearly, some rescaling is needed: if all steps had unit length, the random walk would soon diverge to infinity. Hence, let us partition the interval $[0, T]$ into n equally spaced sub-intervals of length Δ (so that $\Delta n = T$) and define:

$$U_{k\frac{T}{n}} = \sqrt{\frac{T}{n}}S_k = \sqrt{\frac{T}{n}}\sum_{i=1}^{k} X_i \tag{1.11}$$

The (discrete) process defined by (1.11) is called a *rescaled* random walk. To reach a continuous framework, one needs to increase the number of steps n while adjusting k so that $k\Delta$ (i.e., the length of the interval over which S_k is defined) does not move excessively. Since the set $\mathbb{R}$ of real numbers is uncountable, for all $t \in [0, T]$, there exists $k \in \mathbb{R}$ such that $k\Delta \le t < (k+1)\Delta$. As a result, if the number of steps n in the interval $[0, T]$ increases, then k must also increase to counterbalance the decreasing Δ.

So far the analysis took place in a discrete framework with finitely many points $k = 1, \ldots, n$, whereas Brownian motion "lives" in a continuous time space. Hence, one may define U_t^n as a linear interpolation between $U_{k\frac{T}{n}}$ and $U_{(k+1)\frac{T}{n}}$:

$$U_t^n := U_{k\frac{T}{n}} + \left(t - k\frac{T}{n}\right)\left(U_{(k+1)\frac{T}{n}} - U_{k\frac{T}{n}}\right) = U_{k\Delta} + (t - k\Delta)\left(U_{(k+1)\Delta} - U_{k\Delta}\right) \tag{1.12}$$

As the number of steps n approaches ∞ the distance between $k\Delta$ and $(k+1)\Delta$ becomes infinitesimal and $k\Delta$ goes arbitrarily close to t, a fact denoted as $k\Delta \approx t$. By Theorem 1.1,

$$\frac{\frac{1}{k}\sum_{i=1}^k X_k - 0}{1/\sqrt{K}} = \frac{1}{\sqrt{k}}\sum_{i=1}^{k} X_k \xrightarrow{d} \mathrm{N}(0,1) \implies U_t^n \approx \sqrt{\frac{Tk}{n}}\frac{1}{\sqrt{k}}S_k \approx \sqrt{t}\frac{1}{\sqrt{k}}S_k \xrightarrow{d} \mathrm{N}(0,t) \tag{1.13}$$

Thus, as $n \to \infty$, U_t^N is confined within a region described by the distribution $\mathrm{N}(0, t)$.

One may also wonder how the difference $U_t^n - U_s^n$ behaves as $n \to \infty$ for some $t > s$. To this end, let us choose $k, l \in \mathbb{R}$ such that $k\Delta \le t < (k+1)\Delta$ and $l\Delta \le s < (l+1)\Delta$. As in the previous analysis, $k\Delta \approx t$ and $l\Delta \approx s$. Hence,

$$U_t - U_s \approx U_{k\frac{T}{n}} - U_{l\frac{T}{n}} = \sqrt{\frac{T}{n}} \sum_{i=l+1}^{k} X_i \tag{1.14}$$

Since all the X_is are independent and identically distributed, one may shift the index of the summation backwards so that it starts at $i = 1$. In order to do so, it is sufficient to notice that between $(l+1)$ and k there are exactly $(k-l)$ RVs X_i. As a result, by Theorem 1.1

$$U_t^N - U_s^N \approx \sqrt{\frac{T(k-l)}{N}} \frac{1}{\sqrt{(k-l)}} \sum_{i=1}^{k-l} X_i \approx \sqrt{t-s} \underbrace{\frac{1}{\sqrt{k-l}} \sum_{i=1}^{k-l} X_i}_{\xrightarrow{d} \mathrm{N}(0,1)} \xrightarrow{d} \mathrm{N}(0, t-s) \tag{1.15}$$

Therefore, the increments considered behave like a Normal distribution with zero mean and variance $t-s$ (i.e., the time interval over which such increments are measured). Furthermore, increments of a rescaled random walk are said to be stationary since their distributional properties hold irrespectively of the time interval chosen.

Finally, one may ask what is the relationship between the increments $U_{t_2}^n - U_{t_1}^n$ and $U_{t_4}^n - U_{t_3}^n$ for $t_1 < t_2 < t_3 < t_4$. By choosing some value l_i such that for $n \to \infty$ one has $l_i\Delta \approx t_i$, for n large enough it holds that X_i for $i = l_1+1, \dots, l_2$ and X_j for $j = l_3+1, \dots, l_4$ are independent. As a result, the increments considered are distributed as

$$U_{t_2}^n - U_{t_1}^n \approx U_{l_2\frac{T}{n}} - U_{l_1\frac{T}{n}} = \sqrt{\frac{T}{n}} \sum_{i=l_1+1}^{l_2} X_i \xrightarrow{d} \mathrm{N}(0, t_2 - t_1)$$
$$U_{t_4}^n - U_{t_3}^n \approx U_{l_4\frac{T}{n}} - U_{l_3\frac{T}{n}} = \sqrt{\frac{T}{n}} \sum_{i=l_3+1}^{l_4} X_i \xrightarrow{d} \mathrm{N}(0, t_4 - t_3)$$

Furthermore, these increments are independent since $\mathbb{C}[U_{t_2}^N - U_{t_1}^N, U_{t_4}^N - U_{t_3}^N] = 0$. Hence, a rescaled random walk U_t^n defined on the interval $[0, T]$ exhibits the following properties:

1. It is continuous.
2. It starts, without loss of generality, at zero.
3. It has independent and stationary increments.
4. At time t it behaves as a Normal distribution $\mathrm{N}(0, t)$.

The limit process arising as $n \to \infty$ and satisfying all these properties is known as Brownian motion, or Wiener process in honour of Wiener (1923). A limit process of a rescaled random walk satisfying only the second and third conditions belongs to the class of *Lévy processes*, allowed to exhibit jumps, as discussed in the next chapter.

The appropriateness of choosing Brownian motion as the *data generating process* is best explained by Figure 1.4, containing the (log-)price of the ACB stock over the last five years and five independent Brownian motion paths with the same drift and diffusive components. If the actual path had not been highlighted in blue, it would have been difficult to distinguish the actual stock price dynamics from the simulations.

Arithmetic Brownian motion, however, comes with several pitfalls that do not allow it to be used as a realistic model of stock prices. Suppose for the moment that the infinitesimal increment of the price of stock S, over an infinitesimal time interval denoted dt, is governed by the *stochastic differential equation* (SDE) below

$$dS_t = \mu\, dt + \sigma\, dW_t \tag{1.16}$$

where μ and σ are the expected drift and volatility of the process, respectively, and dW_t stands for the infinitesimal increment of a (standard) Brownian motion.

Clearly, (1.16) does not represent a sensible assumption for the price process. Indeed, integrating both sides on the interval $[0, T]$ one finds that:

$$\int_0^T dS_k = \int_0^T \mu\, dk + \int_0^T \sigma\, dW_k \implies S_T = S_0 + \mu T + \sigma W_T \implies S_T \sim \mathrm{N}(S_0 + \mu T, \sigma T) \tag{1.17}$$

Such configuration allows the stock price to become negative (the Normal distribution has support $\mathbb{R}$). Moreover, if S follows an *arithmetic Brownian motion* (ABM) volatility is constant regardless of the current stock price. None of these assumptions make particular sense, as the shareholders' personal belongings are shielded by limited liability and constant volatility would imply larger price movements as the latter decreases. A sensible alternative

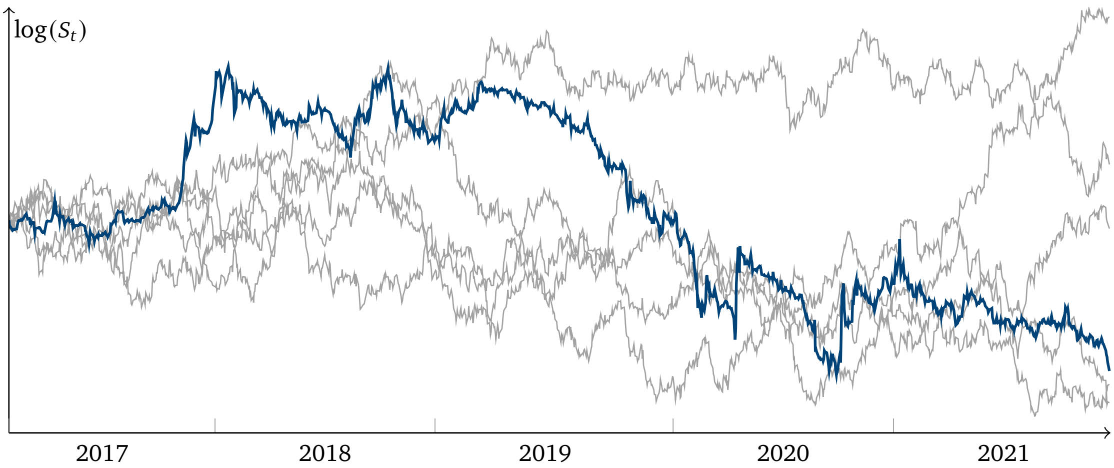


Figure 1.4 Evolution of the log-price of the ACB stock in the 2017-2021 period, compared with five independent Brownian motion paths rescaled to have the same annualised mean and volatility.

is to predict that the instantaneous rate of return follows an arithmetic Brownian motion:

$$\frac{dS_t}{S_t} = \mu \, dt + \sigma \, dW_t \iff dS_t = \mu S_t \, dt + \sigma S_t \, dW_t \tag{1.18}$$

The solution to (1.18) is known as *geometric Brownian motion* (GBM). To solve such SDE, however, one must rely on the analogous of the *chain rule* for *stochastic calculus*: the *Itô-Doeblin formula*. Such a need comes from the observation that Brownian motion paths are nowhere differentiable almost surely, violating the conditions for the existence of a "standard" Riemann-Stieltjes integral and thus requiring a theory of stochastic integration.

First, notice that the discretised version of (1.16) with $S_t = X_t$ can be expressed as

$$X_{t+\Delta} - X_t = \mu_t \Delta + \sigma_t (W_{t+\Delta} - W_t) \tag{1.19}$$

To assess the dynamics of some smooth function $f(X_t, t)$ of class $C^\infty(\mathbb{R})$, one should compute $f(X_{t+\Delta}, t + \Delta)$ by relying on a second-order Taylor expansion around $(x_0, y_0) = (X_t, t)$:

$$\begin{aligned} f(X_{t+\Delta}, t+\Delta) \approx{} & f(X_t, t) + \frac{\partial f}{\partial X} \overbrace{(X_{t+\Delta} - X_t)}^{A} + \frac{\partial f}{\partial t}\Delta \\ & + \frac{1}{2}\left[\frac{\partial^2 f}{\partial X^2} \underbrace{(X_{t+\Delta} - X_t)^2}_{B} + 2\frac{\partial^2 f}{\partial X \partial t} \underbrace{(X_{t+\Delta} - X_t)\Delta}_{C} + \frac{\partial^2 f}{\partial t^2}\Delta^2 \right] \end{aligned} \tag{1.20}$$

Let us now evaluate the terms highlighted by keeping those that are of the same order of magnitude of Δ or $(W_{t+\Delta} - W_t)$, while neglecting all the others:

$$\begin{aligned} A: \quad & X_{t+\Delta} - X_t = \mu_t \Delta + \sigma_t (W_{t+\Delta} - W_t) \\ B: \quad & (X_{t+\Delta} - X_t)^2 = \mu_t^2 \Delta^2 + \sigma_t^2 (W_{t+\Delta} - W_t)^2 + 2\mu_t \sigma_t (W_{t+\Delta} - W_t)\Delta \approx \sigma_t^2 \Delta \\ C: \quad & (X_{t+\Delta} - X_t)\, \Delta = \mu_t \Delta^2 + \sigma_t (W_{t+\Delta} - W_t)\Delta \approx 0 \end{aligned}$$

Substituting A and B back into (1.20), one has

$$f(X_{t+\Delta}, t+\Delta) - f(X_t, t) \approx \frac{\partial f}{\partial X}\left[\mu_t \Delta + \sigma_t (W_{t+\Delta} - W_t)\right] + \frac{\partial f}{\partial t}\Delta + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}\sigma_t^2 \Delta \tag{1.21}$$

Finally, the Itô-Doeblin formula arises as the limit of (1.21) for Δ approaching zero:

$$df_t = \left[\frac{\partial f}{\partial t} + \frac{\partial f}{\partial X}\mu_t + \frac{1}{2}\frac{\partial^2 f}{\partial X}\sigma_t^2 \right] dt + \frac{\partial f}{\partial X}\sigma_t \, dW_t \tag{1.22}$$

An alternative formulation, equivalent to (1.22), is as follows

$$df_t = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial X} dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}(dX_t)^2 \tag{1.23}$$

One can now rely on the Itô-Doeblin formula to solve (1.18). In particular, a clever choice of $f(S_t, t)$ would be a function whose derivatives allow to cancel the S_t terms on the right-hand side. Hence, a natural choice would be the natural logarithm of S_t:

$$\begin{aligned} d\log(S_t) &= \frac{\partial \log(S_t)}{\partial t}\,dt + \frac{\partial \log(S_t)}{\partial S_t}\,dS_t + \frac{1}{2}\frac{\partial^2 \log(S_t)}{\partial S_t^2}(dS_t)^2 \\ &= \mu\,dt + \sigma\,dW_t - \frac{1}{2S_t^2}(\mu^2 S_t^2\,dt^2 + \sigma^2 S_t^2\,dW_t^2 + 2\mu\sigma S_t^2\,dt\,dW_t) \\ &= \mu\,dt + \sigma\,dW_t - \frac{1}{2}\mu^2\,dt^2 - \frac{1}{2}\sigma^2\,dW_t^2 - \mu\sigma\,dt\,dW_t \end{aligned}$$

Intuitively, since dt is an infinitesimal quantity, one can approximate dt^2 with 0. Similarly, since the Brownian increment is normally distributed with zero mean and variance dt, the term $dt\,dW_t$ can be neglected as well. Finally, the Brownian Motion's increment squared is equal to dt by *Itô isometry* (see Appendix A). The resulting stochastic differential is:

$$d\log(S_t) = \mu\,dt + \sigma\,dW_t - \frac{1}{2}\sigma^2\,dt = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma\,dW_t \tag{1.24}$$

Integrating both sides of the SDE above, one obtains

$$\begin{aligned} \int_0^t \log(S_t) = \left(\mu - \frac{1}{2}\sigma^2\right)\int_0^t dt + \sigma\int_0^t dW_t \implies \log(S_t) &= \log(S_0) + \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t \\ \implies S_t &= S_0 \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right] \end{aligned} \tag{1.25}$$

Geometric Brownian motion represents a superior choice for the price process of the underlying as, at each point in time $t > 0$, the latter is lognormally distributed, ruling out negative prices. As a result, if the underlying follows a Geometric Brownian motion, then (log-)returns, rather than prices, are normally distributed with constant volatility.

1.3 | Martingale Pricing

Years before the appearance of the famous Black-Scholes-Merton (BSM) formula, several authors had assumed the price of the underlying asset moved like a geometric Brownian motion and derived valuation formulas very similar to that of BSM (Boness, 1964; Samuelson, 1965; Sprenkle, 1961; Thorp, 1969). For example, Boness (1964) derived the following equation for the price of a European call option:

$$c_t = S_t\Phi\left[\frac{\ln(S_t/K) + (\mu + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right] - Ke^{-\mu(T-t)}\Phi\left[\frac{\ln(S_t/K) + (\mu - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right]$$

It is then clear that the breakthrough introduced by Black and Scholes (1973) and Merton (1973) was not the pricing formula itself but rather its derivation. As discussed below, the extension of Δ hedging from a discrete to a continuous framework, coupled with the no-arbitrage principle, allowed Black and Scholes (1973) and Merton (1973) to exclude the underlying's drift from the formula and thus apply *risk-neutral valuation*.

Having specified a stochastic process for the evolution of the underlying asset's price, the Δ-hedging argument in Black and Scholes (1973) is based on the following assumptions:

- The market is frictionless — that is, no transaction costs are involved
- It is possible to short sell fractional amounts of the underlying stock
- Fractional amounts can be lent and borrowed at the risk-free rate
- The underlying stock does not pay dividends

Let us now denote by $f(S_t, t)$ the price of a derivative whose payoff depends on the price of the underlying S_t and time t. The dynamics of f can be found by means of (1.22):

$$df(S_t, t) = \left[\frac{\partial f(S_t, t)}{\partial t} + \frac{\partial f(S_t, t)}{\partial S_t}\mu S_t + \frac{1}{2}\frac{\partial^2 f(S_t, t)}{\partial S_t^2}\sigma^2 S_t^2\right] dt + \frac{\partial f(S_t, t)}{\partial S_t}\sigma S_t \, dW_t \tag{1.26}$$

Consider a portfolio long one unit of the derivative and short Δ units of the underlying:

$$V_t = f_t(S_t, t) - \Delta S_t \implies dV_t = df_t(S_t, t) - \Delta \, dS_t \tag{1.27}$$

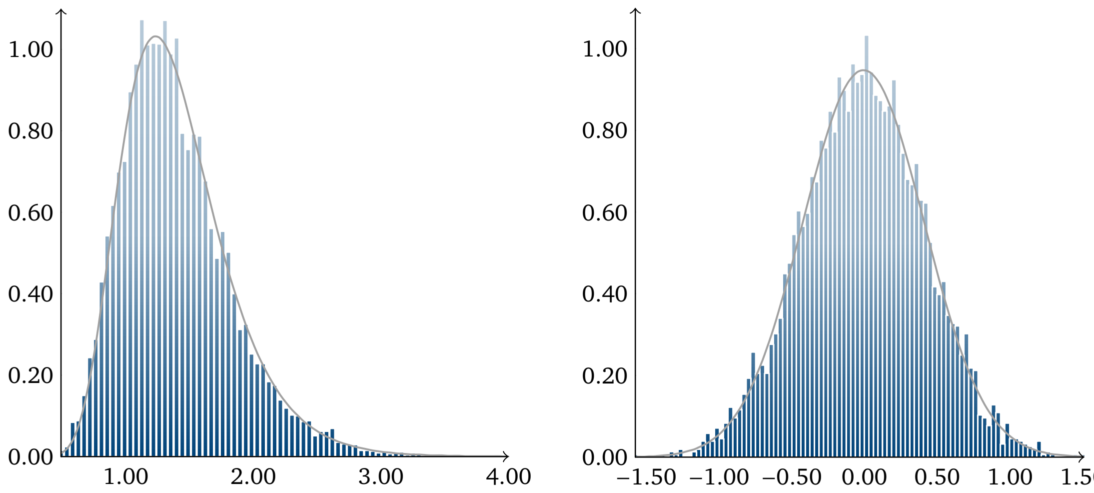


Figure 1.5 Estimated probability density of terminal stock prices with $S_0 = 1$ (left) and log-returns (right) through an histogram of 10 000 independent realisations of a geometric Brownian motion. The theoretical densities are superimposed on the respective histograms.

Substituting (1.18) and (1.26) into (1.27), one has

$$\begin{aligned} dV_t &= \left[\frac{\partial f(S_t,t)}{\partial t} + \frac{\partial f(S_t,t)}{\partial S_t}\mu S_t + \frac{1}{2}\frac{\partial^2 f(S_t,t)}{\partial S_t^2}\sigma^2 S_t^2 \right] dt + \frac{\partial f(S_t,t)}{\partial S_t}\sigma S_t\, dW_t - \Delta\left(\mu S_t\, dt + \sigma S_t\, dW_t\right) \\ &= \left[\frac{\partial f(S_t,t)}{\partial t} + \frac{\partial f(S_t,t)}{\partial S_t}\mu S_t + \frac{1}{2}\frac{\partial^2 f(S_t,t)}{\partial S_t^2}\sigma^2 S_t^2 - \Delta \mu S_t \right] dt + \left(\frac{\partial f(S_t,t)}{\partial S_t}\sigma S_t - \Delta \sigma S_t \right) dW_t \end{aligned}$$

At this point, setting Δ equal to the first derivative of the option's value with respect to the underlying makes the diffusive component disappear. In particular, substituting such Δ and making the appropriate simplifications, the dynamics of the portfolio become:

$$dV_t = \left[\frac{\partial f(S_t,t)}{\partial t} + \frac{1}{2}\frac{\partial^2 f(S_t,t)}{\partial S_t^2}\sigma^2 S_t^2 \right] dt \tag{1.28}$$

Since the portfolio just constructed does not bear any risk, to avoid arbitrage its instantaneous rate of return must coincide with the risk-free rate r. In particular,

$$dV_t = rV_t\, dt = r\left(f_t(S_t,t) - \frac{\partial f(S_t,t)}{\partial S_t} S_t \right) dt \tag{1.29}$$

Equating (1.28) and (1.29), one reaches the BSM *fundamental partial differential equation*:

$$\frac{\partial f(S_t,t)}{\partial t} + \frac{\partial f(S_t,t)}{\partial S_t} r S_t + \frac{1}{2}\frac{\partial^2 f(S_t,t)}{\partial S_t^2}\sigma^2 S_t^2 = r f(S_t,t) \tag{1.30}$$

The explicit solution of this *partial differential equation* (PDE) depends on the boundary conditions related to the terms of the derivative contract. In particular, the terminal payoff defines the boundary conditions needed to solve the FPDE.

The absence of arbitrage opportunities is closely linked to the existence of a probability measure $\mathbb{Q}$, *equivalent* to $\mathbb{P}$, under which the process followed by the discounted underlying's price defines a martingale. Two probability measures $\mathbb{P}$ and $\mathbb{Q}$ are said equivalent (denoted $\mathbb{P} \sim \mathbb{Q}$) if they share the same impossible events, that is:

$$\mathbb{P} \sim \mathbb{Q} \iff [\mathbb{P}(A) = 0 \iff \mathbb{Q}(A) = 0] \quad \forall A \in \mathcal{F} \tag{1.31}$$

Consider a measurable space $(\Omega, \mathcal{F})$ of scenarios referring to the prices of $d+1$ assets in the time interval $[0,T]$. Suppose the dynamics of the assets are given by

$$\begin{aligned} S\colon\ & [0,T] \times \Omega \to \mathbb{R}^{d+1} \\ :\ & (t,\omega) \mapsto \left(S_t^0(\omega), S_t^1(\omega), \ldots, S_t^d(\omega)\right) \end{aligned}$$

where $S_t^0 = e^{rt}$ represents the value at time t of one monetary unit invested in the *money market account*. The latter is used for discounting purposes: for any claim whose value is

V_t, the discounted value will be denoted by $\tilde{V}_t := V_t/S_t^0$. The discount factor $B(t,T) := S_T^0/S_t^0$ reduces to $B(t,T) = e^{-r(T-t)}$ if the *numéraire* S^0 corresponds to the money market account.

One possible way to represent a contingent claim maturing at time T would be to specify its payoff at maturity $H(\omega)$ for each realisation $\omega \in \Omega$. Note that, since by definition H is known at maturity, the terminal payoff is a $\mathcal{F}_t$-measurable map $H\colon \Omega \to \mathbb{R}$. With the notion of *pricing rule*, one refers to a procedure which assigns to a contingent claim H a value $V_t(H)$ at each point in time $t \in [0,T]$. A proper valuation rule should satisfy the following:

- A rule must be adapted — that is, $V_t(H)$ must be $\mathcal{F}_t$-measurable $\forall t \in [0,T]$
- Because of the contingent nature of the claim to be priced, a rule must be nonnegative

$$H(\omega) \geq 0 \;\forall \omega \in \Omega \implies V_t(H) \geq 0 \;\forall t \in [0,T] \tag{1.32}$$

- The rule *should* be linear — that is, for a portfolio of N contingent claims

$$V_t\left(\sum_{n=1}^{N} H_n\right) = \sum_{n=1}^{N} V_t(H_n) \tag{1.33}$$

For any measurable event $A \in \mathcal{F}$, let us denote by the RV $\mathbb{1}_A$ the payoff at T of a claim which pays one if A occurs and zero otherwise. For instance, $\mathbb{1}_\Omega$ represents a ZCB paying one monetary unit at time T. Hence, its value is given by $V_t(\mathbb{1}_\Omega) = e^{-r(T-t)}$.

Consider the following map $\mathbb{Q}\colon \mathcal{F} \to \mathbb{R}$

$$\mathbb{Q} := \frac{V_0(\mathbb{1}_A)}{V_0(\mathbb{1}_\Omega)} = e^{rT} V_0(\mathbb{1}_\Omega) \tag{1.34}$$

Then, by (1.32) and (1.33) one has that:

- Since $0 \leq \mathbb{1}_A \leq 1$, $0 \leq \mathbb{Q} \leq 1$
- If A and B are two disjoint measurable events in $\mathcal{F}$, $\mathbb{1}_{A+B} = \mathbb{1}_A + \mathbb{1}_B$ and by the linearity of the pricing rule $\mathbb{Q}(A \cup B) = \mathbb{Q}(A) + \mathbb{Q}(B)$

Allowing for infinite sums in (1.33), one concludes that σ-additivity is satisfied and thus $\mathbb{Q}$ defines a probability measure on the measurable space $(\Omega, \mathcal{F})$. There exists a strict correspondence between the valuation rule V and the probability measure $\mathbb{Q}$:

$$V_t(H) = e^{-r(T-t)} \mathbb{E}_{\mathbb{Q}}[H \mid \mathcal{F}_t] \quad \text{and} \quad \mathbb{Q}(A) = e^{-r(T-t)} V_t(\mathbb{1}_A) \tag{1.35}$$

Suppose now a measurable event $A \in \mathcal{F}$ is such that under the *true-world* probability $\mathbb{P}$ one has $\mathbb{P}(A) = 0$. Clearly, in this scenario a random payoff $\mathbb{1}_A$ is worth nothing to the investor.

Under the pricing rule corresponding to $\mathbb{Q}$, however, $V_0(\mathbb{1}_A) = e^{-rT}\mathbb{E}_{\mathbb{Q}}[\mathbb{1}_A] = e^{-rT}\mathbb{Q}(A)$. Thus, it must be that $\mathbb{Q}(A) = 0$ in order for measure $\mathbb{Q}$ to be consistent with the true-world views on the scenarios contained in Ω. Suppose this is not the case: if $\mathbb{Q}(A) = 0$ while $\mathbb{P}(A) \neq 0$, one has that the claim on A bears no premium but a positive probability of yielding a nonnegative payoff — that is, a bet on A would represent an arbitrage. Therefore, to exclude the possibility of a "free lunch", $\mathbb{P}$ and $\mathbb{Q}$ must be equivalent probability measures.

Consider an asset S^i currently trading at S^i_t. Holding the stock until maturity leads to a payoff of S^i_T, whereas selling the stock now to invest the proceeds at the risk-free rate returns an amount $e^{r(T-t)S^i_t}$. By the *law of one price* implied by the no-arbitrage principle, these two strategies should share the same value at time t (under $\mathbb{Q}$); namely:

$$\mathbb{E}_{\mathbb{Q}}\left[S^i_T \mid \mathcal{F}_t\right] = \mathbb{E}_{\mathbb{Q}}\left[e^{-r(T-t)}S^i_t \mid \mathcal{F}_t\right] = e^{-r(T-t)}S^i_t \implies \mathbb{E}_{\mathbb{Q}}\left[\tilde{S}^i_T \mid \mathcal{F}_t\right] = \tilde{S}^i_t \tag{1.36}$$

The absence of arbitrage opportunities and the existence of $\mathbb{Q}$ are linked by means of the *first fundamental theorem of asset pricing*, due to Harrison and Pliska (1981).

Theorem 1.2. *Consider a market model defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathcal{F}_t$. Let $\{S^i_t\}_{t\in[0,T]}$ represent the price process followed by asset S^i. Then, the market does not admit arbitrage opportunities if and only if there exists a probability measure $\mathbb{Q} \sim \mathbb{P}$ such that the discounted price process(es) $\{\tilde{S}^i_t\}_{t\in[0,T]}$ define $\mathbb{Q}$-martingales.*

Figure 1.6 is included to stress that the martingale properties are satisfied only under the *risk-neutral measure* $\mathbb{Q}$, whereas the discounted price process still exhibits a positive drift under $\mathbb{P}$. In particular, suppose the stock price follows a geometric Brownian motion

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW^{\mathbb{P}}_t$$

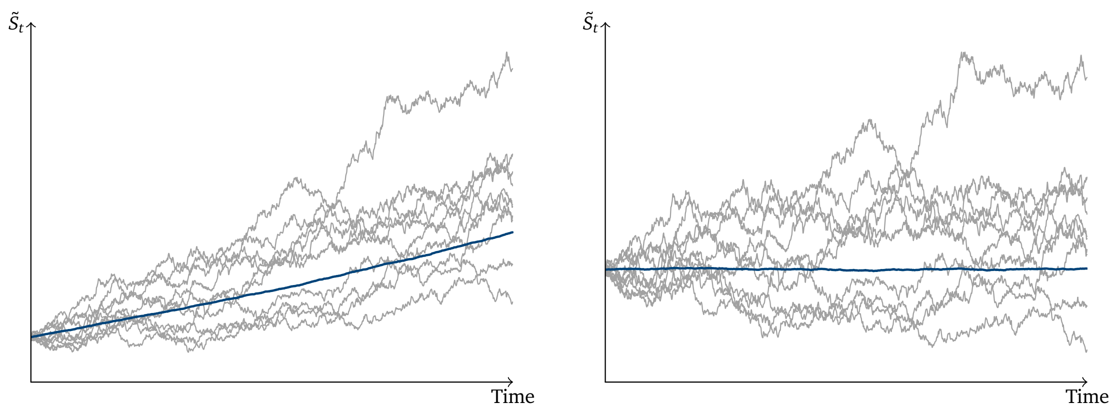


Figure 1.6 The discounted price process on the left defines a submartingale with respect to its natural filtration and the real-world probability measure $\mathbb{P}$. The graph on the right shows how the discounted price process defines a $\mathbb{Q}$-martingale once the change of measure has been applied.

where $dW_t^{\mathbb{P}}$ denotes the increment of a Brownian motion under the true-world probability measure $\mathbb{P}$. One can then apply the Itô-Doeblin formula to find the dynamics of $\{\tilde{S}_t\}_{t\in[0,T]}$:

$$\begin{aligned} d\tilde{S}_t &= \frac{\partial \tilde{S}}{\partial t}\,dt + \frac{\partial \tilde{S}}{\partial S}\,dS_t + \frac{1}{2}\frac{\partial^2 \tilde{S}}{\partial S^2}(dS_t)^2 \\ &= -re^{-rt}S_t\,dt + e^{-rt}(\mu S_t\,dt + \sigma S_t\,dW_t) \\ &= \tilde{S}_t(\mu - r)\,dt + \sigma\tilde{S}_t\,dW_t^{\mathbb{P}} \end{aligned} \tag{1.37}$$

It is then clear that the dynamics of the discounted price process given by (1.37) do not define a martingale. In particular, assuming the drift μ is greater than the risk-free rate r it is easy to see that $\mathbb{E}_{\mathbb{P}}[\tilde{S}_t \mid \mathcal{F}_s] > \tilde{S}_s$ for $s < t$: the discounted price process is said to define a *submartingale* under $\mathbb{P}$. In light of Theorem 1.2, however, one knows that (1.37) should be driftless under the risk-neutral measure $\mathbb{Q}$. Then, it must be that

$$\tilde{S}_t(\mu - r)\,dt + \sigma\tilde{S}_t\,dW_t^{\mathbb{P}} = \sigma\tilde{S}_t\,dW_t^{\mathbb{Q}} \implies dW_t^{\mathbb{Q}} = dW_t^{\mathbb{P}} + \left(\frac{\mu - r}{\sigma}\right)dt := dW_t^{\mathbb{P}} + \xi\,dt \tag{1.38}$$

where $dW_t^{\mathbb{Q}}$ denotes the increment of a (standard) Brownian motion under $\mathbb{Q}$ and ξ stands for the market price of risk, here given by the *Sharpe ratio* of the underlying. As a result, the risk-neutral dynamics of the stock price are given by

$$dS_t = \mu S_t\,dt + \sigma S_t\left[dW_t^{\mathbb{Q}} - \left(\frac{\mu - r}{\sigma}\right)dt\right] = rS_t\,dt + \sigma S_t\,dW_t^{\mathbb{Q}} \tag{1.39}$$

and the discounted price process whose dynamics are described by (1.37), once the change of measure from $\mathbb{P}$ to $\mathbb{Q}$ is performed, defines a $\mathbb{Q}$-martingale:

$$\begin{aligned} d\tilde{S}_t &= \tilde{S}_t(\mu - r)\,dt + \sigma\tilde{S}_t\,dW_t^{\mathbb{P}} \\ &= \tilde{S}_t(\mu - r)\,dt + \sigma\tilde{S}_t\left[dW_t^{\mathbb{Q}} - \left(\frac{\mu - r}{\sigma}\right)dt\right] \\ &= \sigma\tilde{S}_t\,dW_t^{\mathbb{Q}} \end{aligned} \tag{1.40}$$

In light of (1.40), the pricing rule $\mathbb{Q}$ is also called *equivalent martingale measure* (EMM).

Another fundamental notion stemming from Black and Scholes (1973) and Merton (1973) is that of market completeness. The latter amounts to stating that the payoff of any derivative instrument can be replicated by means of an appropriate trading strategy. Formally, a market is said complete if for any derivative contract H there exists a *self-financing* strategy $(\{\phi_t^0\}, \{\phi_t\})_{t\in[0,T]}$ such that:

$$H = V_0 + \int_0^T \phi_t\,dS_t + \int_0^T \phi_t^0\,dS_t^0 \tag{1.41}$$

almost surely under $\mathbb{P}$ and where V_0 represents the initial commitment needed to setup the replicating portfolio. It is clear that if (1.41) is verified with $\mathbb{P}$ equal to one, it must also hold under an equivalent martingale measure $\mathbb{Q} \sim \mathbb{P}$ and one has that:

$$\tilde{H} = V_0 + \int_0^T \phi_t \, d\tilde{S}_t \tag{1.42}$$

almost surely under $\mathbb{Q}$. Assuming the trading strategy ϕ_t for the underlying is such that $\int_0^T \phi_t \, d\tilde{S}_t$ defines a martingale, one concludes that $\mathbb{E}_{\mathbb{Q}}\left[\tilde{H}\right] = V_0$ — that is, under the pricing rule $\mathbb{Q}$ the value of any contingent claim H corresponds to the initial cost of a portfolio replicating H. Since every equivalent measure would lead to the same pricing rule, one may state that market completeness implies the uniqueness of a given equivalent martingale measure. The opposite is equally true, as shown in Harrison and Pliska (1983).

Theorem 1.3. *Consider a market model defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathcal{F}_t$. Let $\left\{S_t^i\right\}_{t \in [0,T]}$ represent the price process followed by asset S^i. The market is complete if and only if there exists a unique martingale measure $\mathbb{Q}$ equivalent to $\mathbb{P}$.*

However, it is well-known that perfect hedges are mostly a theoretical construct; as a result, a market model which assumes completeness is likely to return biased results. As discussed in the next chapter, most market models built on Lévy processes do not feature completeness, thereby providing a more accurate description of reality.

Chapter 2

Lévy Processes

The previous chapter introduced the concept of risk-neutral valuation, wherein the value of any contingent claim H_T written on an underlying S_t in an arbitrage-free market can be computed as an expectation under the equivalent martingale measure $\mathbb{Q}$:

$$V_t(H_T) = e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[H_T \mid \mathcal{F}_t] \tag{2.1}$$

Computing such expectation amounts to specifying the risk-neutral dynamics dictating the underlying's price movements. As mentioned in the discussion of Risk-Neutral Valuation, if the log-price $X_t := \log(S_t)$ follows an arithmetic Brownian motion:

$$dX_t = \left(r - \frac{\sigma^2}{2}\right) dt + \sigma \, dW_t^{\mathbb{Q}} \tag{2.2}$$

where $W_t^{\mathbb{Q}}$ is a standard Brownian motion under the equivalent martingale measure, then the price S_t follows a geometric Brownian motion. However, such an assumption comes with several implications that are hardly met when looking at the common properties characterising the time series of stock returns. In the BSM framework, the (log-)return on the underlying at any time $t \in [0, T]$ is normally distributed:

$$X_t \sim \mathrm{N}\left(X_0 + \left(r - \frac{1}{2}\sigma^2\right) t, \sigma^2 t\right) \tag{2.3}$$

It is well-known that stock returns do not obey a Gaussian probability law, as Mandelbrot (1963) pointed out. The departure from Gaussianity is already evident on a graphical level: Figure 2.1 shows the daily (log-)returns on six of the major stock indices globally. It is clear from a glance that a leptokurtic distribution with negative skewness characterises the time series under consideration. Alternatively, one could rely on econometric tests designed to assess the Gaussianity of a given sample. Among these, the technique with the best statistical power for a given significance level is the one proposed by Shapiro and Wilk (1965). Based on the p-values reported in Table 2.1, one concludes that all the indices considered had returns leading to a non-Gaussian distribution.

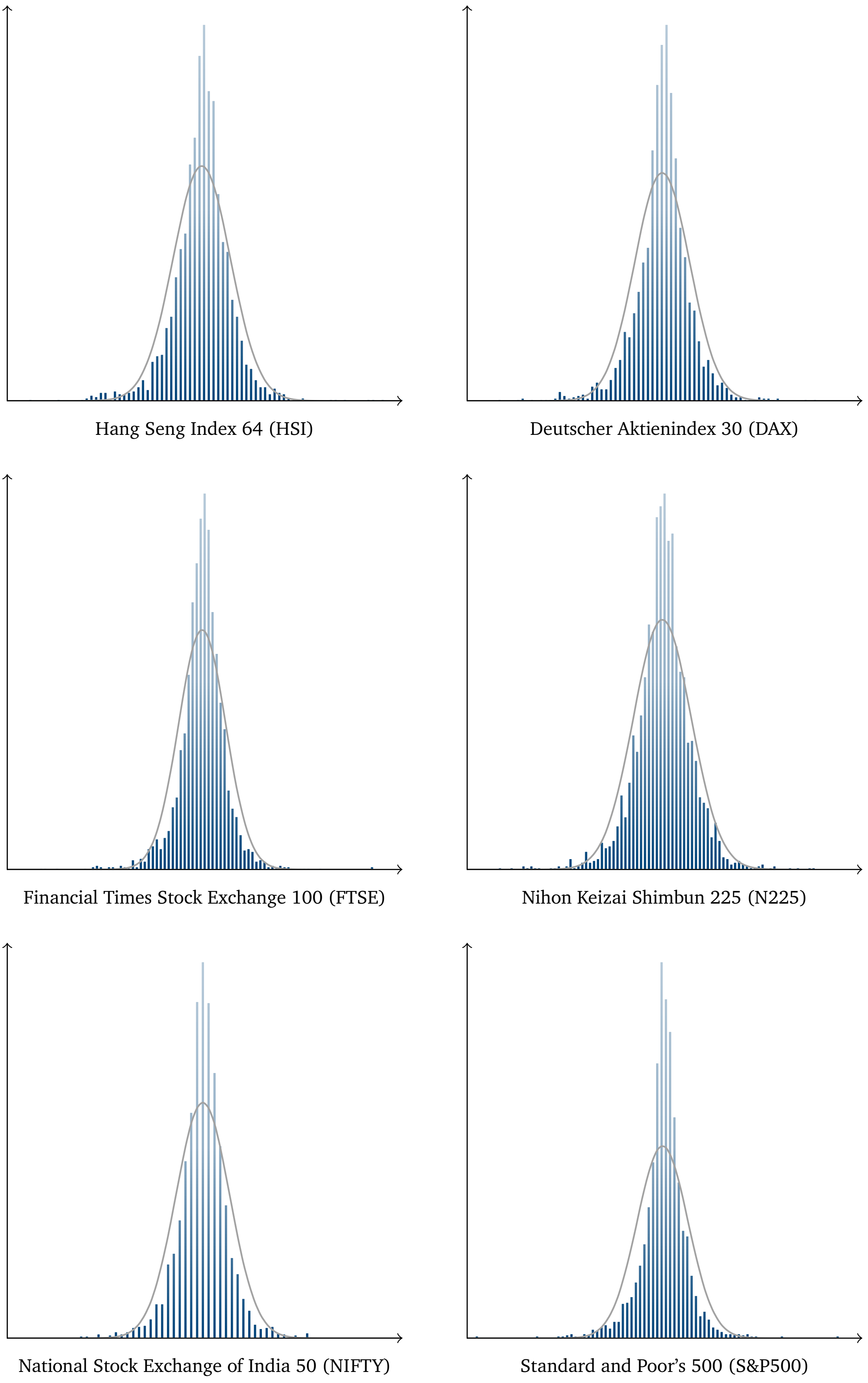


Figure 2.1 Histograms of the daily returns on six of the major stock indices globally. A Normal density with the same moments as the series considered is superimposed in grey in each plot. The time series range from 19/09/2007 to 13/04/2022: in such a period, all the empirical distributions of daily returns are leptokurtic and characterised by a negative asymmetry.

	μ	σ	$\mathcal{S}$	$\mathcal{K}$	$\mathbb{P}(W)$
HSI	-0.053	0.232	-0.323	7.522	0.000
DAX	-0.000	0.224	-0.567	7.467	0.000
FTSE	-0.057	0.191	-0.628	9.544	0.000
N225	-0.009	0.235	-0.743	6.605	0.000
NIFTY	0.068	0.218	-0.006	14.187	0.000
S&P500	0.046	0.207	-1.159	12.747	0.000

Table 2.1 Results of the Shapiro-Wilk test for normality applied to the six time series of stock a considered. The mean (μ) and standard deviation (σ) reported are annualised, contrary to skewness ($\mathcal{S}$) and kurtosis ($\mathcal{K}$). Finally, the last column shows the p-value associated with the Shapiro-Wilk test: a p-value below 5.00% allows one to reject the null of normality. Hence, all the empirical distributions are far from following a Gaussian distribution law.

Although deviations from normality were already documented in the 1960s, it is believed that the framework proposed by Black and Scholes (1973) and Merton (1973) has been "broken" since the market crash in 1987. Indeed, somewhat ironically, the first substantial deviation from the BSM model predictions was recorded in the aftermath of Black Monday 1987. One of the most significant implications of a specification of the form (2.2) for the process followed by the logarithm of the underlying is that volatility σ should be constant between different maturities and strike prices. Furthermore, solving the fundamental PDE (1.30) with the terminal condition $f(S_T,T) = [S_T - K]^+$, the resulting price of a European call option maturing in $\tau := T - t$ is given by:

$$V_t^{BS} = S_t \Phi\left[\frac{\log(S_t/K) + (r + \sigma^2/2)\tau}{\sigma\sqrt{\tau}}\right] - Ke^{-r\tau}\Phi\left[\frac{\log(S_t/K) + (r - \sigma^2/2)\tau}{\sigma\sqrt{\tau}}\right] \tag{2.4}$$

This equation, caeteris paribus, defines a monotonically increasing function of the underlying's volatility σ, mapping $(0,+\infty)$ into $([S_t - Ke^{-r\tau}]^+, S_t)$*. Hence, given a market price V_t^M one can invert (2.4) and recover the unique value of Σ_t reconciling the BSM and market prices through some root-finding algorithm (e.g., *Newton-Raphson* or *Brent*). Analytically,

$$\exists!\ \Sigma_t(K,T) > 0 : V_t^{BS}(S_t, K, \tau, r, \Sigma_t(K,T)) = V_t^M(K,T) \tag{2.5}$$

The mapping $\Sigma_t : (K,T) \longrightarrow \Sigma_t(K,T)$ is known as the *implied volatility surface* at time t; for a given maturity T and exercise price K, $\{\Sigma_t\}_{t\geq 0}$ defines a stochastic process. If the market followed the assumptions set forth by Black and Scholes (1973) and Merton (1973), one would expect a flat surface across strikes and maturity dates. However, this is not the case once one looks at the empirically extracted volatility surface. As shown in Figure 2.2, for a given maturity T, implied volatility tends to exhibit a *skew*: out-of-the-money puts (calls)

*Independently of the model specified for the price process, one can easily derive the last interval by considering no-arbitrage relationships for the value of a European call.

are worth more (less) than their BSM counterparts. Moreover, the skew tends to flatten out as the time to maturity increases. Finally, it should also be noted that the skew is typical of equity markets, where out-of-the-money put options are used extensively as insurance. If one looks, for instance, at the volatility implied by options on a given exchange rate, one sees symmetry that gives rise to a genuine volatility smile.

Given the above premises, it is natural to move on to a generalisation of the BSM framework. This chapter will focus on including discontinuity points in the process followed by the underlying. Therefore, it is necessary to abstract from Brownian motion to progress to the more general class of Lévy processes.

Definition 2.1. *Let $X\colon [0,T]\times\Omega \longrightarrow \mathbb{R}$ be a real-valued (càdlàg) stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, $\{X_t\}_{t\in[0,T]}$ defines a Lévy process if and only if:*

1. *It starts, without loss of generality, at zero.*
2. *It has independent and stationary increments.*

   Independence: $\forall (t_0 < \cdots < t_n), X_{t_0} \perp\!\!\!\perp \left(X_{t_1} - X_{t_1}\right) \perp\!\!\!\perp \ldots \perp\!\!\!\perp \left(X_{t_n} - X_{t_{n-1}}\right)$

   Stationarity: The probability law followed by $X_{t+h} - X_t$ *is independent of* t

3. *It is stochastically continuous — that is,* $\forall \varepsilon > 0, \lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| \geq \varepsilon) = 0.$

It is at this point essential to stress that the last property in no way implies that the paths generated by a Lévy process are continuous. The assumption of stochastic continuity is included to ensure that any jumps and points of discontinuity occur at random times, thus excluding any "calendar effects" (Cont and Tankov, 2003). While assuming a Lévy process

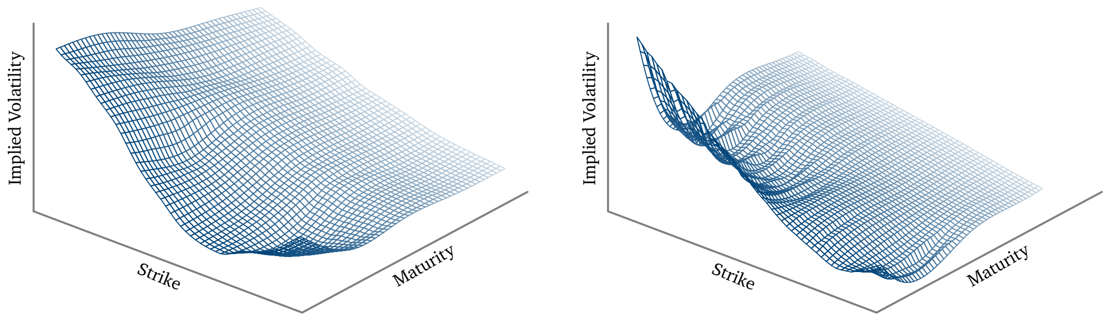


Figure 2.2 Implied volatility surface(s) as of 14/04/2022 on call options written on the AAPL stock and the S&P500 ETF Trust. The maturity of the options considered ranges from a minimum of fifteen trading days to a maximum of ninety-five trading days. The plots highlight how the premiums charged exhibit a lower deviation from the BSM predictions as the maturity increases.

for the (log-)price process provides a sufficient generalisation of the BSM framework, some constraints exist on the distributional properties of dX_t. In particular, at each time t, the distribution of X_t must belong to the class of *infinitely divisible distributions*.

Definition 2.2. *The distribution measure induced by a random variable X is said infinitely divisible if, for all $n \in \mathbb{N}$, there exists a collection of i.i.d. RVs $X_1^{(1/n)}, \dots, X_n^{(1/n)}$ such that:*

$$X \stackrel{d}{=} X_1^{(1/n)} + \dots + X_n^{(1/n)}$$

where $\stackrel{d}{=}$ indicates equality in distribution.

Since the CF of a RV completely characterises its distribution, the distribution measure of X is infinitely divisible if, for all $n \in \mathbb{N}$, there exists a random variable $X^{(1/n)}$ such that:

$$\Phi_X(z) = \left(\Phi_X^{(1/n)}(z)\right)^n \tag{2.6}$$

These two equivalent definitions can be used to show that a given distribution is infinitely divisible. For instance, consider a random variable $X \sim \mathrm{N}(\mu, \sigma^2)$; the Fourier transform of a normally distributed RV is known in closed form and given by:

$$\Phi_X(z) = \exp\left(iz\mu - \frac{z^2\sigma^2}{2}\right) = \exp\left[n\left(iz\frac{\mu}{n} - \frac{z^2\sigma^2}{2n}\right)\right] = \left[\exp\left(iz\frac{\mu}{n} - \frac{z^2\sigma^2}{2n}\right)\right]^n = \left(\Phi_X^{(1/n)}(z)\right)^n$$

where $X^{(1/n)} \sim \mathrm{N}(n^{-1}\mu, n^{-1}\sigma^2)$. Similarly, in line with Definition 2.2, one can easily see that $X \stackrel{d}{=} X_1^{(1/n)} + \dots + X_n^{(1/n)}$ where $X_i^{(1/n)} \sim \mathrm{N}(n^{-1}\mu, n^{-1}\sigma^2)$ for $i = 1, \dots, n$. In general, it can be shown that if μ_X is an infinitely divisible distribution then there exists a Lévy process $\{Y_t\}_{t\geq 0}$ where the distribution of Y_1 is given by μ_X (Sato, 1999). Well-known examples of infinitely divisible distributions include the gamma and Poisson distributions, whereas a uniform distributed RV does not induce an infinitely divisible probability law.

Consider now a Lévy process $\{X_t\}_{t\geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The characteristic function of the process at each point in time t is defined as:

$$\Phi_t(z) := \Phi_{X_t}(z) = \mathbb{E}[\exp(izX_t)] \quad \forall z \in \mathbb{R} \tag{2.7}$$

Since $\{X_t\}_{t\geq 0}$ is defined to have stationary independent increments, it holds that:

$$\Phi_{t+s}(z) := \Phi_{X_{t+s}}(z) = \Phi_{X_s}(z)\Phi_{X_{t+s}-X_s}(z) = \Phi_{X_s}(z)\Phi_{X_t}(z) := \Phi_s(z)\Phi_t(z) \quad \forall s < t \tag{2.8}$$

where the second equality holds because the characteristic function of the sum of i.i.d. RVs equals the product of the single characteristic functions. Furthermore, the stochastic continuity of the process ensures that X_s converges in distribution to X_t as s approaches t.

As a result, $\Phi_s(z) \to \Phi_t(z)$ as $s \to t$ and $t \to \Phi_t(z)$ is continuous in t. Hence, there exists a continuous map $\psi\colon \mathbb{R} \to \mathbb{C}$, known as the *characteristic exponent* of X, such that:

$$\Phi_{X_t}(z) = \mathbb{E}[\exp(izX_t)] = \exp(t\psi(z)) \quad \forall z \in \mathbb{R} \tag{2.9}$$

The assumption of stationary increments implies that the CF of X_t is linear in time t. As a result, the increments of a Lévy process sampled over time intervals of the same length will share the same characteristic function (and hence distribution).

2.1 | Poisson Process

Brownian motion, discussed in the previous chapter, is the only example of continuous Lévy process; the other "building block" needed to reach a general formulation of Lévy processes, thereby allowing for discontinuities in the resulting paths, is the *Poisson process*. The latter is closely related to exponentially distributed random variables: consider a RV τ following exponential distribution with probability density given by $\lambda e^{-\lambda\tau}\mathbb{1}_{\tau\geq 0}$ for $\lambda \in \mathbb{R}_+$.

Definition 2.3. *Let $\{\tau_i\}_{i\geq 1}$ be a collection of i.i.d. exponentially distributed random variables and define $T_n := \sum_{i=1}^{n} \tau_i$. Then, the process $\{N_t\}_{t\geq 0}$ given by*

$$N_t = \sum_{n\geq 1} \mathbb{1}_{t\geq T_n} = \#\{n \geq 1 : T_n \in [0,t]\} \tag{2.10}$$

is known as Poisson process with parameter λ.

By definition, the Poisson process exhibits càdlàg paths — that is, at any discontinuity point $N_t = N_{t+}$. Since $\{N_t\}_{t\geq 0}$ is built upon a sequence of exponentially distributed RVs, one can easily see that the process has independent and stationary increments. Furthermore, the Poisson process verifies the Markov property:

$$\mathbb{E}\left[N_t \mid N_u\right] = \mathbb{E}\left[N_t \mid N_s\right] \quad \forall u \leq s < t \tag{2.11}$$

One can impose the martingale property on the process by subtracting a given *compensator* (here, the first moment), thereby defining its *compensated* version:

$$\tilde{N}_t = N_t - \lambda t \tag{2.12}$$

whose characteristic function is then given by

$$\Phi_{\tilde{N}_t}(z) = \mathbb{E}\left[\exp(iz\tilde{N}_t)\right] = \exp\left[\lambda t(e^{iz} - 1 - iz)\right] \quad \forall z \in \mathbb{R} \tag{2.13}$$

Since the Poisson process has independent increments one concludes that

$$\begin{aligned}\mathbb{E}\left[N_t \mid \mathcal{F}_s\right] &= \mathbb{E}\left[N_t - N_s + N_s \mid \mathcal{F}_s\right] \\ &= \mathbb{E}\left[N_t - N_s \mid \mathcal{F}_s\right] + N_s = \lambda(t-s) + N_s \implies \mathbb{E}\left[\tilde{N}_t \mid \mathcal{F}_s\right] = \tilde{N}_s \quad \forall s \leq t\end{aligned} \tag{2.14}$$

Figure 2.3 presents sample paths of the Poisson process and its compensated version; note that the latter is no longer integer-valued and does not thus define a counting process.

Since the Poisson process counts the number of points on $[0, \infty)$ in a given interval $[0, t]$, it "defines" a *random jump measure* M such that for any measurable set $A \subset \mathbb{R}_+$

$$M(\omega, A) = \#\{n \geq 1 : T_n(\omega) \in A\} \tag{2.15}$$

The expectation of such measure is dictated by both the intensity λ of the Poisson process and the measurable set A considered. In particular, $\mathbb{E}\left[M(\omega, A)\right] = \lambda|A|$ where $|A|$ stands for the Lebesgue measure associated to set A. The Poisson process may be expressed in terms of the random measure just defined, namely:

$$N_t(\omega) = M(\omega, [0, t]) = \int_0^t M(\omega, ds) \tag{2.16}$$

Moreover, for all random measures M with intensity μ, measurable sets A such that $\mu(A) < \infty$, and functions f such that $\int_A \exp(f(x))\mu(dx) < \infty$ it holds that

$$\mathbb{E}\left[\exp\left(\int_A f(x)M(dx)\right)\right] = \exp\left[\int_A \left(e^{f(x)} - 1\right)\mu(dx)\right] \tag{2.17}$$

In general, one can associate to any càdlàg process $\{X_t\}_{t\geq 0}$ a so-called (random) jump measure J_X such that, for any measurable set $A \subset \mathbb{R}$, $J_X([0, t] \times A)$ counts the number of jumps occurring in the interval $[0, t]$ whose size lies in A:

$$J_X([0, t] \times A) = \#\{s \in [0, t] : (X_s - X_{s-}) \in A\} \tag{2.18}$$

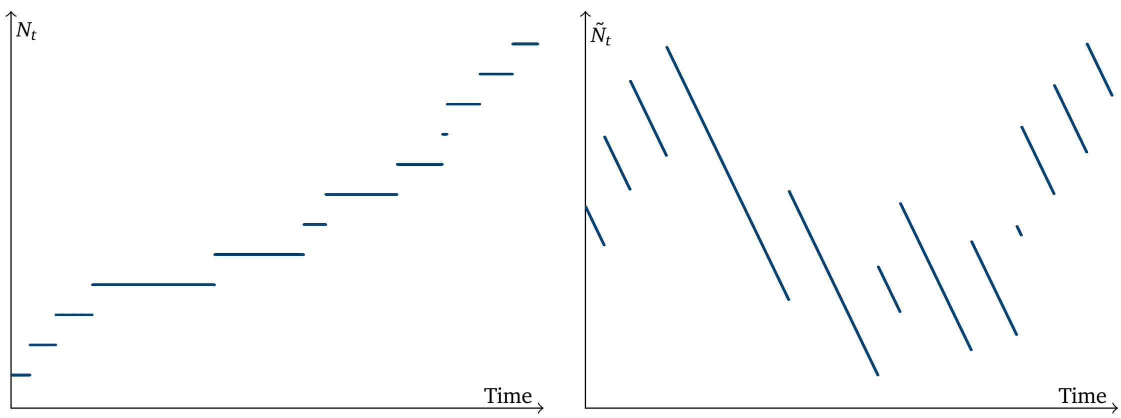


Figure 2.3 Left: Sample path of a Poisson process $\{N_t\}_{t\geq 0}$ with jump intensity $\lambda = 1$. Right: Compensated Poisson process $\{\tilde{N}_t\}_{t\geq 0}$ associated to the realisation on the left. The compensated version $\{\tilde{N}_t\}_{t\geq 0}$ shall not be interpreted as a counting process as it is no longer integer-valued.

In order to obtain a more accurate description of asset price dynamics it is essential to abstract from jumps in unit value, the only ones that a Poisson process can generate, and introduce an arbitrary probability law for the jumps.

Definition 2.4. *Let $\{N_t\}_{t\geq 0}$ be a Poisson process with jump intensity λ and $\{Y_i\}_{i\in\mathbb{N}}$ a collection of i.i.d. RVs following a distribution f and independent from the Poisson process. Then,*

$$X_t = \sum_{i=1}^{N_t} Y_i \tag{2.19}$$

defines a compound Poisson process with jump intensity λ and jump size distribution f.

Hence, X_t may be seen as the value of a random walk after a random number of steps given by N_t; the characteristic function of a compound Poisson process is given by

$$\Phi_{X_t}(z) = \mathbb{E}\left[\exp(izX_t)\right] = \exp\left[\lambda t \int_{-\infty}^{+\infty} (e^{izx} - 1) f(dx)\right] \quad \forall z \in \mathbb{R} \tag{2.20}$$

Such CF may be further simplified by defining the measure $\nu(A) = \lambda f(A)$; (2.20) becomes

$$\Phi_{X_t}(z) = \mathbb{E}\left[\exp(izX_t)\right] = \exp\left[t \int_{-\infty}^{+\infty} (e^{izx} - 1) \nu(dx)\right] \quad \forall z \in \mathbb{R} \tag{2.21}$$

Note that ν does not always define a probability measure: indeed, $\int_{\mathbb{R}} \nu(dx) = \lambda$ does not necessarily equal one. Furthermore, it can be shown that the random jump measure defined on $\mathbb{R} \times [0, \infty)$ associated to a compound Poisson process $\{X_t\}_{t\geq 0}$ has intensity $\mu(dx \times dt) = \nu(dx)dt = \lambda f(dx)dt$ (Cont and Tankov, 2003). As a result, every compound Poisson process $\{X_t\}_{t\geq 0}$ may be represented in terms of the jump measure just defined:

$$X_t = \sum_{s\in[0,t]} (X_s - X_{s-}) = \int_0^t \int_{\mathbb{R}} x J_X(ds \times dx) \tag{2.22}$$

where J_X is a jump measure with intensity $\lambda f(dx)dt$.

Definition 2.5. *Let $\{X_t\}_{t\geq 0}$ be a real-valued Lévy process. One can define a measure ν on $\mathbb{R}$:*

$$\nu(A) = \mathbb{E}\left[\#\{t \in [0,1] : (X_t - X_{t-}) \in A\}\right] \quad \forall A \in \mathcal{B}(\mathbb{R} \setminus \{0\}) \tag{2.23}$$

Here, $\nu(A)$ is known as the Lévy measure of the process and provides a representation of the discontinuous component of $\{X_t\}_{t\geq 0}$ — that is, $\nu(A)$ counts the expected number of jumps whose size belongs to A in a unitary time interval.

Finally, using (2.17) one can show that a Lévy measure ν integrates $(|x|^2 \wedge 1)$:

$$\int_{\mathbb{R}} (|x|^2 \wedge 1)\nu(dx) < \infty \tag{2.24}$$

where $(a \wedge b)$ denotes the minimum between a and b.

2.2 | Lévy-Itô Decomposition

Given a càdlàg process $\{Y_t\}_{t\geq 0}$ and an arithmetic Brownian motion $\{\gamma t + W_t\}_{t\geq 0}$ independent of $\{Y_t\}_{t\geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the process described by $X_t = \gamma t + W_t + Y_t$ defines a Lévy process and can be thus decomposed as:

$$X_t = \gamma t + W_t + \sum\nolimits_{s\in[0,t]} (Y_s - Y_{s-}) = \gamma t + W_t + \int_0^t \int_{\mathbb{R}} y J_Y(ds \times dy) \tag{2.25}$$

One could still define the Lévy measure ν on $\mathcal{B}(\mathbb{R} \setminus \{0\})$ of the process associated to (2.25) following Definition 2.5. However, such measure may not be necesarily finite as the process may exhibit an infinite number of "small" jumps in the time interval considered. This would result in an infinite series, whose convergence relies on some restrictions on the Lévy measure of $\{X_t\}_{t\geq 0}$, giving rise to the well-known *Lévy-Itô decomposition*.

Theorem 2.1. *Let $\{X_t\}_{t\geq 0}$ be a real-valued Lévy process with Lévy measure ν on $\mathcal{B}(\mathbb{R} \setminus \{0\})$ such that $\int_{\mathbb{R}} (|x|^2 \wedge 1)\nu(dx) < \infty$. Then, there exists a $\gamma \in \mathbb{R}$, a Brownian motion $\{W_t\}_{t\geq 0}$ with variance σ^2, and a jump measure J_X on $[0, \infty) \times \mathbb{R}$ with intensity $\nu(dx)\, dt$ such that:*

$$X_t = \gamma t + W_t + \int_0^t \int_{|x|\geq 1} x J_X(ds \times dx) + \lim_{\varepsilon \downarrow 0} \int_0^t \int_{\varepsilon \leq |x| < 1} x \left[J_X(ds \times dx) - \nu(dx)\, ds\right] \tag{2.26}$$

The triplet (γ, σ^2, ν) completely characterises the distribution of $\{X_t\}_{t\geq 0}$, and is therefore known as the characteristic (or Lévy) triplet of the process.

The decomposition in (2.26), first proposed by Lévy (1934) and then formalised by Itô (1941), shows how any Lévy process may be thought of as a combination of a Brownian motion and several independent (compound) Poisson processes. In particular, $\gamma t + W_t$ is included to represent the continuous dynamics of the process paths, whereas the remaining terms account for discontinuity points. The integrability condition imposed on ν guarantees the process exhibits a finite number of jumps whose absolute magnitude is greater than one, thereby ensuring the finiteness of the first integral in (2.26). Finally, for any $\varepsilon > 0$ one may also define an additional compound Poisson process where the magnitude of jumps lies in $[\varepsilon, 1)$; here, however, the jump measure should be replaced by its compensated analogous $\tilde{J}_X := (J_X - \nu(dx)\, dt)$ in order to guarantee a bounded sum.

As already mentioned, given a Lévy process $\{X_t\}_{t\geq 0}$ there exists a continuous map $\psi \colon \mathbb{R} \to \mathbb{C}$, known as the characteristic exponent of $\{X_t\}_{t\geq 0}$, such that:

$$\Phi_{X_t}(z) = \mathbb{E}[\exp(izX_t)] = \exp(t\psi(z)) \quad \forall z \in \mathbb{R} \tag{2.27}$$

A result known as *Lévy-Khinchin representation* shows that such characteristic exponent is completely determined by the Lévy triplet (γ, σ^2, ν) associated to the process.

Theorem 2.2. *Let $\{X_t\}_{t\geq 0}$ be a real-valued Lévy process with characteristic triplet given by (γ, σ^2, ν). Then, the characteristic function of the process at time t is given by:*

$$\Phi_{X_t}(z) = \mathbb{E}[\exp(izX_t)] = \exp\left\{t\left[i\gamma z - \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}}\left(e^{izx} - 1 - izx\mathbb{1}_{|x|<1}\right)\nu(dx)\right]\right\} \tag{2.28}$$

for all $z \in \mathbb{R}$.

The indicator $\mathbb{1}_{|x|<1}$ in the characteristic exponent appearing in (2.28) serves as ac *truncation function*, accounting for jumps in the process whose magnitude is larger than one. As a result, if the Lévy measure also satisfies $\int_{|x|\geq 1}|x|\nu(dx) < \infty$ there is no need to introduce truncations through $\mathbb{1}_{|x|<1}$ in the integrand of (2.28).

Both the decomposition in (2.26) and the representation in (2.28) may be greatly simplified if the paths of the process exhibit *finite variation*.

Definition 2.6. *The total variation of a function $f\colon [a,b] \to \mathbb{R}$ is given by*

$$TV(f) = \sup\sum_{i=1}^{n}|f(t_i) - f(t_{i-1})| \tag{2.29}$$

where the supremum runs over all the partitions $a = t_0 < \cdots < t_n = b$ of the domain on which f is defined. Every increasing or decreasing function exhibits finite variation and every function of finite variation is a difference of two increasing functions (Cont and Tankov, 2003).

One can show that a Lévy process exhibits finite (total) variation if and only if

$$\sigma^2 = 0 \quad \text{and} \quad \int_{|x|\leq 1}|x|\nu(dx) < \infty \tag{2.30}$$

It is indeed well-known that Brownian motion paths have infinite variation and accumulate quadratic variation at the rate of one unit per time. Hence, the absence of a purely diffusive component and the restriction on the Lévy measure for small jumps ensure the process is of finite variation. If this is the case, (2.26) and (2.28) may be rewritten as

$$X_t = \left(\gamma - \int_{|x|\leq 1} x\nu(dx)\right)t + \int_0^t\int_{\mathbb{R}} xJ_X(ds\times dx) \tag{2.31}$$

$$\Phi_{X_t}(z) = \mathbb{E}[\exp(izX_t)] = \exp\left\{t\left[i\left(\gamma - \int_{|x|\leq 1} x\nu(dx)\right)z + \int_{\mathbb{R}}(e^{izx} - 1)\nu(dx)\right]\right\} \tag{2.32}$$

For option pricing purposes, it is essential to mention that any Lévy process satisfies the *strong Markov property* — that is, it is homogenous in both time and space. In particular, defining the transition probability as in Cont and Tankov (2003):

$$P_{s,t}(x,B) := \mathbb{P}\{X_t \in B \mid X_s = x\} \quad \forall B \in \mathcal{B}(\mathbb{R}) \tag{2.33}$$

Then, it holds that

$$P_{s,t}(x,B) = P_{0,t-s}(0,B-X) \tag{2.34}$$

It shall be mentioned that such property of the transition kernel completely characterises Lévy processes, the only homogenous processes in both time and space. Finally, Sato (1999) shows the following with respect to Lévy processes and martingales.

Definition 2.7. *Given a real-valued stochastic process $\{X_t\}_{t\geq 0}$ characterised by independent increments, one has that:*

- $\left\{e^{izX_t}/\,\mathbb{E}\left[e^{izX_t}\right]\right\}_{t\geq 0}$ *defines a martingale for all $z \in \mathbb{R}$.*
- *If the expectation $\mathbb{E}\left[e^{zX_t}\right]$ is finite for some $z \in \mathbb{R}$ and for all t in the interval considered, then $\left\{e^{zX_t}/\,\mathbb{E}\left[e^{zX_t}\right]\right\}_{t\geq 0}$ defines a martingale.*
- *If the process has finite expectation for all $t \geq 0$, then $\{M_t\}_{t\geq 0} = \{X_t - \mathbb{E}[X_t]\}_{t\geq 0}$ defines a martingale with independent increments.*
- *If the process has finite variance for all $t \geq 0$ and M is as above, then the process given by $\left\{M_t^2 - \mathbb{E}[M_t^2]\right\}_{t\geq 0}$ defines a martingale.*

If $\{X_t\}_{t\geq 0}$ verifies the properties of a Lévy process, then for the processes listed above to define martingales it is enough for the central moments mentioned to be finite for at least one time t.

The Lévy-Itô decomposition in (2.26), coupled with Definition 2.7, may be used to draw the following conclusions about a Lévy process with characteristic triplet (γ, σ^2, ν):

- The process $\{X_t\}_{t\geq 0}$ is a martingale if and only if $\int_{|x|\geq 1} |x|\nu(dx)$ is finite, and
$$\gamma + \int_{|x|\geq 1} x\nu(dx) = 0 \tag{2.35}$$
- The process $\{\exp(X_t)\}_{t\geq 0}$ is a martingale if and only if $\int_{|x|\geq 1} e^x \nu(dx)$ is finite, and
$$\gamma + \frac{\sigma^2}{2} + \int_{\mathbb{R}} (e^x - 1 - x\mathbb{1}_{|x|\leq 1})\nu(dx) = 0 \tag{2.36}$$

Finally, one may distinguish between Lévy processes with a finite number of jumps in the realised paths and those with an infinite number of discontinuities in a given time interval. The first scenario includes the so-called finite-activity, or *jump-diffusion* models; here, the expected evolution of the underlying is dictated by a purely diffusive process, such as a Brownian motion with drift, while the points of discontinuity represent out-of-the-ordinary events such as market crashes and booms. Alternatively, if the underlying dynamics were to be modelled by a process extracted from the second scenario, we would speak of a model with *infinite activity*. In this case, the inclusion of a diffusive component is redundant as the process can already describe the underlying dynamics accurately (Carr et al., 2002).

2.3 | Jump-Diffusion

The first author to take account of discontinuities in price paths is was Merton (1976). In his derivation, a (compound) Poisson process, independent of the diffusive component, is included into the dynamics of the underlying.

Assume that, in an infinitesimal time increment dt, the price jumps from S_t to $Y_t S_t$; then, $Y_t - 1$ represents the percentage size of the discontinuity — that is, $Y_t = 0.25$ entails that the stock will trade at 25% of its value before the jump. Regarding the frequency of arrivals of such discontinuities, Merton (1976) assumed it to be dictated by a Poisson process $\{N_t\}_{t\geq 0}$ with intensity λ. Hence, in an infinitesimal increment dt the discontinuous component is given by $(S_t - Y_t S_t)\, dN_t = S_t(Y_t - 1)\, dN_t$, or, in terms of instantaneous rate of return, $(Y_t - 1)\, dN_t$. Moreover, Merton (1976) imposed a log-normal distribution for the jump size Y_t — that is, $\log(Y_t) \sim \mathrm{N}(\alpha, \beta^2)$. Equivalently,

$$Y_t \sim \mathrm{LogN}\left(\exp\left(\alpha + \frac{\beta^2}{2}\right), \exp(2\alpha + \beta^2)(\exp(\beta^2) - 1)\right) \tag{2.37}$$

The discontinuous component of the instantaneous rate of return does still entail a predictable part, which should be subtracted from the drift of the Brownian motion to preserve the randomness of the process. In particular, one has:

$$\kappa := \mathbb{E}[Y_t - 1] = \exp\left(\alpha + \frac{\beta^2}{2}\right) - 1 \tag{2.38}$$

Since $\{Y_t\}_{t\geq 0}$ is independent of the Poisson process $\{N_t\}_{t\geq 0}$, one then has:

$$\mathbb{E}[(Y_t - 1)\, dN_t] = \mathbb{E}[Y_t - 1]\mathbb{E}[dN_t] = \lambda\kappa\, dt \tag{2.39}$$

Hence, the instantaneous return on the underlying is given by the following SDE:

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t + (Y_t - 1)\, dN_t - \lambda\kappa\, dt = (\mu - \lambda\kappa)\, dt + \sigma\, dW_t + (Y_t - 1)\, dN_t \tag{2.40}$$

Finally, one has to allow for more than one jump to occur at each time. Since the Poisson counter dictates the number of discontinuities, (2.40) becomes:

$$\frac{dS_t}{S_t} = (\mu - \lambda\kappa)\, dt + \sigma\, dW_t + \left(\prod_{j=1}^{dN_t} Y_j - 1\right) \tag{2.41}$$

To reach a solution of (2.41), however, one needs to rely on an extension of the Itô-Doeblin formula capable of accounting for discontinuities in the process paths. In particular, let

$\{X_t\}_{t\geq 0}$ be a *jump-diffusion* (JD) process, and consider a jump of size ΔX_t; then, the dynamics of a smooth function $f(X_t, t)$ of class $C^2(\mathbb{R})$ are given in Cont and Tankov (2003):

$$df_t = \left[\frac{\partial f}{\partial t} + \frac{\partial f}{\partial X}\mu_t + \frac{1}{2}\frac{\partial^2 f}{\partial X}\sigma_t^2\right] dt + \frac{\partial f}{\partial X}\sigma_t\, dW_t + [f(X_{t-} + \Delta X_t) - f(X_{t-})] \tag{2.42}$$

where $f(X_{t-})$ represents the value of the function just before the jump. Equivalently,

$$df_t = \frac{\partial f}{\partial t}\, dt + \frac{\partial f}{\partial X}\, dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}(dX_t)^2 + [f(X_{t-} + \Delta X_t) - f(X_{t-})] \tag{2.43}$$

Just as in the case of the SDE defining geometric Brownian motion, a clever choice for f is given by the natural logarithm of the process — that is, $f(X_t, t) = f(X_t) = \log(X_t)$:

$$\begin{aligned} d\log(S_t) &= \frac{1}{S_t}\, dS_t - \frac{1}{2S_t^2}(dS_t)^2 + \log\left(S_t \prod_{j=1}^{dN_t} Y_j\right) - \log(S_t) \\ &= (\mu - \lambda\kappa)\, dt + \sigma\, dW_t - \frac{\sigma^2}{2}\, dt + \log\left(\prod_{j=1}^{dN_t} Y_j\right) \\ &= \left(\mu - \lambda\kappa - \frac{\sigma^2}{2}\right) dt + \sigma\, dW_t + \sum_{j=1}^{dN_t} \log(Y_j) \end{aligned} \tag{2.44}$$

Finally, integrating (2.44) on the interval $[0, t]$ one has:

$$\begin{aligned} &\int_0^t d\log(S_t) = \int_0^t \left(\mu - \lambda\kappa - \frac{\sigma^2}{2}\right) dt + \int_0^t \sigma\, dW_t + \int_0^t \sum_{j=1}^{dN_t} \log(Y_j) \\ \Longrightarrow\ &\log(S_t) - \log(S_0) = \left(\mu - \lambda\kappa - \frac{\sigma^2}{2}\right) t + \sigma W_t + \sum_{j=1}^{N_t} \log(Y_j) \\ \Longrightarrow\ &S_t = S_0 \exp\left[\left(\mu - \lambda\kappa - \frac{\sigma^2}{2}\right) t + \sigma W_t + \sum_{j=1}^{N_t} \log(Y_j)\right] \end{aligned} \tag{2.45}$$

Hence, the model proposed by Merton (1976) specifies an *exponential Lévy process* of the form $\{S_t\}_{t\geq 0} = S_0 \exp(\{X_t\}_{t\geq 0})$ for the dynamics of the underlying under $\mathbb{P}$, where:

$$X_t = \left(\mu - \lambda\kappa - \frac{\sigma^2}{2}\right) t + \sigma W_t + \sum_{j=1}^{N_t} \log(Y_j) \tag{2.46}$$

A sample path generated by a jump-diffusion process is presented in Figure 2.4. The inclusion of jumps in the underlying price process clearly leads to much higher volatility of the latter. Consequently, the distribution of returns will inevitably be characterised by heavy tails and thus be more representative of the financial reality.

The conclusion of the first chapter mentions how such a configuration for the dynamics of the underlying leads to an incomplete market. In other words, an agent cannot hedge his position on an option against the risks dictated by both the diffusive and discontinuous components by relying exclusively on the underlying and a risk-free asset.

In a complete market, an option's value equals the cost of the trading strategy that replicates its cash flows. In an incomplete market, however, an option's price will be composed of the cost of hedging and a risk premium dictated by the writer of the contract to cover his unhedgeable risk. Still, this second component may be close to zero in well-functioning, competitive derivative markets, and, most importantly, incomplete frameworks are much more resemblant to lifelike option trading. If markets were complete, hedging an option would be coping with the Δ-risk by trading the underlying asset. Once one neglects market imperfections, such risk can be hedged away completely. At this point, one could argue that reducing the uncertainty attached to an options contract to Δ-risk is a bolder assumption than neglecting bid-ask spreads, and this is indeed the case. The fundamental exposure in the evolution of an option's value is, in fact, mainly due to variables such as Γ-risk and ν-risk rather than distortions arising from market inefficiencies.

As shown in Cont and Tankov (2003), one can always derive an equivalent martingale measure for an exponential Lévy model through a procedure similar to the drift change of Brownian motion, known as the *Esscher transform*. While exponential Lévy models are arbitrage-free, by Theorem 1.3, one knows that there will be more than one risk-neutral measure under which no arbitrage opportunities are created.

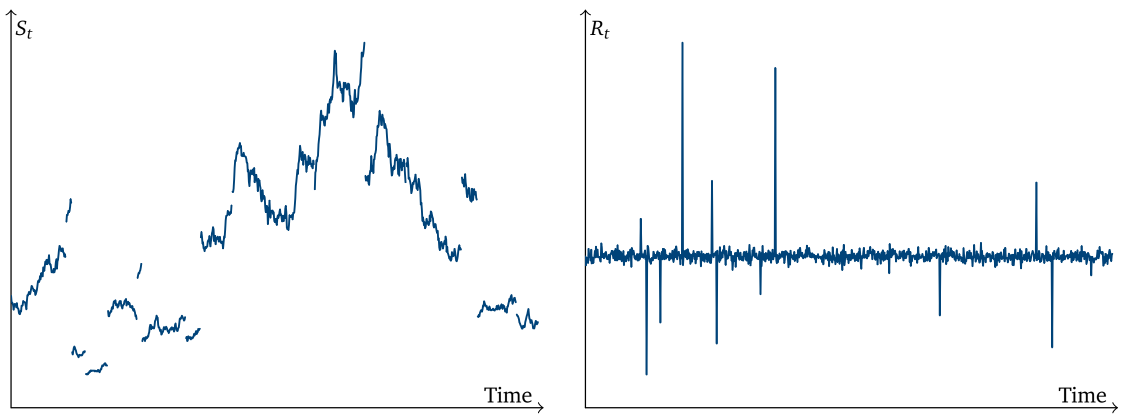


Figure 2.4 Left: Sample path of a jump-diffusion process over an interval of ten years, with $\mu = 0.05$, $\sigma = 0.20$, $\alpha = 0.00$, $\beta = 0.50$, and $\lambda = 1.5$. The last parameter entails that one can expect three jumps every two years. Right: Logarithmic returns associated to the realisation shown on the left.

Although a JD process for the underlying leads to an incomplete market, in his derivation, Merton (1976) dramatically simplifies the analysis by assuming that the jump risk is diversifiable and, as such, should bear no risk premium. This assumption is hardly compatible with empirical observations since large price movements identifiable as jumps can safely occur in highly diversified indices such as those shown in Figure 2.1. Nevertheless, this makes it possible to obtain the risk-neutral dynamics of the underlying asset in the same way as one would with a purely diffusive process — that is, by imposing the drift of the Brownian motion equal to the risk-free rate without altering the discontinuous component of the process. In other words, the log-price at time t is given by:

$$X_t = \left(r - \lambda\kappa - \frac{\sigma^2}{2}\right) t + \sigma W_t^{\mathbb{Q}} + \sum_{j=1}^{N_t} \log(Y_j) \tag{2.47}$$

where $W_t^{\mathbb{Q}}$ is a (standard) Brownian motion under the equivalent martingale measure proposed by Merton (1976). By conditioning on the Poisson process being equal to $N_t = n$, one can express (2.47) in terms of a single source of randomness — namely, the Brownian motion under $\mathbb{Q}$. Letting $N_t = n$, (2.47) becomes:

$$X_t = \left(r - \lambda\kappa - \frac{\sigma^2}{2}\right) t + \sigma W_t^{\mathbb{Q}} + \sum_{j=1}^{n} \log(Y_j) \tag{2.48}$$

Since the (log-)jumps are independent and identically (normally) distributed, and the Poisson process is independent of the Brownian motion, one has that

$$\sigma W_t^{\mathbb{Q}} + \sum_{j=1}^{n} \log(Y_j) = n\alpha + \underbrace{\sqrt{\sigma^2 + \frac{n\beta^2}{t}}}_{:=\,\sigma_n} \sqrt{t} Z = n\alpha + \sigma_n W_t \sim \mathrm{N}\left(n\alpha, \sigma_n^2 t\right) \tag{2.49}$$

where $Z \sim \mathrm{N}(0, 1)$. Hence, one can rewrite (2.47) as

$$\begin{aligned}
X_t &= \left(r - \lambda\kappa - \frac{\sigma^2}{2}\right) t + n\alpha + \sqrt{\sigma^2 + \frac{n\beta^2}{t}} W_t \\
&= \left(r - \lambda\kappa - \frac{\sigma^2}{2} - \frac{n\beta^2}{2t} + \frac{n\beta^2}{2t}\right) t + n\alpha + \sqrt{\sigma^2 + \frac{n\beta^2}{t}} W_t \\
&= \left(r - \lambda\kappa - \frac{\sigma_n^2}{2} + \frac{n\beta^2}{2t}\right) t + n\alpha + \sigma_n W_t = \left(\frac{n\beta^2}{2t} - \lambda\kappa\right) t + n\alpha + \left(r - \frac{\sigma_n^2}{2}\right) t + \sigma_n W_t
\end{aligned} \tag{2.50}$$

Therefore, the SDE of the underlying in a jump-diffusion model can be expressed as

$$S_t = S_0 \exp\left[\left(\frac{n\beta^2}{2t} - \lambda\kappa\right) t + n\alpha + \left(r - \frac{\sigma_n^2}{2}\right) t + \sigma_n W_t\right] = S_0^n \left[\left(r - \frac{\sigma_n^2}{2}\right) t + \sigma_n W_t\right] \tag{2.51}$$

where

$$S_n := S_0 \exp\left[\left(\frac{n\beta^2}{2t} - \lambda\kappa\right)t + n\alpha\right] \quad \text{and} \quad \sigma_n^2 := \sigma^2 + \frac{n\beta^2}{t} \tag{2.52}$$

A call option written on the stock with dynamics given by (2.51) is associated with a premium that can be calculated through (2.4) with the appropriate changes to spot price and volatility. By the law of total probability, one has to account for all possible numbers of jumps in the Poisson process; the price of a call option is then given by:

$$V_t = \sum_{n=0}^{\infty} \mathbb{Q}(N_t = n) V_t^{BS}(S_n, \sigma_n) = \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} V_t^{BS}(S_n, \sigma_n) \tag{2.53}$$

The infinite series above converges exponentially and can be truncated after few values for accurate results (Cont and Tankov, 2003). As an exponential Lévy process, the jump-diffusion proposed by Merton (1976) is characterised by a Lévy triplet (γ, σ^2, ν), where

$$\gamma = r - \frac{\sigma^2}{2} - \lambda\kappa \quad \text{and} \quad \nu(dx) = \lambda f(dx) = \frac{\lambda}{\beta\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{dx - \alpha}{\beta}\right)^2\right] \tag{2.54}$$

Therefore, the characteristic exponent of the log-price process $\{X_t\}_{t\geq 0}$ is:

$$\psi(z) = i\gamma z - \frac{\sigma^2 z^2}{2} + \lambda\left[\exp\left(i\alpha z - \frac{\beta^2 z^2}{2}\right) - 1\right] \tag{2.55}$$

Because of the stationary and independent increments, the characteristic function is then linear in time with respect to (2.55). As noted in Chapter 1, one can compute the k-th moment of a RV by differentiating the characteristic function k times and evaluate it at the

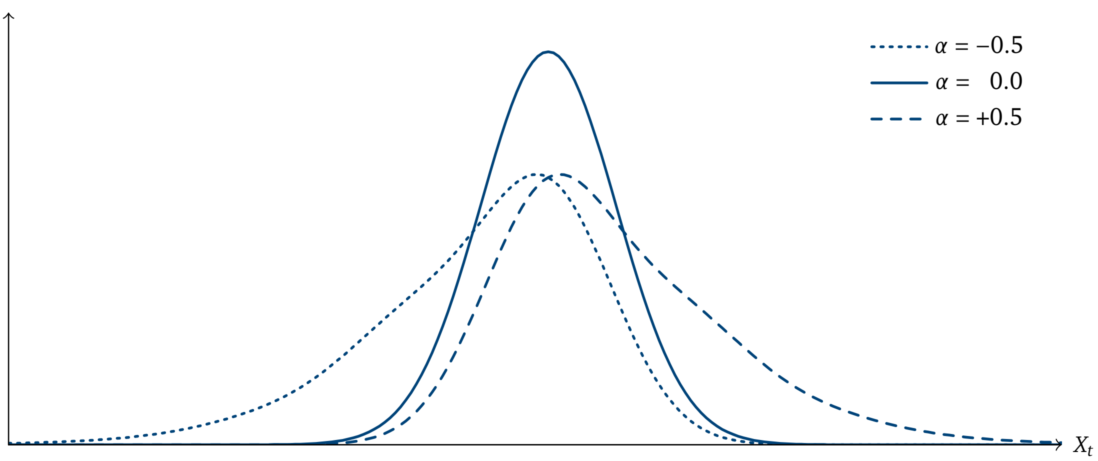


Figure 2.5 Density of the log-price X_t in one year under a jump-diffusion model with log-normally distributed jumps. The parameters kept fixed are $\mu = 0.05$, $\sigma = 0.20$, $\lambda = 0.5$, and $\beta = 0.10$.

point $z = 0$. Doing so for the model proposed by Merton (1976), one obtains:

$$\mathbb{E}[X_t] = t(\gamma + \alpha\lambda) \tag{2.56}$$

$$\mathbb{V}[X_t] = t(\sigma^2 + \lambda\alpha^2 + \lambda\beta^2) \tag{2.57}$$

$$\mathbb{E}[X_t^3] = t\lambda(3\alpha\beta^2 + \alpha^3) \tag{2.58}$$

$$\mathbb{E}[X_t^4] = t\lambda(3\beta^3 + 6\alpha^2\beta^2 + \alpha^4) \tag{2.59}$$

Since t, λ, and β all belong to $\mathbb{R}_0^+$, it is clear from (2.58) that the skewness of the distribution is entirely determined by the sign of the mean jump size α. Such observation is confirmed once the density of the log-price X_t is plotted as in Figure 2.5. Similarly, the intensity of jump arrivals λ is positively related with the fourth moment of the process.

Finally, Figure 2.6 shows the implied volatility surface $\Sigma_t\colon (K,T) \to \Sigma_t(K,T)$ in the Merton (1976) model for different values of the JD process parameters. In line with the observations above, changing the (log-)jump size α affects the shape of the volatility skew, especially at short maturities. On the other hand, caeteris paribus, increasing the jump intensity λ leads to a more pronounced volatility smile. Regardless of the parameter set $\Theta = (\mu, \sigma, \lambda, \alpha, \beta)$ chosen, however, the JD model with log-normally distributed jumps does not adequately reproduce the smile at longer maturities.

2.4 | Variance-Gamma

As already mentioned, one does not need to introduce a purely diffusive component to reach an accurate representation of the evolution of asset prices. This section is devoted to describing one of such pure jump processes — namely, the *variance-gamma* (VG) model proposed by Madan et al. (1998) and extending the work of Madan and Seneta (1990) and Madan and Milne (1991). Although such a model has no diffusive part, it can be derived through a procedure known as Brownian subordination — the VG model may be interpreted as a Brownian motion evaluated on a different time scale dictated by a gamma process. Here, the gamma process is called *subordinator*: a Lévy process with almost surely increasing paths popularised by Clark (1973) for financial applications. This feature allows one to interpret the subordinator as a "distorted" time scale for the increments of some other Lévy process, as shown by the following theorem (Cont and Tankov, 2003).

Theorem 2.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{X_t\}_{t\geq 0}$ a $\mathbb{R}$-valued Lévy process with characteristic exponent $\psi_X(z)$ and triplet (γ, σ^2, ν). Let $\{S_t\}_{t\geq 0}$ be a subordinator with Laplace exponent $l(z)$ and triplet $(b, 0, \rho)$. Then, the process $\{Y_t\}_{t\geq 0}$ defined as:*

$$Y_t(\omega) = X_{S_t(\omega)}(\omega) \tag{2.60}$$

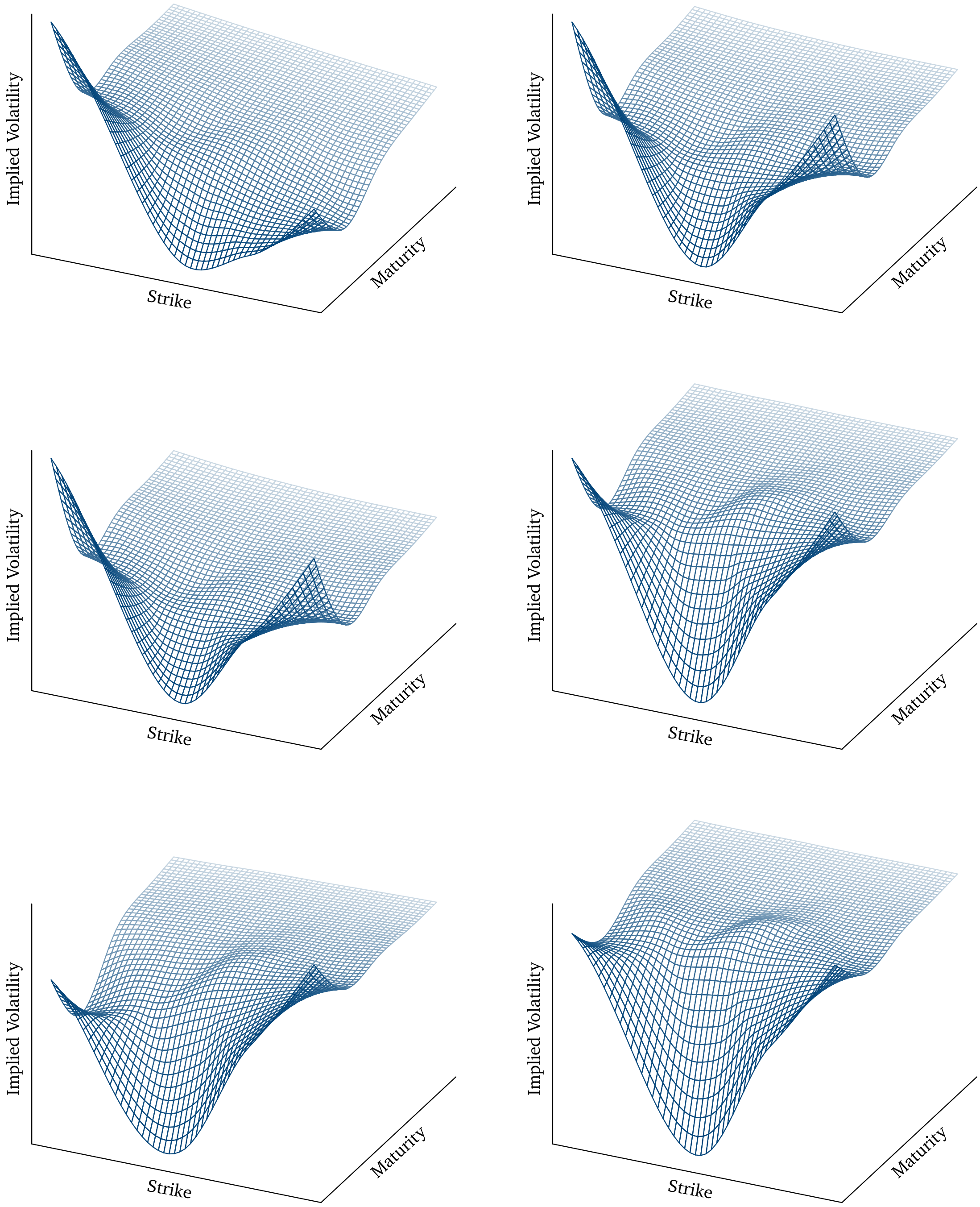


Figure 2.6 Volatility surfaces generated by the Merton (1976) model under $\mathbb{Q}$. The left column shows how the volatility profile changes for different values of the (log-)jump size — that is, from top to bottom $\alpha = -0.50$, 0.00, and 0.50. The right column does the same for three distinct values of the jump intensity λ — that is, from top to bottom $\lambda = 1.00$, 3.00, and 5.00. The volatility of the price process and that of the (log)-jumps are kept constant at $\sigma = 0.20$ and $\beta = 0.50$.

is still a Lévy process with characteristic exponent given by:

$$\psi_Y(z) = l(\psi_X(z)) \quad \text{where } l(z) = bz + \int_0^\infty (e^{zx} - 1)\rho(dx) \tag{2.61}$$

so that the Fourier transform of Y_t at a given time t is found through the composition of the characteristic exponent of X_t with the Laplace exponent of S_t. Finally, the characteristic triplet $(\gamma_Y, \sigma_Y^2, \nu_Y)$ of $\{Y_t\}_{t\geq 0}$ is given by:

$$\sigma_Y^2 = b\sigma^2 \tag{2.62}$$

$$\gamma_Y = b\gamma + \int_0^\infty \rho(ds) \int_0^\infty x \mathbb{1}_{|x|\leq 1} f_{X_s}(dx) \tag{2.63}$$

$$\nu_Y(B) = b\nu(B) + \int_0^\infty f_{X_s}(B)\rho(ds) \quad \forall B \in \mathcal{B}(\mathbb{R}) \tag{2.64}$$

where f_{X_s} is the density of X_t at time $t = s$.

In the framework presented in Theorem 2.3, $\{Y_t\}_{t\geq 0}$ is said *subordinate* to $\{X_t\}_{t\geq 0}$. The VG process described by Madan et al. (1998) is simply a Brownian motion subordinate to a gamma process. Here, rather than prescribing a diffusion process for the underlying, realistic reproductions of market activity are obtained by allowing an infinite number of (small) jumps in any time interval. From Section 2.2, one knows that the absence of a Brownian component ensures the process has finite variation; by Definition 2.6, this entails the VG process can be expressed as the difference of two increasing processes. As shown by Madan et al. (1998), these two increasing processes are themselves gamma processes with mean rates μ_p and μ_n accounting for gains and losses, respectively:

$$X_t = \gamma_t(\mu_p, \mu_p^2\nu) - \gamma_t(\mu_n, \mu_n^2\nu) \tag{2.65}$$

where the mean rates μ_p and μ_n are defined as:

$$\mu_p := \frac{1}{2}\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}} + \frac{\theta}{2} \quad \text{and} \quad \mu_n := \frac{1}{2}\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}} - \frac{\theta}{2} \tag{2.66}$$

Alternatively, letting $\{Z_t\}_{t\geq 0}$ be a Brownian motion with drift θ, and using the notation of (2.60), the variance-gamma process $\{X_t\}_{t\geq 0}$ is defined as:

$$X_t(\omega) = Z_{\gamma_t(1,\nu)}(\omega) = \theta\gamma_t(1, \nu) + \sigma W_{\gamma_t(1,\nu)} \tag{2.67}$$

where $\gamma_t(1, \nu)$ is a gamma process with unit mean rate — that is, a process with independent increments over non-overlapping intervals $[t, t+h]$ with density given by:

$$\left(\frac{1}{\nu}\right)^{h/\nu} \frac{x^{(h/\nu)-1} \exp\left(-x/\nu\right)}{\Gamma\left(h/\nu\right)} \quad \forall x > 0 \tag{2.68}$$

where $t+h-t=h$ is the interval length and $\Gamma(\cdot)$ the Gamma function. The representation of the VG process as a subordinate to a gamma process implies that the density of X_t is Normal conditionally on the realisation of the time-change dictated by $\gamma_t(1,\nu)$:

$$\begin{aligned} f_{X_t}(x) &= \int_{\mathbb{R}} f_{X_t,\gamma_t}(x,g)\,dg = \int_{\mathbb{R}} f_{X_t|\gamma_t}(x \mid g) f_{\gamma_t}(g)\,dg \\ &= \int_0^\infty \frac{1}{\sigma\sqrt{2\pi g}} \exp\left[-\frac{(x-\theta g)^2}{2\sigma^2 g}\right] \left(\frac{1}{\nu}\right)^{t/\nu} \frac{g^{(t/\nu)-1}\exp\left(-g/\nu\right)}{\Gamma\left(t/\nu\right)}\,dg \\ &= \left(\frac{1}{\nu}\right)^{t/\nu} \frac{2\exp(\theta x/\sigma^2)}{\sigma\sqrt{2\pi}\Gamma(t/\nu)} \left(\frac{x^2}{2\sigma^2/\nu+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}} \boldsymbol{K}_{\frac{t}{\nu}-\frac{1}{2}}\left[\frac{1}{\sigma^2}\sqrt{x^2\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}\right] \end{aligned} \tag{2.69}$$

where $\boldsymbol{K}$ is a modified Bessel function of the second kind defined in Madan et al. (1998). Following a similar reasoning — that is, conditioning on the realisation of the gamma process — one sees that the characteristic exponent $\psi(z)$ of $\{X_t\}_{t\geq 0}$ is:

$$\psi(z) = -\frac{1}{\nu}\log\left(1 - i\nu\theta z + \frac{\sigma^2 z^2 \nu}{2}\right) \tag{2.70}$$

As (2.67) defines a Lévy process, its characteristic function is linear in time t and equal to:

$$\Phi_t(z) = \exp\left\{t\left[-\frac{1}{\nu}\log\left(1 - i\nu\theta z + \frac{\sigma^2 z^2 \nu}{2}\right)\right]\right\} = \left(1 - i\nu\theta z + \frac{\sigma^2 z^2 \nu}{2}\right)^{-\frac{t}{\nu}} \tag{2.71}$$

One can find the k-th moment of (2.67) by evaluating the k-th derivative of (2.71) at $z=0$; here, the first four moments of the VG process at time t are given by:

$$\mathbb{E}[X_t] = t\theta \tag{2.72}$$

$$\mathbb{V}[X_t] = t(\sigma^2 + \theta^2\nu) \tag{2.73}$$

$$\mathbb{E}[X_t^3] = t(2\theta^3\nu^2 + 3\sigma^2\theta\nu) \tag{2.74}$$

$$\mathbb{E}[X_t^4] = t(3\sigma^4\nu + 12\sigma^2\nu^2\theta^2 + 6\theta^4\nu^3) \tag{2.75}$$

From (2.74), one sees that $\theta = 0$ implies a symmetric distribution of X_t as in Madan and Seneta (1990) and Madan and Milne (1991); when $\theta \neq 0$, the sign of the asymmetry corresponds to that of θ. Furthermore, dividing (2.75) by the variance of X_t, thereby obtaining its kurtosis, one sees that this equals $3(1+\nu)$; in other words, ν represents the percentage excess kurtosis over the Normal distribution.

As noted at the beginning of this section, the VG process proposed by Madan et al. (1998) does not include any diffusive component. Such statement may be proven by finding the characteristic triplet of the process. Madan et al. (1998) show that the Lévy measure $\overline{\nu}$ is given by the following expression:

$$\overline{\nu}(dx) = \frac{\exp(\theta x/\sigma^2)}{\nu|x|}\exp\left(-\frac{\sqrt{2/\nu+\theta^2/\sigma^2}}{\sigma}|x|\right)dx \tag{2.76}$$

By (2.31), one knows that the Lévy-Itô decomposition of a process with finite variation is:

$$X_t = \left(\gamma - \int_{|x|\le 1} x\nu(dx)\right)t + \int_0^t\int_{\mathbb{R}} xJ_X(ds\times dx) := \overline{\gamma}t + \int_0^t\int_{\mathbb{R}} xJ_X(ds\times dx) \tag{2.77}$$

To find γ it is sufficient to impose the expectation of (2.77) equal to (2.72):

$$\mathbb{E}[X_t] = \mathbb{E}\left[\overline{\gamma}t + \int_0^t\int_{\mathbb{R}} xJ_X(ds\times dx)\right] = t\left(\overline{\gamma} + \int_{\mathbb{R}} x\overline{\nu}(dx)\right) = t\theta \tag{2.78}$$

The integral above may be solved by plugging in the VG Lévy measure (2.76);, let:

$$A = \frac{\theta}{\sigma^2} \quad \text{and} \quad B = \frac{|\theta|}{\sigma^2}\sqrt{1+\frac{2\sigma^2}{\nu\theta^2}} \tag{2.79}$$

Then, the integral involving the Lévy measure has solution:

$$\begin{aligned}\int_{\mathbb{R}} x\overline{\nu}(dx) &= \int_{\mathbb{R}} x\frac{\exp(\theta x/\sigma^2)}{\nu|x|}\exp\left(-\frac{\sqrt{2/\nu+\theta^2/\sigma^2}}{\sigma}|x|\right)dx \\ &= \int_{\mathbb{R}}\frac{x}{\nu|x|}e^{Ax-B|x|} = \int_0^\infty \frac{1}{\nu}e^{(A-B)x} - \int_{-\infty}^0 \frac{1}{\nu}e^{(A+B)x} = \frac{2A}{\nu(B^2-A^2)} = \theta\end{aligned} \tag{2.80}$$

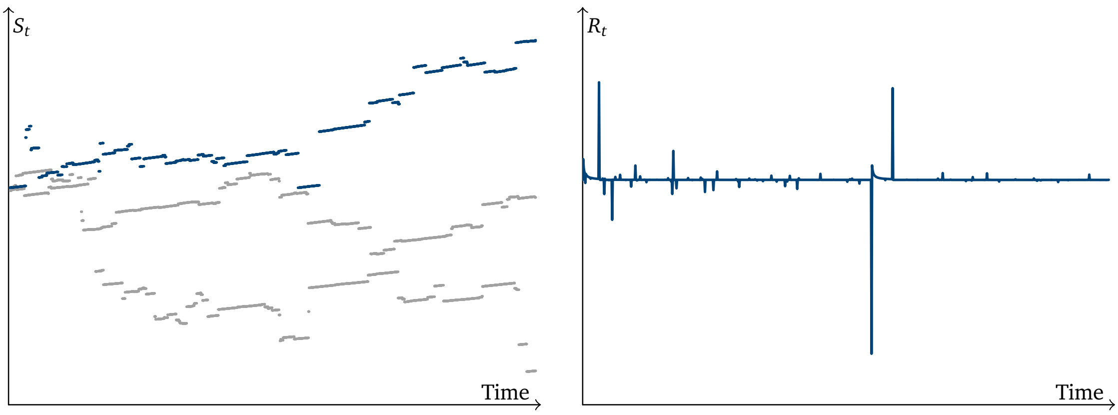


Figure 2.7 Left: Sample paths of a variance-gamma process over an interval of one year, with $\theta = -0.10$, $\sigma = 0.20$, and $\nu = 0.10$. The last parameter entails the distribution of (log-)prices has an excess kurtosis around 10%, or 3.30. Right: Log-returns associated to the (blue) realisation.

As a result, $\overline{\gamma} = 0$ and the Lévy-Itô decomposition for the VG process is simply:

$$X_t = \int_0^t \int_{\mathbb{R}} x J_X(ds \times dx) \tag{2.81}$$

Despite being the result of the subordination of a purely diffusive process — that is, Brownian motion — all the information regarding $\{X_t\}_{t\geq 0}$ is contained in its Lévy measure, making the variance-gamma model a pure jump process. The characteristic triplet is then:

$$\left(\int_{|x|\leq 1} x\overline{\nu}(dx), 0, \nu \right) \tag{2.82}$$

The absence of a diffusive component is evident in Figure 2.7, presenting one sample path of a VG process along with the (logarithmic) returns corresponding to such realisation.

When it comes to option pricing, the framework proposed by Madan et al. (1998) corresponds to an exponential Lévy model — that is, the log-price of the underlying is assumed to follow the dynamics dictated by (2.67). In particular, under an equivalent martingale measure $\mathbb{Q}$, at any time t one has:

$$S_t = S_0 \exp(rt + X_t) \implies \tilde{S}_t = S_0 \exp(X_t) \tag{2.83}$$

For $\tilde{S}_t$ to define a martingale under $\mathbb{Q}$, one needs to subtract a correction term ω which can be found by evaluating the Lévy exponent of the VG process at $z = i^{-1}$:

$$\omega = \psi\left(\frac{1}{i}\right) = -\frac{1}{\nu}\log\left(1 - \nu\theta - \frac{\sigma^2\nu}{2}\right) \tag{2.84}$$

Then, $\tilde{S}_t = S_0 \exp(X_t - \omega t)$ defines a $\mathbb{Q}$-martingale:

$$\mathbb{E}_{\mathbb{Q}}[\tilde{S}_t \mid \mathcal{F}_0] = \mathbb{E}_{\mathbb{Q}}[S_0 \exp(X_t - \omega t) \mid \mathcal{F}_0] = S_0 \tag{2.85}$$

Under these dynamics, Madan et al. (1998) were able to derive a closed-formula for the premium charged on European options. For instance, the price of a call when the underlying follows an exponential VG process is given by:

$$V_t = S_t e^{-r\tau} \Psi\left(\alpha_1, \beta_1, \tau/\nu\right) - K e^{-r\tau} \Psi\left(\alpha_2, \beta_2, \tau/\nu\right) \tag{2.86}$$

where $\Psi(\cdot)$ is defined by Madan et al. (1998) in terms of the modified Bessel function of the second kind and the degenerate hypergeometric function of two variables, and

$$\alpha_1 = \frac{\log(S_t/K) - \omega t}{\sigma}\sqrt{\frac{1-\nu(\theta+\sigma^2/2)}{\nu}} \qquad \beta_1 = \frac{\theta+\sigma^2}{\sigma}\sqrt{\frac{\nu}{1-\nu(\theta+\sigma^2/2)}} \tag{2.87}$$

$$\alpha_2 = \frac{\log(S_t/K) - \omega t}{\sigma}\sqrt{\frac{1}{\nu}} \qquad \beta_2 = \frac{\theta\sqrt{\nu}}{\sigma} \tag{2.88}$$

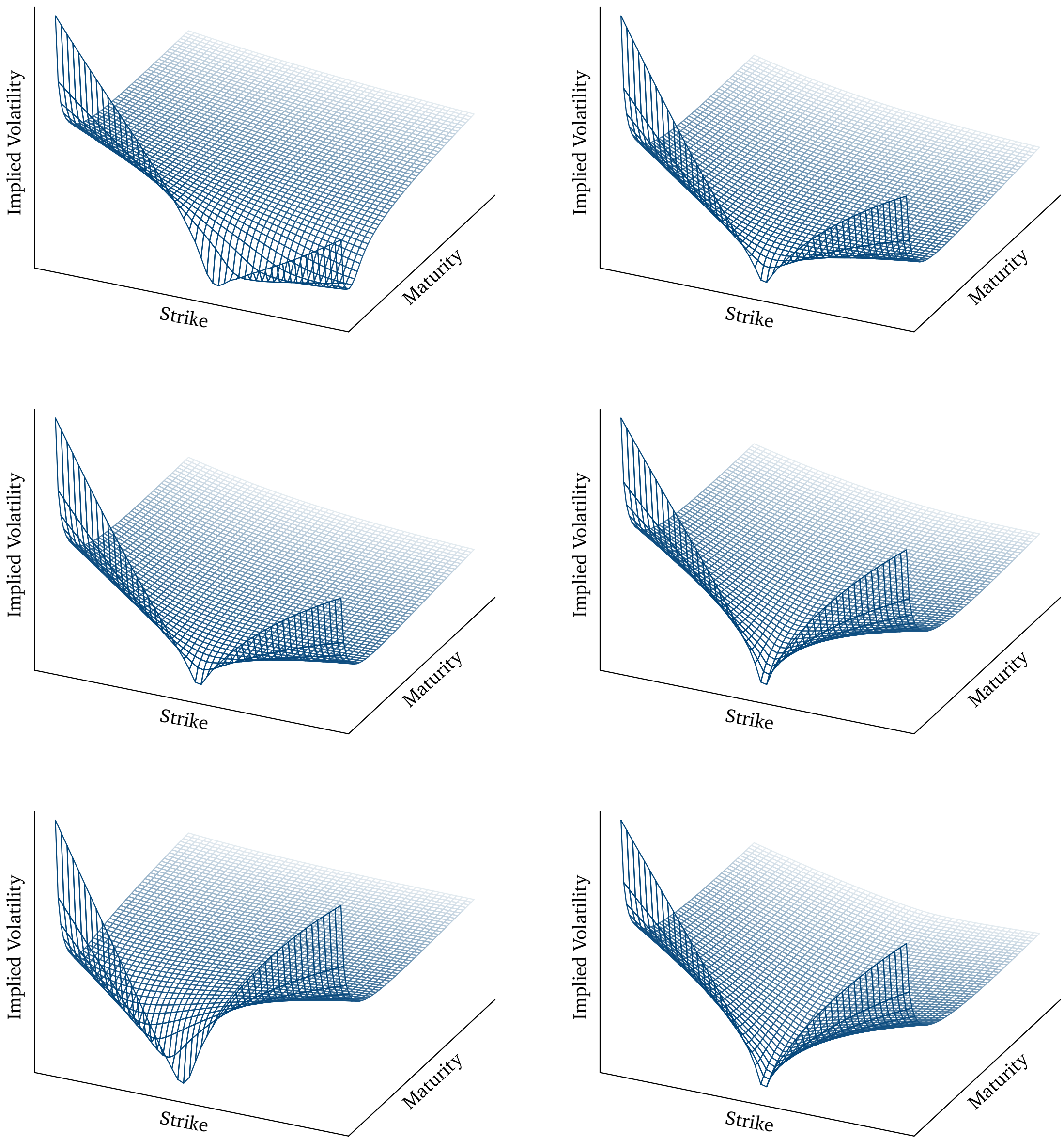


Figure 2.8 Volatility surfaces generated by the variance-gamma model under $\mathbb{Q}$. The left column shows how the volatility profile changes for different values of the Brownian drift — that is, from top to bottom $\theta = -0.50$, 0.00, and 0.50. The right column does the same for three distinct values of the gamma process's rate ν — that is, from top to bottom $\nu = 0.10$, 0.50, and 1.00. The (annualised) volatility of the subordinate Brownian motion is kept constant at $\sigma = 0.20$.

Finally, Figure 2.8 shows the implied volatility surface $\Sigma_t : (K, T) \longrightarrow \Sigma_t(K, T)$ in the VG model of Madan et al. (1998) for different values of θ and ν. In line with the observations above, changing the skewness θ of the (log-)price θ affects the shape of the volatility smile, especially at short maturities. On the other hand, caeteris paribus, increasing the ν parameter leads to a more pronounced volatility skew which, in any case, tends to be more peaked than that arising from a jump-diffusion model. Regardless of the parameter set $\Theta = (\theta, \sigma, \nu)$ chosen, however, the asymmetric VG model proposed by Madan et al. (1998) does not adequately reproduce the surface at longer maturities. In order to obtain a good fit even for contracts with an expiration further out in time, it is essential to relax the assumption of independent increments at the basis of each exponential Lévy process discussed so far. The leptokurtic distributions of returns are, in fact, only a tiny part of the deviations from the BSM model, which include a positive correlation between the magnitude of changes in returns. The next chapter is devoted to discussing such additional stylised facts, along with option pricing models which explicitly account for them.

Chapter 3

Time-Varying Volatility

The previous chapter focused on the relaxation of the assumption of continuous price paths inherent in the geometric Brownian motion employed by Black and Scholes (1973) and Merton (1973). The BSM model, however, builds on another fundamental assumption: namely, a constant volatility for the underlying asset. Just as the exponential Lévy processes discussed in Chapter 2 improve on the BSM framework by allowing for discontinuity points in the price process, *stochastic volatility* (SV) models are based on the idea that volatility itself follows a stochastic process. The additional randomness introduced by these class of models eventually leads to leptokurtic return distributions resembling the empirically observed ones (Gatheral, 2011) such as those presented in Figure 2.1.

Before delving into a discussion of models designed to relax the assumption of constant volatility imposed by Black and Scholes (1973), it is appropriate to proceed with a brief digression on some characteristic properties of the variability of returns; in particular:

- Volatility tends too *cluster* — that is, significant returns follow wide price swings while more minor variations often predict negligible returns. From an econometric point of view, this empirical observation implies that the volatility time series is characterised by the presence of significant *autocorrelation* $\rho(h)$, defined as the correlation of a time series with itself lagged by h periods, for several lags h:

$$\rho(h) := \mathrm{Corr}[X_t, X_{t-h}] = \frac{\mathbb{C}[X_t, X_{t-h}]}{\mathbb{V}[X_t]} \tag{3.1}$$

  The value of such function, for a series made up of n observations, may be computed up to a lag of $h = \lfloor n/4 \rfloor$. Furthermore, one may consider the autocorrelation at a given lag statistically significant should its value lie outside the confidence bands defined by $\pm 1.96/n^{-1/2}$; the latter ($n^{-1/2}$) being the standard deviation of $\rho(h)$ for a so-called *white noise* process, unpredictable by definition. In order to keep things simple at this stage, the (annualised) volatility time series of Apple returns presented in Figure 3.1b is estimated through the method proposed by Rogers and Satchell (1991) based on the opening, closing, minimum, and maximum daily prices of the stock. Specifically, on a given day t the (daily) volatility is calculated as:

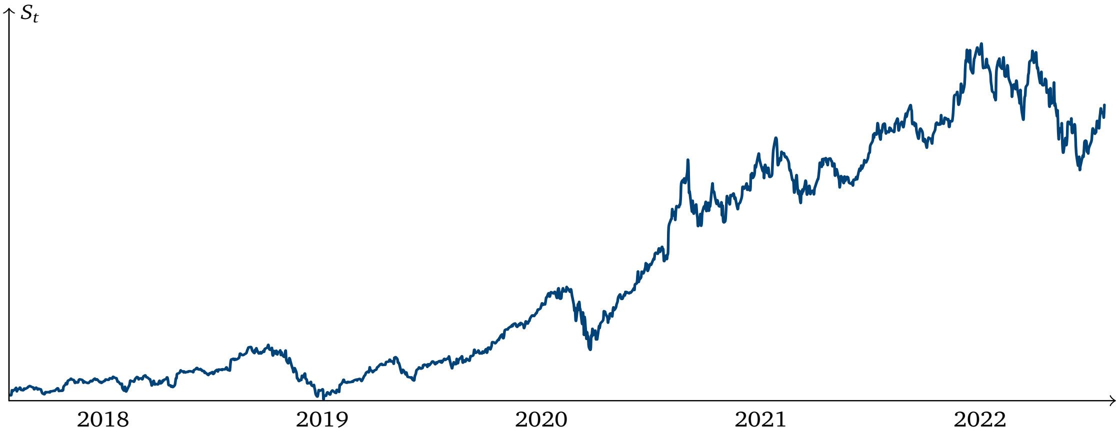
$$v_t = \sqrt{\log\left(\frac{H_t}{S_t}\right)\log\left(\frac{H_t}{O_t}\right) + \log\left(\frac{L_t}{S_t}\right)\log\left(\frac{L_t}{O_t}\right)} \tag{3.2}$$

where H_t, L_t, O_t, and S_t are the maximum, minimum, opening, and closing price, respectively, at time t. The volatility clusters are evident even by just an analysis of the graph presented, while more rigorous evidence is provided by the autocorrelation function for the first 50 lags in Figure 3.1c. It is thus obvious how the observed volatility at day t is (significantly) affected by its value recorded up to forty days earlier. Furthermore, both Figure 3.1b and Figure 3.1c reveal the mean-reverting nature of the volatility process, which indeed tends to fluctuate around a "long-run" value as shown by the (slowly) decaying autocorrelation function.

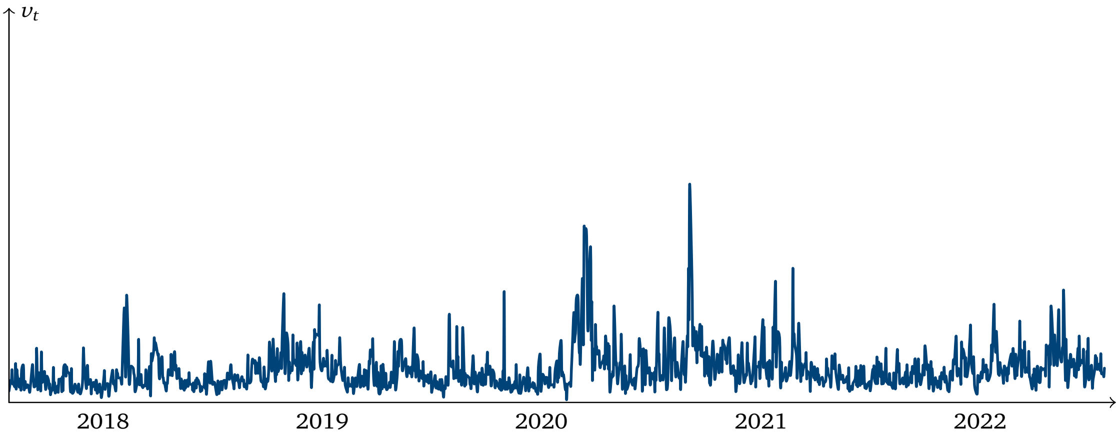
- Volatility and returns share a negative correlation — that is, volatility is larger during market crashes and the latter have a greater impact on the riskiness of an asset than bullish markets. Such phenomenon is known as *leverage effect* and was first pointed out by Black (1976), which pointed out how a decreasing spot price leads to a rise in the leverage ratio of the company thereby increasing its perceived riskiness. Such explanation, however, is neither capable of justifying the observed magnitude of the relationship (Schwert, 1988) nor the presence of the leverage effect in market indices aggregating the performance of several companies. As a result, a number of explanations based on behavioural finance theories and heterogeneous agent models have been proposed over the years; for instance, the presence of *noise traders* subject to herding behaviour might explain the increased volatility in bearish markets (Avramov et al., 2004). The occurrence of a leverage effect is evident from the data presented in Table 3.1, showing the correlation of volatility v_t with positive ($R^+_{t-h} := [R_{t-h}]^+$) and negative ($R^-_{t-h} := [-R_{t-h}]^+$) shocks up to eight lags $h = 1, \ldots, 8$. In particular, it is clear how the correlation between volatility and returns is negative and the fact that negative shocks have a greater impact on v_t than positive shocks.

The frequency with which these properties occur make them true stylised facts of volatility, much like the leptokurtic distributions of returns presented in Chapter 2.

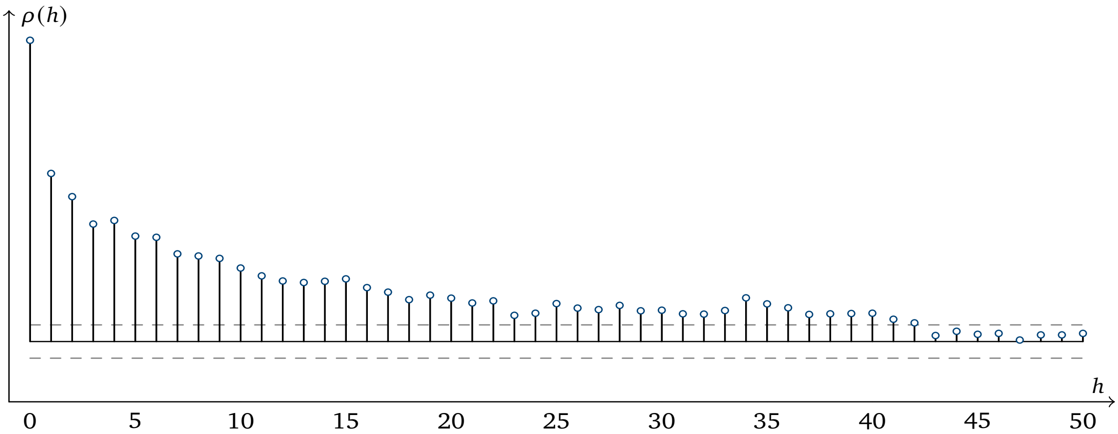
The widespread presence of the empirical observations just discussed clarifies the inadequacy of a constant volatility specification for the underlying price process; an



(a) Time series of the (adjusted) closing prices recorded for the *Apple* stock in the five-year span ranging from August 2017 to August 2022.



(b) Annualised volatility estimated each day through the approach proposed by based on the opening, closing, minimum, and maximum daily prices of the stock.



(c) Sample autocorrelation of the volatility time series plotted in (b) for the first fifty lags, along with the 95% confidence bands under the null hypothesis of a white noise process.

Figure 3.1 Estimated annualised volatility of the *Apple* stock through the method proposed by Rogers and Satchell (1991), along with the corresponding sample autocorrelation function for the first fifty lags. The plots in (b) and (c) make clear how volatility itself is stochastic and characterised by a strong degree of autocorrelation leading to clustering.

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
$\text{Corr}[v_t, R_{t-h}]$	-0.168	-0.127	-0.063	-0.084	-0.082	-0.066	-0.107	-0.043
$\text{Corr}[v_t, R^+_{t-h}]$	0.145	0.092	0.143	0.108	0.120	0.097	0.055	0.087
$\text{Corr}[v_t, R^-_{t-h}]$	0.411	0.293	0.239	0.239	0.248	0.200	0.225	0.153

Table 3.1 Correlation between volatility v_t and log-returns R_{t-h} for various lags h. The negative relationship is evident for all the lags considered and negative shocks R^-_{t-h} have a (much) greater impact on volatility than positive shocks R^+_{t-h}.

immediate improvement would then consist in the introduction of a deterministic function of time in the geometric Brownian motion underpinning the BSM framework. In such a case, the SDE dictating the dynamics of the underlying price under $\mathbb{Q}$ would read:

$$dS_t = rS_t\,dt + \sigma_t S_t\,dW_t^{\mathbb{Q}} \tag{3.3}$$

A straightforward application of the Itô-Doeblin formula with $f_t(S_t) = \log(S_t)$, however, reveals that the terminal stock price is given by:

$$S_T = S_t \exp\left[r\tau - \frac{1}{2}\int_t^T \sigma_s^2\,ds + \int_t^T \sigma_s\,dW_s\right] \tag{3.4}$$

Clearly, such configuration still implies that log-prices are normally distributed:

$$X_T \sim \mathrm{N}\left(X_t + \left(r - \frac{1}{2}\bar{\sigma}^2\right)\tau, \bar{\sigma}^2\tau\right) \tag{3.5}$$

where $\bar{\sigma}^2 := \tau\int_t^T \sigma_s^2\,ds$. It follows that one can simply price options relying on (2.4) by simply replacing the constant σ with the root mean squared volatility $\bar{\sigma}^2$. Furthermore, just allowing for time-dependent volatility keeps the reasoning in a complete market framework, leading to a higher mathematical tractability while providing a poor representation of options markets. A slight generalisation of (3.3) would make volatility dependent on both time and the spot price of the underlying, as discussed in the next section.

3.1 | Local Volatility

The approach proposed by Cox and Ross (1976) is among the first frameworks allowing the volatility of the underlying to depend on both the current time and the spot price of the underlying. In particular, Cox and Ross (1976) focused on the so-called class of *constant elasticity of variance* (CEV) models, wherein volatility is given by $\sigma(t, S_t) = \delta S_t^{\alpha-1}$ with $\delta \in \mathbb{R}$ and $\alpha \in (0, 1)$. Such class of models also allows for option pricing through closed formulas, as shown by Schroder (1989); moreover, such values lead to a skewed volatility surface thereby improving on the constant volatility paradigm of the BSM framework.

Before delving into pricing models that specify a functional form for the process followed by the underlying's volatility, it is worth briefly discussing one of the best-known non-parametric approaches that purport to reproduce the volatility surface (erroneously) predicted constant by the BSM assumptions. In particular, the remainder of this section is concerned with the discussion of the (forward) PDE proposed by Dupire (1994), who started from the assumption of a time- and stock-dependent volatility in the GBM SDE. In other words, based on Dupire's assumptions, (1.18) becomes:

$$dS_t = rS_t\,dt + \sigma(t,S_t)S_t\,dW_t^{\mathbb{Q}} := rS_t\,dt + \tilde{\sigma}S_t\,dW_t^{\mathbb{Q}} \tag{3.6}$$

where it is understood that the dynamics are considered under the equivalent martingale measure, and the underlying is assumed to pay no dividends. In this context, $\sigma(t,S_t)$ is known as *local volatility* as it is proxied by a deterministic function of both the current time and the underlying's spot price. In particular, Dupire (1994) showed the existence of a (local) volatility function $\sigma(t,S_t)$ consistent with the surface observed on a given day.

One derivation of the Dupire's PDE starts from the application of the Itô-Doeblin formula to the payoff of a European call — $f = [S_T - K]^+$ — when the dynamics of the underlying are given by (3.6). As remarked in Chapter 1, however, the application of the Itô-Doeblin formula is subject to the function considered being a smooth (twice) differentiable function; this is clearly not the case for the payoff function of an option, which is intrinsically nonlinear as shown in Figure 3.2 below. Nevertheless, the relevant derivatives may be calculated by resorting to special functions such as the Dirac's measure discussed in Appendix A. In particular, the first two derivatives of the call payoff, with respect to both the underlying and the exercise price, are given by:

$$\frac{\partial}{\partial S_T}[S_T - K]^+ = \frac{\partial}{\partial S_T}\mathbb{1}_{S_T>K}(S_T - K) = \mathbb{1}_{S_T>K} \qquad \frac{\partial}{\partial K}[S_T - K]^+ = -\mathbb{1}_{S_T>K} \tag{3.7}$$

$$\frac{\partial^2}{\partial S_T^2}[S_T - K]^+ = \frac{\partial}{\partial S_T}\mathbb{1}_{S_T>K} = \delta(S_T - K) \qquad \frac{\partial^2}{\partial K^2}[S_T - K]^+ = \delta(S_T - K) \tag{3.8}$$

As a result, applying the Itô-Doeblin formula reveals that the dynamics of the call's payoff are described by the following SDE:

$$df_T = \frac{\partial f}{\partial T}\,dT + \frac{\partial f}{\partial S_T}\,dS_T + \frac{1}{2}\frac{\partial^2 f}{\partial S_T^2}(dS_T)^2 \tag{3.9}$$

$$= \left[\mathbb{1}_{S_T>K} rS_T + \frac{\delta(S_T - K)}{2}\tilde{\sigma}_T^2 S_T^2\right] dT + \mathbb{1}_{S_T>K}\tilde{\sigma}S_T\,dW_T \tag{3.10}$$

The partial derivative with respect to the maturity T of the expectation of df_T is then:

$$\mathbb{E}_{\mathbb{Q}}[df_T] = \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{S_T>K} r S_T + \frac{\delta(S_T - K)}{2}\tilde{\sigma}_T^2 S_T^2\right] dT \tag{3.11}$$

$$\frac{\partial}{\partial T}\mathbb{E}_{\mathbb{Q}}[df_T] = \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{S_T>K} r S_T + \frac{\delta(S_T - K)}{2}\tilde{\sigma}_T^2 S_T^2\right] \tag{3.12}$$

To reach the Dupire's PDE, one should express (3.12) as a function of the call price $V(K,T)$:

$$\frac{\partial}{\partial T}\mathbb{E}_{\mathbb{Q}}[df_T] = r\underbrace{\mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{S_T>K} S_T\right]}_{A} + \frac{1}{2}\underbrace{\mathbb{E}_{\mathbb{Q}}\left[\delta(S_T - K)\tilde{\sigma}_T^2 S_T^2\right]}_{B} \tag{3.13}$$

As presented in Chapter 1 and remarked in Chapter 2, the (first) fundamental theorem of asset pricing ensures the premium charged for a call option is given by the expectation of its terminal payoff under the risk neutral measure — that is, $C(K,T) = e^{-rT}\mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{S_T>K}(S_T - K)\right]$. Evaluating A and B separately by leveraging on (3.7) and (3.8), one sees that:

$$A: \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{S_T>K} S_T\right] = e^{rT}C(K,T) - Ke^{rT}\frac{\partial}{\partial K}C(K,T) \tag{3.14}$$

$$B: \mathbb{E}_{\mathbb{Q}}\left[\delta(S_T - K)\tilde{\sigma}_T^2 S_T^2\right] = K^2\mathbb{E}_{\mathbb{Q}}\left[\delta(S_T - K)\tilde{\sigma}_T^2\right] = K^2\mathbb{E}_{\mathbb{Q}}\left[\tilde{\sigma}_T^2 \mid S_T = K\right]\frac{\partial^2 e^{rT}C(K,T)}{\partial K^2} \tag{3.15}$$

where the first equality in (3.15) follows from the fact that $\delta(S_T - K) \neq 0 \iff S_T = K$. Letting $C^c(K,T) := e^{rT}C(K,T)$, one can then plug (3.14) and (3.15) into (3.13) to reach:

$$\frac{\partial C^c(K,T)}{\partial T} = rC^c(K,T) - rK\frac{\partial C^c(K,T)}{\partial K} + \frac{1}{2}K^2\mathbb{E}_{\mathbb{Q}}\left[\tilde{\sigma}_T^2 \mid S_T = K\right]\frac{\partial^2 C^c(K,T)}{\partial K^2} \tag{3.16}$$

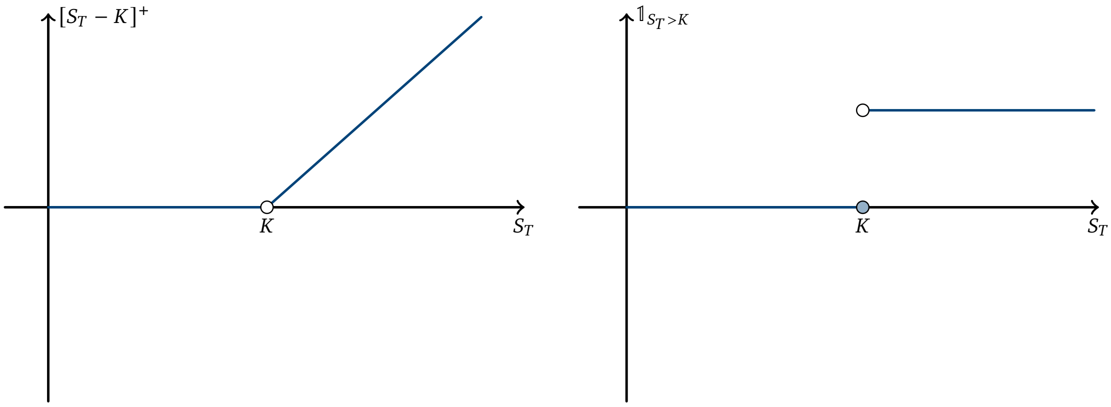


Figure 3.2 Left: Plot of the max function $[S_T - K]^+ := \max\{S_T - K, 0\}$ corresponding to the payoff at the maturity of a European call option. Right: Indicator function corresponding to the first derivative of $[S_T - K]^+$, where differentiation is understood to be in the generalised functions sense.

Finally, "solving" for the local volatility in the equation above the PDE becomes:

$$\mathbb{E}_{\mathbb{Q}}\left[\tilde{\sigma}_T^2 \mid S_T = K\right] = \frac{2}{K^2}\left[\frac{\partial C^c(K,T)}{\partial T} - rC^c(K,T) + rK\frac{\partial C^c(K,T)}{\partial K}\right]\left[\frac{\partial^2 C^c(K,T)}{\partial K^2}\right]^{-1} \tag{3.17}$$

While (3.17) provides an expression for the (local) volatility surface in terms of the observed market prices, the framework proposed by Dupire (1994) is built on a non-parametric approach, and as such it does not specify the time evolution of the (local) volatility surface. In other words, while providing a perfect fit to the existing surface, the model should be calibrated each time the option prices quoted on the market move. Finally, calibrating the Dupire's model assumes the existence of a continuum of expirations and exercise prices; as a result, for a pratical implementation one would need to resort to interpolation techniques.

3.2 | Stochastic Volatility

The disadvantages of the local volatility models discussed in the previous section, along with the stylised facts presented at the beginning of this chapter, prompt the need for a functional, parametric specification of the dynamics for the underlying's volatility process. Over the years, several authors proposed modelling volatility as a random variable. For instance, Scott (1987) and Hull and White (1987) were among the first to generalise the BSM framework by allowing for stochastic volatility; the authors, however, were not able to provide closed-form solutions to price options contracts and resorted to numerical methods to solve the two-dimensional pricing PDEs. Later on, Eisenberg and Jarrow (1991) as well as Stein and Stein (1991) proposed a different approach wherein the premium charged on a European option is obtained by averaging the BSM prices over all possible volatility paths. The derivation of the two models, however, was built on the assumption of zero correlation between returns and volatility. The first framework explicitly accounting for both the leverage effect and the mean reverting nature of v_t is due to Heston (1993). Consider a generic SV framework based on the bivariate diffusion described by:

$$dS_t = \mu S_t\,dt + \sqrt{v_t}S_t\,dW_{1,t} \tag{3.18}$$

$$dv_t = a(t,S_t,v_t)S_t\,dt + \sigma b(t,S_t,v_t)\sqrt{v_t}\,dW_{2,t} \tag{3.19}$$

where $\{W_{1,t}\}_{t\geq 0}$ and $\{W_{2,t}\}_{t\geq 0}$ are two correlated BMs so that $\mathbb{E}[W_{1,t}W_{2,t}] = \rho t$, or equivalently $\mathbb{E}[dW_{1,t}\,dW_{2,t}] = \rho\,dt$. Indeed, one has:

$$\rho = \frac{\mathbb{E}[W_{1,t},W_{2,t}] - \mathbb{E}[W_{1,t}]\mathbb{E}[W_{2,t}]}{\sqrt{\mathbb{V}[W_{1,t}]\mathbb{V}[W_{2,t}]}} = \frac{\mathbb{E}[W_{1,t},W_{2,t}]}{\sqrt{t^2}} \implies \rho t = \mathbb{E}[W_{1,t}W_{2,t}] \tag{3.20}$$

In his original derivation, Heston (1993) assumed the dynamics of volatility $\sqrt{v_t}$ are dictated by a *Ornstein-Uhlenbeck* (OU) process as in Stein and Stein (1991):

$$d\sqrt{v_t} = -\beta\sqrt{v_t}\,dt + \delta\,dW_{2,t} \tag{3.21}$$

Applying the Itô-Doeblin formula with $f(t, v_t) = \sqrt{v_t}^2$ then reveals that the process followed by the variance of the spot price is described by the following SDE:

$$\begin{aligned} dv_t &= \frac{\partial v}{\partial t}\,dt + \frac{\partial v}{\partial\sqrt{v}}\,d\sqrt{v_t} + \frac{1}{2}\frac{\partial^2 v}{\partial\sqrt{v}^2}(d\sqrt{v_t})^2 \\ &= 2\sqrt{v_t}(\delta\,dW_{2,t} - \beta\sqrt{v_t}\,dt) + \delta^2\,dt \\ &= (\delta^2 - 2\beta v_t)\,dt + 2\delta\sqrt{v_t}\,dW_{2,t} \end{aligned} \tag{3.22}$$

which may be also expressed in the form of the well-known square-root process firstly employed by Cox et al. (1985) to model interest rates:

$$dv_t = \kappa(\theta - v_t)\,dt + \sigma\sqrt{v_t}\,dW_{2,t} \tag{3.23}$$

Here, $\theta := \delta^2/2\beta$ denotes the long-term mean of the variance process — that is, by construction one has $\lim_{t\to\infty}\mathbb{E}[v_t] = \theta$. In general, the variance of the spot price oscillates

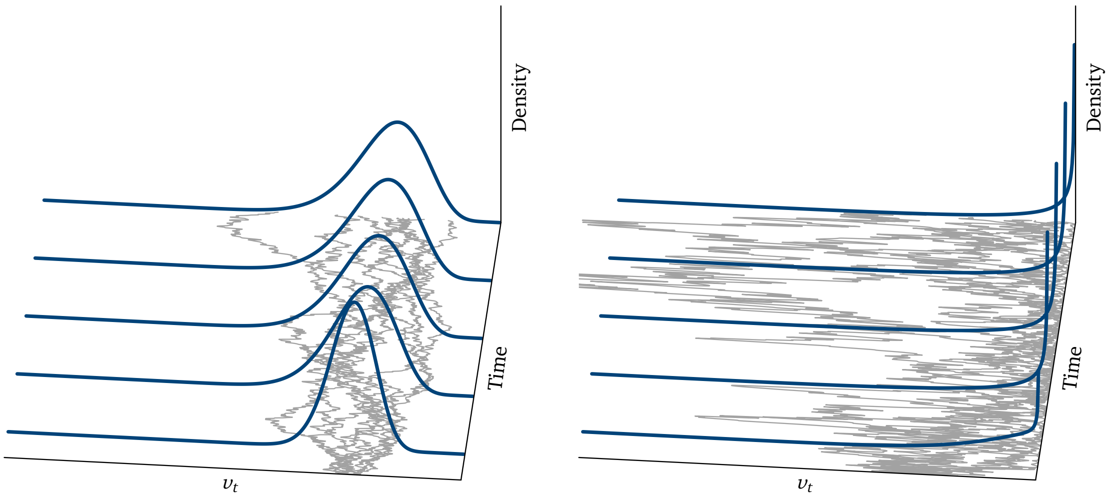


Figure 3.3 Monte Carlo simulation of ten volatility sample paths dictated by the square-root process described above with $\kappa = 3.0$, $\theta = 0.04$, and $\sigma = 0.20$ (0.50) on the left (right) with corresponding non-central χ^2 distributions at five different times. In the plot showed on the right — where the Feller condition does not hold — the volatility process tends to cluster around zero.

around θ and reverts to it at a speed dictated by $\kappa := 2\beta$ with oscillations whose magnitude is specified by the volatility of volatility, $\sigma = 2\delta$. The representation (3.23) for the variance process ensures the latter will be strictly positive as long as the Feller condition ($\sigma^2 < 2\kappa\theta$) is verified (Feller, 1951). Overall, it is known the variance process $\{v_t\}_{t\geq 0}$ at t is a random variable which induces a non-central χ^2 distribution scaled by $c \in \mathbb{R}$ with k degrees of freedom and noncentrality parameter λ given by:

$$c = \frac{(1-e^{-t\kappa})\sigma^2}{4\kappa} \quad \wedge \quad k = \frac{4\kappa\theta}{\sigma^2} \quad \wedge \quad \lambda = \frac{4\kappa v_0 e^{-t\kappa}}{(1-e^{-t\kappa})\sigma^2} \tag{3.24}$$

The plots in Figure 3.3 present paths of the variance process otained by discretising (3.23) both with and without the Feller condition being verified. Clearly, in the latter case the density tends to be concentrated around the origin — that is, $v_t = 0$.

One may follow a procedure similar to that of Chapter 1 to derive the pricing PDE needed to value an option under the Heston's dynamics. Here, however, the methodology will be slightly more involved as the presence of an additional source of randomness as the diffusive component in the dynamics of the underlying's variance leads to an incomplete market. As a result, rebalancing a position in the underlying is not sufficient to replicate the option and an additional (longer-maturity) contract is required to hedge the risks away. Suppose one is interested in valuing an option worth V_t at time t, and by standard assumptions consider another option worth U_t and a risk-free money market account which evolves according to $dB_t = rB_t dt$. Just as the inclusion of a discontinuous component in the underlying price process prompted the need for an extension of the Itô-Doeblin formula discussed in Chapter 2, with a bivariate diffusion one needs to account for the additional source of randomness driving the (variance) process. In particular, for two separate stochastic processes $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$, (1.23) becomes:

$$df_t = \frac{\partial f}{\partial t}\,dt + \frac{\partial f}{\partial X}\,dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}(dX_t)^2 + \frac{\partial f}{\partial Y}\,dY_t + \frac{1}{2}\frac{\partial^2 f}{\partial Y^2}(dY_t)^2 + \frac{\partial^2 f}{\partial X \partial Y}\,dX_t\,dY_t \tag{3.25}$$

In other words, the formula remains fairly simple up to the inclusion of the relevant derivatives for the second process considered as well as a cross-term accounting for the (quadratic) covariation of $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$. Therefore, in the framework proposed by Heston (1993), the value V_t of the option obeys the dynamics described by:

$$\begin{aligned} dV_t &= \frac{\partial V}{\partial t}\,dt + \frac{\partial V}{\partial S}\,dS_t + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(dS_t)^2 + \frac{\partial V}{\partial v}\,dv_t + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}(dv_t)^2 + \frac{\partial^2 V}{\partial S \partial v}\,dS_t\,dv_t \\ &= \frac{\partial V}{\partial t}\,dt + \frac{\partial V}{\partial S}\,dS_t + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}v_t S_t^2\,dt + \frac{\partial V}{\partial v}\,dv_t + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\sigma^2 v_t\,dt + \frac{\partial^2 V}{\partial S \partial v}\rho\sigma S_t v_t\,dt \\ &= \left[\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}v_t S_t^2 + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\sigma^2 v_t + \frac{\partial^2 V}{\partial S \partial v}\rho\sigma S_t v_t\right]dt + \frac{\partial V}{\partial S}\,dS_t + \frac{\partial V}{\partial v}\,dv_t \end{aligned} \tag{3.26}$$

To ease notation further on, denote the differential operator above by $\mathcal{L}V(t, S, v)$:

$$\mathcal{L}V(t, S, v) := \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2} v_t S_t^2 + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\sigma^2 v_t + \frac{\partial^2 V}{\partial S \partial v}\rho\sigma S_t v_t \tag{3.27}$$

The application of the no-arbitrage principle requires the construction of a replicating (self-financing) portfolio; consider investing Δ units in the underlying, Σ units in the longer maturity option, and α units in the money market account so that:

$$V_t = \Delta S_t + \Sigma U_t + \alpha B_t \implies dV_t = \Delta\, dS_t + \Sigma\, dU_t + \alpha\, dB_t \tag{3.28}$$

Substituting for the dynamics for the two options and the money market account, one gets:

$$\mathcal{L}V(t, S, v)\, dt + \frac{\partial V}{\partial S}\, dS_t + \frac{\partial V}{\partial v}\, dv_t = \Delta\, dS_t + \Sigma\left[\mathcal{L}U(t, S, v)\, dt + \frac{\partial U}{\partial S}\, dS_t + \frac{\partial U}{\partial v}\, dv_t\right] + \alpha r B\, dt \tag{3.29}$$

Similarly to the derivation of the BSM fundamental PDE, one can then remove the randomness in the portfolio arising from the two Brownian motions by imposing:

$$\frac{\partial V}{\partial S} = \Delta + \Sigma\frac{\partial U}{\partial S} \implies \Delta = \frac{\partial V}{\partial S} - \Sigma\frac{\partial U}{\partial S} \tag{3.30}$$

$$\frac{\partial V}{\partial v} = \Sigma\frac{\partial U}{\partial v} \implies \Sigma = \frac{\partial V}{\partial v}\Bigg/\frac{\partial U}{\partial v} \tag{3.31}$$

Plugging the values just found into (3.29) the latter simplifies significantly to:

$$\begin{aligned}\mathcal{L}V(t, S, v)\, dt &= \Sigma\mathcal{L}U(t, S, v)\, dt + \alpha r B\, dt \\ &= \Sigma\mathcal{L}U(t, S, v)\, dt + r\left[V_t - \frac{\partial V}{\partial S}S_t + \Sigma\frac{\partial U}{\partial S}S_t - \Sigma U_t\right] dt\end{aligned} \tag{3.32}$$

Moving all the terms related to the first option's value on the left-hand side, and substituting for the units Σ invested in the longer maturity option from (3.31), one reaches:

$$\frac{\mathcal{L}V(t, S, v) - rV_t + rS_t\dfrac{\partial V}{\partial S}}{\partial V/\partial v} = \frac{\mathcal{L}U(t, S, v) - rU_t + rS_t\dfrac{\partial U}{\partial S}}{\partial U/\partial v} \tag{3.33}$$

As the (sub-)PDE on the left-hand side of (3.33) only concerns the option to be valued while the terms on the right-hand side only refer to the option used for vega-hedging: it follows that the identity must equal a function $f(t, S_t, v_t)$ of the independent variables. In particular, focusing on the left-hand side and expanding the differential operator one has:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2} v_t S_t^2 + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\sigma^2 v_t + \frac{\partial^2 V}{\partial S \partial v}\rho\sigma S_t v_t - rV_t + rS_t\frac{\partial V}{\partial S} := \frac{\partial V}{\partial v} f(t, S_t, v_t) \tag{3.34}$$

The $f(t, S_t, v_t)$ function is multiplied by the first partial derivative of the option's value with respect to the underlying's variance; hence, in analogy with the BSM fundamental PDE,

$f(t, S_t, \upsilon_t)$ should be some function of the variance process's risk-neutral drift. To start, the bivariate diffusion governing the spot price and its variance may be equally expressed relying on two uncorrelated BMs $\{Z_{1,t}\}_{t\geq 0}$ and $\{Z_{2,t}\}_{t\geq 0}$; in particular, let:

$$dW_{1,t} = \sqrt{1-\rho^2}\, dZ_{1,t} + \rho\, dZ_{2,t} \tag{3.35}$$

$$dW_{2,t} = dZ_{2,t} \tag{3.36}$$

for all $t \geq 0$. As a result, the two SDEs become:

$$dS_t = \mu S_t\, dt + \sqrt{(1-\rho^2)(\upsilon_t)} S_t\, dZ_{1,t} + \sqrt{\upsilon_t}\rho S_t\, dZ_{2,t} \tag{3.37}$$

$$d\upsilon_t = \kappa(\theta - \upsilon_t)\, dt + \sigma\sqrt{\upsilon_t}\, dZ_{2,t} \tag{3.38}$$

Under a given EMM* $\mathbb{Q}_\xi \in \mathbf{Q}$, in analogy with (1.38), the two BMs will be replaced by their risk-neutral counterparts given by:

$$dZ_{1,t}^{\mathbb{Q}_\xi} = dZ_{1,t} + \frac{\mu - r - \xi\rho\sqrt{\upsilon_t}}{\sqrt{(1-\rho^2)(\upsilon_t)}}\, dt \tag{3.39}$$

$$dZ_{2,t}^{\mathbb{Q}_\xi} = dZ_{2,t} + \xi\, dt \tag{3.40}$$

where ξ stands for the market price of volatility, and the modified Sharpe ratio appearing in (3.39) ensures the stock price will grow at the risk-free rate:

$$\begin{aligned}
\frac{dS_t}{S_t} &= \mu\, dt + \sqrt{(1-\rho^2)(\upsilon_t)}\left(dZ_{1,t}^{\mathbb{Q}_\xi} - \frac{\mu - r - \xi\rho\sqrt{\upsilon_t}}{\sqrt{(1-\rho^2)(\upsilon_t)}}\, dt\right) + \sqrt{\upsilon_t}\rho\left(dZ_{2,t}^{\mathbb{Q}_\xi} - \xi\, dt\right) \\
&= (\mu - \mu + r + \xi\rho\sqrt{\upsilon_t} - \sqrt{\upsilon_t}\rho\xi)\, dt + \sqrt{(1-\rho^2)(\upsilon_t)}\, dZ_{1,t}^{\mathbb{Q}_\xi} + \sqrt{\upsilon_t}\rho\, dZ_{2,t}^{\mathbb{Q}_\xi} \\
&= r\, dt + \sqrt{(1-\rho^2)(\upsilon_t)}\, dZ_{1,t}^{\mathbb{Q}_\xi} + \sqrt{\upsilon_t}\rho\, dZ_{2,t}^{\mathbb{Q}_\xi} = r\, dt + \sqrt{\upsilon_t}\, dW_{1,t}^{\mathbb{Q}_\xi}
\end{aligned} \tag{3.41}$$

As a result, the bivariate diffusion under the chosen EMM $\mathbb{Q}_\xi$ is characterised by:

$$dS_t = rS_t\, dt + \sqrt{\upsilon_t} S_t\, dW_{1,t}^{\mathbb{Q}_\xi} \tag{3.42}$$

$$d\upsilon_t = [\kappa(\theta - \upsilon_t) - \xi\sigma\sqrt{\upsilon_t}]\, dt + \sigma\sqrt{\upsilon_t}\, dW_{2,t}^{\mathbb{Q}_\xi} \tag{3.43}$$

A replication argument similar to the one discussed above shows that the absence of arbitrage opportunities implies the following pricing PDE:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\upsilon_t S_t^2 + \frac{1}{2}\frac{\partial^2 V}{\partial \upsilon^2}\sigma^2\upsilon_t + \frac{\partial^2 V}{\partial S\partial \upsilon}\rho\sigma S_t\upsilon_t - rV_t + rS_t\frac{\partial V}{\partial S} = -\frac{\partial V}{\partial \upsilon}[\kappa(\theta - \upsilon_t) - \xi\sigma\sqrt{\upsilon_t}] \tag{3.44}$$

*If the stock price is driven by a bivariate diffusion, the market is not complete and the set of risk-neutral measures has a cardinality greater than one — that is, the EMM is not unique.

It follows that $f(t, S_t, v_t) = \kappa(\theta - v_t) - \xi\sigma\sqrt{v_t}$, confirming the initial guess.

At this stage, Heston proposes a solution for (3.44) of the form:

$$V_t = S_t P_1 - K e^{-r\tau} P_2 \tag{3.45}$$

where P_1 and P_2 represent the (risk-neutral) conditional probabilities of the option ending in the money at expiration under the stock and risk-free asset numéraires, respectively. Then, passing to the log-space with $X_t := \log(S_t)$ and combining (3.44) and (3.45) reveals that the two probabilities must satisfy the following PDEs:

$$\frac{\partial P_j}{\partial t} + \frac{\partial P_j}{\partial x}(r + u_j v_t) + \frac{\partial P_j}{\partial v}(\alpha_j - \beta_j v_t) + \frac{1}{2}\frac{\partial^2 P_j}{\partial x^2} v_t + \frac{1}{2}\frac{\partial^2 P_j}{\partial v^2}\sigma^2 v_t + \frac{\partial^2 P_j}{\partial x \partial v}\rho\sigma v_t = 0 \tag{3.46}$$

where $u_1 = -u_2 = 1/2$, $\alpha_1 = \alpha_2 = \kappa\theta$, $\beta_1 = \kappa - \rho\sigma$, and $\beta_2 = \kappa$.

Even though finding P_1 and P_2 from (3.46) is not immediate, Heston (1993) showed that their Fourier transforms (i.e., characteristic functions) must satisfy the same PDEs subject to the terminal condition $\Phi_T = \mathbb{E}[\exp(izX_T)] = e^{izX_T}$. The expressions for the characteristic functions are particularly tractable as the model considered belongs to the class of affine diffusions, wherein drifts and covariances are linear in the state vector $[x, v]$ (Gatheral, 2011); as a result, the Fourier transform of each probability is simply given by:

$$\mathbb{E}[\exp(izX_T)] = \exp\left[C(\tau, z) + D(\tau, z)v_t + izX_t\right] \tag{3.47}$$

where, employing the dynamics for the underlying's price and variance proposed by Heston (1993), the $C(\tau, z)$ and $D(\tau, z)$ coefficients correspond to the expressions below:

$$C(\tau, z) = rizt + \frac{\alpha}{\sigma^2}\left\{\left(\beta_j - \rho\sigma iz + d\right)\tau - 2\log\left[\frac{1 - ge^{d\tau}}{1 - g}\right]\right\} \tag{3.48}$$

$$D(\tau, z) = \frac{\beta_j - \rho\sigma iz + d}{\sigma^2}\left[\frac{1 - e^{d\tau}}{1 - ge^{d\tau}}\right] \tag{3.49}$$

$$d = \sqrt{(\rho\sigma iz)^2 - \sigma^2\left(2u_j iz - z^2\right)} \tag{3.50}$$

$$g = \frac{\beta_j - \rho\sigma iz + d}{\beta_j - \rho\sigma iz - d} \tag{3.51}$$

Once the characteristic functions for P_1 and P_2 are specified, one can then recover the probabilities by inverting the Fourier transform Φ_j as follows:

$$P_j = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{e^{-iz\log(k)}\Phi_{j,T}(z)}{iz}\right]dz \tag{3.52}$$

A detailed discussion of the Fourier inversion technique, popularised by Heston (1993) and presented in (3.52) to recover the originally sought probabilities can be found in Chapter 4.

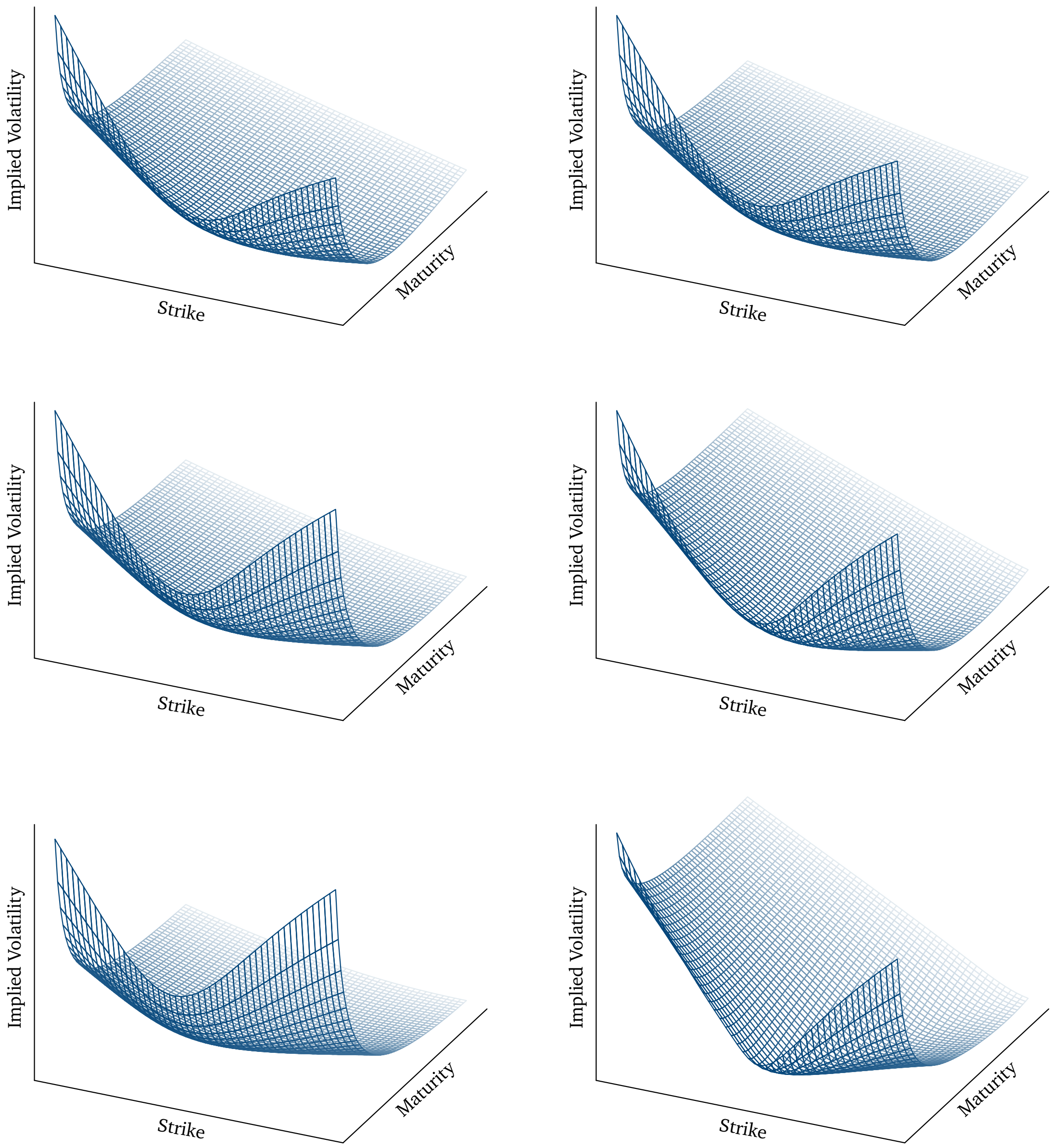


Figure 3.4 Volatility surfaces generated by the stochastic volatility model under $\mathbb{Q}_\xi$. The left column shows how the volatility profile changes for different values of the correlation between the spot price and its volatility — that is, from top to bottom $\rho = -0.75$, -0.25, and 0.00. The right column does the same for three distinct values of the volatility of volatility σ — that is, from top to bottom $\sigma = 0.25$, 0.50, and 1.50. The speed of mean reversion κ and the long-term average variance θ are kept constant at 3.00 and 0.04, respectively.

It is then possible to create a synthetic data set of options prices generated by the SV framework and look at the BSM implied volatility surface Σ_t, presented in Figure 3.4. First of all, one notes how including an additional diffusion driving the (latent) volatility process allows for a closer reproduction of the implied skew for long-term contracts; in particular, the correlation coefficient ρ between the two BMs driving the spot price and its volatility is responsible for the sign of the asymmetry in the implied skew: when $\rho = 0$ one achieves a symmetric smile. At the same time, while returning a good fit for expirations far ahead, the introduction of dependence in the spot price increments is not sufficient to replicate the pronounced smile for short-term contracts. In fact, the bivariate diffusion of Heston (1993) does not generate a sufficient variation in the spot price over a short tenor unless one assumes an excessive value for the volatility of volatility σ; hence, one needs to rely on the inclusion of discontinuities as to achieve such variation. Jumps may be included either in the diffusion driving the spot price or its variance, or in both; the next section is devoted to the discussion of a pricing model proposed by Bates (1996) combining the benefits of the SV model of Heston (1993) and the JD framework of Merton (1976).

3.3 | Stochastic Volatility Jump-Diffusion

As anticipated at the end of the previous section, the pricing model proposed by Bates (1996) "merges" the improvements over the BSM framework implemented by Merton (1976) and Heston (1993). In particular, Bates's *stochastic volatility jump-diffusion* (SVJD) model still assumes a bivariate diffusion for the underlying's spot price and volatility but allows for the underlying's path to exhibit discontinuity points. The latter are introduced through a compound Poisson process with intensity λ and, as the distribution of a given jump Y_t is assumed to be log-normal — that is, $\log(Y_t) \sim \mathrm{N}(\alpha, \beta^2)$, such a framework may be equally interpreted as a generalisation of the JD model or the SV representation. The discontinuous component of the instantaneous rate of return does still entail a predictable part, which should be subtracted from the drift of the Brownian motion to preserve the randomness of the process. In particular, one has:

$$\eta := \mathbb{E}[Y_t - 1] = \exp\left(\alpha + \frac{\beta^2}{2}\right) - 1 \tag{3.53}$$

Since $\{Y_t\}_{t\geq 0}$ is assumed to be independent of the Poisson process $\{N_t\}_{t\geq 0}$, one then has:

$$\mathbb{E}[(Y_t - 1)\, dN_t] = \mathbb{E}[Y_t - 1]\mathbb{E}[dN_t] = \lambda\eta\, dt \tag{3.54}$$

Hence, the instantaneous return on the underlying is given by the following SDE:

$$\frac{dS_t}{S_t} = \mu\, dt + \sqrt{v_t}\, dW_{1,t} + (Y_t - 1)\, dN_t - \lambda\eta\, dt = (\mu - \lambda\eta)\, dt + \sqrt{v_t}\, dW_{1,t} + (Y_t - 1)\, dN_t \tag{3.55}$$

Finally, one has to allow for more than one jump to occur at each time. Since the Poisson counter dictates the number of discontinuities, (2.40) becomes:

$$\frac{dS_t}{S_t} = (\mu - \lambda\eta)\, dt + \sqrt{v_t}\, dW_{1,t} + \left(\prod_{j=1}^{dN_t} Y_j - 1 \right) \tag{3.56}$$

Applying the Itô-Doelin formula for discontinuous processes presented in Chapter 2, one notes that the SDE for the log-spot price X_t is given by:

$$\begin{aligned} dX_t &= \frac{1}{S_t}\, dS_t - \frac{1}{2S_t^2}(dS_t)^2 + \log\left(S_t \prod_{j=1}^{dN_t} Y_j \right) - \log(S_t) \\ &= (\mu - \lambda\eta)\, dt + \sqrt{v_t}\, dW_{1,t} - \frac{v_t}{2}\, dt + \log\left(\prod_{j=1}^{dN_t} Y_j \right) \\ &= \left(\mu - \lambda\eta - \frac{v_t}{2} \right) dt + \sigma\, dW_{1,t} + \sum_{j=1}^{dN_t} \log(Y_j) \end{aligned} \tag{3.57}$$

By standard replication arguments, the diffusion for $\{S_t\}_{t\geq 0}$ in the log-space under a given EMM $\mathbb{Q}_\xi \in \mathcal{Q}$ is described by the following SDEs:

$$dX_t = \left(r - \lambda\eta - \frac{v_t}{2} \right) dt + \sigma\, dW_{1,t}^{\mathbb{Q}_\xi} + \sum_{j=1}^{dN_t} \log(Y_j) \tag{3.58}$$

$$dv_t = [\kappa(\theta - v_t) - \xi\sigma\sqrt{v_t}]\, dt + \sigma\sqrt{v_t}\, dW_{2,t}^{\mathbb{Q}_\xi} \tag{3.59}$$

where ξ stands for the market price of volatility risk as in Heston (1993). Even in the model proposed by Bates (1996), the premia charged for European options is found through the inversion of the (risk-neutral) characteristic function of the log-spot price of the underlying. As the Poisson counter $\{N_t\}_{t\geq 0}$ for jumps is assumed independent from the two BMs driving (3.58) and (3.59), the Fourier transform for the SVJD model can be obtained by simply multiplying that of the SV framework times the characteristic function of the JD process. Following Schoutens et al. (2004), one has:

$$\mathbb{E}[\exp(izX_T)] = \exp\{[iz(X_t + rt)] \tag{3.60}$$

$$+ \left\{ \frac{\theta\kappa}{\sigma^2} \left[(\kappa - \rho\sigma iz - d)t - \frac{2\log(1 - ge^{-dt})}{(1-g)} \right] \right\} \tag{3.61}$$

$$+ \left[\frac{v_0^2(\kappa - \rho\sigma iz - d)(1 - e^{-dt})}{\sigma^2(1 - ge^{-dt})} \right] \tag{3.62}$$

$$+ \left\{ -\lambda\alpha izt + \lambda t \left[(1+\alpha)^{iz} \exp\left(\frac{\beta^2 iz(iz-1)}{2} - 1 \right) \right] \right\} \} \tag{3.63}$$

where d and g are two auxiliary variables given by:

$$d = \sqrt{(\rho\sigma iz)^2 - \sigma^2(-iz - z^2)} \tag{3.64}$$

$$g = \frac{\kappa - \rho\sigma iz - d}{\kappa - \rho\sigma iz + d} \tag{3.65}$$

Once the CF of the log-price is specified, one can then price options contracts in the Bates (1996) framework by relying on the Fourier inversion technique proposed by Heston (1993). The volatility profiles generated by the SVJD model of Bates (2006) are not reported for they are qualitatively similar to those in Figure 3.4, except one can rely on much more reasonable values for the volatility of volatility σ and the spot-volatility correlation ρ. In particular, the left asymmetry typical of the distribution of returns can be obtained both through a leverage effect (i.e., $\rho < 0$) and average negative jump size (i.e., $\alpha < 0$). Similarly, a more pronounced smile in the Bates (1996) model may be the result of either a larger frequency of jumps λ or a greater volatility of volatility σ as in Heston (1993).

While the (log-normal) risk-neutral density of the terminal underlying price is still tractable as the latter is driven by a simple GBM as in Black and Scholes (1973), making a direct computation of the risk-neutral expectation of the payoff a feasible approach, the expressions for the EMM $\mathbb{Q}(S_T \mid \mathcal{F}_t)$ in the three other frameworks considered in this thesis are either much more involved or not directly available. The use of Fourier inversion techniques within financial modelling was introduced by Stein and Stein (1991) and popularised by the work of Heston (1993) just discussed. As one shall see in the next chapter, the premium charged for a European option can be decomposed into a portfolio of Arrow-Debreu securities — assets paying off a single unit of the numéraire chosen conditional on a prespecified event occurring — to be valued under the appropriate equivalent martingale measure. Such intuition leads to a pricing formula resembling that employed in Black and Scholes (1973) which, however, are also applicable in the case of more complex dynamics for the underlying asset. Moreover, authors such as Lewis (2001) proposed a technique relying on a single inversion of a generalised CF defined for the payoff function of an option in some *strip of regularity* in the complex plane.

Chapter 4

Fourier Transform Methods

As introduced in the first chapter, under the martingale pricing approach (or risk-neutral valuation), the premium V_t of a derivative contract can be found by solving a conditional expectation under the equivalent martingale measure. Harrison and Pliska (1981) showed that such a probability measure rules out the presence of arbitrage opportunities; moreover, in the (particular) case of complete markets, the risk-neutral measure is unique, and a single price for the options contract can be established (Harrison and Pliska, 1983). A European call option pays off $[S_T - K]^+$ at the expiration date T; as a result, its value at an intermediate time $t \in [0, T]$ is given by:

$$V_t(H_T) = e^{-r\tau}\mathbb{E}_{\mathbb{Q}}[H_T \mid \mathcal{F}_t] = e^{-r\tau} \int_0^{\infty} [S_T - K]^+ \mathbb{Q}(S_T \mid \mathcal{F}_t)\, dS_T \tag{4.1}$$

where $\tau := T - t$ and $\mathbb{Q}(S_T \mid \mathcal{F}_t)$ stands for the conditional risk-neutral transition probability of the underlying reaching state S_T at the expiration T given the filtration $\mathcal{F}_t$ as of the valuation date t. Such density is entirely determined by the risk-neutral dynamics specified for the underlying; for instance, if S_t follows a geometric Brownian motion, then one can find the price charged for a call option on it by solving the following integral:

$$C_t(S_t, T, K) = \frac{e^{-r\tau}}{S_T \sigma \sqrt{2\pi\tau}} \int_K^{\infty} (S_T - K) \exp\left\{ -\frac{1}{2} \left[\frac{X_T - \left[X_0 + \left(r - \sigma^2/2\right)\tau\right]}{\sigma\sqrt{\tau}} \right]^2 \right\} dS_T \tag{4.2}$$

In fact, in such a case, the log-price X_t follows a Normal distribution. Unfortunately, for all the other underlying asset dynamics discussed in this thesis, a functional form of the risk-neutral density is either more involved or not directly available. For instance, the mere superposition of a (compound) Poisson process to a geometric Brownian motion leads to an infinite sum in the expression of the risk-neutral density as discussed in the JD model of Chapter 2. Nevertheless, a one-to-one relationship holds between the distribution induced by a random variable X and its characteristic function presented in Definition 1.6.

Furthermore, the foundations provided by Lévy (1925) make it possible to recover a probability density $f(x)$ starting from the CF of X; this is particularly convenient when dealing with Lévy processes, where the characteristic function is linear in time and easily obtainable through the decomposition described in Theorem 2.2.

Overall, the characteristic function of a given random variable X is related to (i.e., defined as) the Fourier transform of the probability density function $f(x)$ it induces, as will be further discussed in the next section.

4.1 | Fourier Transform

The concept of Fourier transform is named after French mathematician Joseph Fourier, who introduced it in 1822. This decomposition is widely employed in studying differential equations and signal processing. Within financial modelling, the Fourier transform was introduced by Stein and Stein (1991) and popularised by the work of Heston (1993).

As already discussed, the characteristic function of a random variable X equals the expectation of the complex exponential e^{izX} — that is, $\Phi_X(z)$ is given by:

$$\Phi_X(z) = \mathbb{E}\left[\exp(izX)\right] = \int_{\mathbb{R}} e^{izx}\, df_X(x) := \mathfrak{F}[f(x)] \tag{4.3}$$

where $\mathfrak{F}[f(x)]$ denotes the Fourier transform of the random variable's density $f(x)$. Equivalently, one can rely on Euler's formula to rewrite the complex exponential as:

$$\exp(izX) = \cos(zX) + i\sin(zX) \tag{4.4}$$

From (4.4), it is clear that, for any $z \in \mathbb{R}$, e^{izX} corresponds to a point on the unit circle in the complex plane. As a result, $\Phi_X(z)$ is symmetric around $z = 0$, its norm cannot exceed one, and $\Phi_X(0) := 1$ since $0 \cdot X$ is deterministic. The symmetry verified by the characteristic function of a random variable entails that one can focus on the positive half axis ($z > 0$) and still be able to characterise the distribution induced by X fully. Moreover, the function's complex conjugate $\overline{\Phi_X(z)}$ is simply given by $\Phi_X(-z)$. Finally, two useful properties of the Fourier transform concern differentiability and convolutions of densities. In fact, as shown in Appendix A, both are mapped to multiplications in the Fourier space:

$$\text{Differentiation:} \qquad \mathfrak{F}\left[\frac{\partial^n f(x)}{\partial x^n}\right] = -(iz)^n \mathfrak{F}\left[f(z)\right] \tag{4.5}$$

$$\text{Convolution:} \qquad \mathfrak{F}\left[(f * g)(x)\right] = \mathfrak{F}\left[f(x)\right]\mathfrak{F}\left[g(x)\right] \tag{4.6}$$

Given the premises above, and building on the results of Lévy (1925), Gil-Pelaez (1951) provides the following representation for a cumulative distribution function (CDF):

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(x)\,dx = \frac{1}{2} - \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{-izx}\Phi_X(z)}{iz}\,dz \tag{4.7}$$

A derivation of the integral above can be found in Appendix A. The density function is then obtained by differentiating the CDF once:

$$f_X(x) = \frac{\partial F_X(x)}{x} = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-izx}\Phi_X(z)\,dz \tag{4.8}$$

One can then exploit the symmetry of the characteristic function and see that:

$$\Re[\phi_X(z)] = \frac{\phi_X(z) + \overline{\phi_X(z)}}{2} = \frac{\phi_X(z) + \phi_X(-z)}{2} \tag{4.9}$$

$$\Im[\phi_X(z)] = \frac{\phi_X(z) - \overline{\phi_X(z)}}{2i} = \frac{\phi_X(z) - \phi_X(-z)}{2i} \tag{4.10}$$

Hence, the CDF in (4.7) can be rewritten as:

$$\begin{aligned}
F_X(x) &= \frac{1}{2} - \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{-izx}\Phi_X(z)}{iz}\,dz \\
&= \frac{1}{2} - \frac{1}{2\pi}\int_{-\infty}^{0} \frac{e^{-izx}\Phi_X(z)}{iz}\,dz + \frac{1}{2\pi}\int_{0}^{\infty} \frac{e^{-izx}\Phi_X(z)}{iz}\,dz \\
&= \frac{1}{2} - \frac{1}{2\pi}\int_{0}^{\infty} \left[\frac{-e^{izx}\Phi_X(-z) + e^{-izx}\Phi_X(z)}{iz}\right] dz \\
&= \frac{1}{2} - \frac{1}{\pi}\int_{0}^{\infty} \Re\left[\frac{e^{-izx}\Phi_X(z)}{iz}\right] dz
\end{aligned} \tag{4.11}$$

Before delving into the option pricing techniques based on characteristic functions, it is essential to point out that the Fourier transform of a given function f requires the latter to be (absolutely) integrable — that is, for $\mathfrak{F}[f(x)]$ to exist it must be that:

$$\int_{\mathbb{R}} |f(x)|\,dx < \infty \tag{4.12}$$

Furthermore, if $f(x)$ and $g(x)$ are both square integrable functions, then inner products $\langle f, g\rangle$ are preserved under Fourier transforms:

$$\langle f(x), g(x)\rangle = \int_{\mathbb{R}} f(x)\overline{g}(x)\,dx = \frac{1}{2\pi}\int_{\mathbb{R}} \mathfrak{F}[f]\overline{\mathfrak{F}[g]}\,dz \tag{4.13}$$

The relationship above is known as Parseval identity (or Plancherel Theorem) and provides the basis for the call valuation formula proposed by Lewis (2001). The latter focused on *generalised* characteristic functions where the input variable z belongs to $\mathbb{C}$. In this case, however, the expectation of the complex exponential defining the characteristic function of a random variable is only defined in some *strip of regularity* $\mathbb{S}_x$ parallel to the real z-axis — that is, a region on the complex plane defined by $\alpha < \Im[z] < \beta$. Luckily, most properties discussed so far apply to generalised Fourier transforms as well. For instance, the Parseval identity given by (4.13) with $z \in \mathbb{C}$ and assuming $f = g$ becomes:

$$\int_{\mathbb{R}} |f(x)|^2 \, dx = \frac{1}{2\pi} \int_{\mathbb{R}} |\mathfrak{F}[f(\Re[z] + i\Im[z])]|^2 \, d\Re[z] \tag{4.14}$$

4.2 | Option Pricing

The results of Gil-Pelaez (1951) provide the basis for the closed-form solution of the call option valuation problem presented by Heston (1993). Indeed, the premium charged for a call option on an asset following a stochastic volatility model is obtained through a BSM-style formula involving conditional probabilities of the option being in the money at the expiration date; in turn, such probabilities are found by inverting the relevant characteristic functions. Not even a decade later, Bakshi and Madan (2000) generalised several past results concerning option pricing through Fourier transforms. Letting $k = \log(K)$ and $X_T = \log(S_T)$, one can always decompose the value of a call option as follows:

$$\begin{aligned} C_t(S_t, T, K) &= e^{-r\tau} \int_K^{\infty} (S_T - K)\mathbb{Q}(S_T \mid \mathcal{F}_t) \, dS_T \\ &= e^{-r\tau} \int_k^{\infty} e^{X_T} \mathbb{Q}(X_T \mid \mathcal{F}_t) \, dX_T - K e^{-r\tau} \int_k^{\infty} \mathbb{Q}(X_T \mid \mathcal{F}_t) \, dX_T \\ &= S_t \Pi_1(S_T > K \mid \mathcal{F}_t) - K e^{-r\tau} \Pi_2(S_T > K \mid \mathcal{F}_t) \end{aligned} \tag{4.15}$$

Here, both Π_1 and Π_2 represent the (risk-neutral) conditional probabilities of the option ending in the money at $t = T$. In particular, Π_1 and Π_2 are computed relying on the stock and risk-free asset as numéraire, respectively. Starting from the latter, the characteristic function of Π_2 is given by $\Phi_2(z) = \Phi_{X_T}(z) := \Phi_T(z) = \mathbb{E}_{\mathbb{Q}}[e^{iuX_T}]$; therefore, one has:

$$\begin{aligned} \Pi_2 &= \int_k^{\infty} \left(\frac{1}{2\pi} \int_{\mathbb{R}} e^{-iuX_T} \Phi_T(z) \, dz \right) dX_T \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \Phi_T(z) \left(\int_k^{\infty} e^{-iuX_T} \, dX_T \right) dz \\ &= \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \Re \left[\frac{e^{-izk} \Phi_T(z)}{iz} \right] dz \end{aligned} \tag{4.16}$$

A similar expression holds for Π_1, up to a change of measure accounting for the stock numéraire. Consider a measure $\tilde{\mathbb{Q}}$, linked to $\mathbb{Q}$ through the Radon-Nikodym derivative:

$$\frac{d\tilde{\mathbb{Q}}}{d\mathbb{Q}} = \frac{e^{X_t}}{\mathbb{E}_{\mathbb{Q}}[e^{X_t}]} \tag{4.17}$$

As a result, the probability Π_1 of the underlying being above the strike at maturity under the stock numéraire has a Fourier transform given by:

$$\Phi_1 = \tilde{\Phi}_T(z) = \mathbb{E}_{\tilde{\mathbb{Q}}}[e^{izX_t}] = \mathbb{E}_{\mathbb{Q}}\left[\frac{d\tilde{\mathbb{Q}}}{d\mathbb{Q}}e^{izX_t}\right] = \mathbb{E}_{\mathbb{Q}}\left[\frac{e^{X_t(1+iz)}}{\mathbb{E}_{\mathbb{Q}}[e^{X_t}]}\right] = \frac{\Phi_T(z-i)}{\Phi_T(-i)} \tag{4.18}$$

Therefore, a similar procedure to that used to reach (4.16) can be followed to show that

$$\Pi_1 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{e^{-izk}\Phi_1(z)}{iz}\right]dz = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{e^{-izk}\Phi_T(z-i)}{iz\Phi_T(-z)}\right]dz \tag{4.19}$$

Setting k to equal a particular measure of moneyness given by $k := \log(S_t/K) + r\tau$, the risk-neutral probabilities derived can be simplified to:

$$\Pi_1 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{e^{izk}\Phi_T(z-i)}{iz}\right]dz \tag{4.20}$$

$$\Pi_2 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{e^{izk}\Phi_T(z)}{iz}\right]dz \tag{4.21}$$

Finally, one can substitute these two expressions in the general valuation formula presented in (4.15) and combine the two integrals into one, thereby obtaining:

$$C_t(S_t,T,K) = \frac{S_t - Ke^{-r\tau}}{2} + \frac{1}{\pi}\int_0^\infty \left\{S_t\Re\left[\frac{e^{izk}\Phi_T(z-i)}{iz}\right] - Ke^{-r\tau}\Re\left[\frac{e^{izk}\Phi_T(z)}{iz}\right]\right\}dz \tag{4.22}$$

In the years following Bakshi and Madan's work, other authors (Attari, 2004; Bates, 2006) have proposed alternatives that always lead to a BSM-style formula but exploit the intrinsic relationship between the two probabilities of the option being profitable at expiration. In turn, such observation allows to reduce the computational burden required by the numerical evaluation of the integral(s).

Another approach to the problem of valuing a European option is pursued by Carr and Madan (1999), who provide a valuation formula especially suitable for a discretised version of the Fourier transform, allowing to rely on a single inversion to price options for several exercise prices. Unfortunately, the premium of a call option is not absolutely integrable, and an appropriate transformation must be performed to make the implementation of the *fast Fourier transform* feasible. In fact, by letting $X_t = \log(S_t)$ and $k = \log(K)$, one notices that $C_t(S_t,T,K)$ does not go to zero as $k \to -\infty$ (i.e., $K \to 0$); in particular:

$$\begin{aligned}
\lim_{k\to-\infty} C_t(S_t,T,K) &= \lim_{k\to-\infty} e^{-r\tau}\int_{\mathbb{R}} \left[e^{X_T} - e^k\right]^+ \mathbb{Q}(X_T \mid \mathcal{F}_t)\,dX_T \\
&= e^{-r\tau}\int_{\mathbb{R}} \left[e^{X_T} - e^{-\infty}\right]^+ \mathbb{Q}(X_T \mid \mathcal{F}_t)\,dX_T \\
&= e^{-r\tau}\int_{\mathbb{R}} e^{X_T}\mathbb{Q}(X_T \mid \mathcal{F}_t)\,dX_T = e^{-r\tau}\mathbb{E}_{\mathbb{Q}}[e^{X_t} \mid \mathcal{F}_t] = S_t
\end{aligned} \tag{4.23}$$

where the last equality follows by the martingale condition $\mathbb{E}_{\mathbb{Q}}[S_T \mid \mathcal{F}_t] = e^{r\tau}S_t$. Hence, $C_t(S_t, T, K)$ is clearly not (absolutely) integrable and therefore does not admit a Fourier transform. Nevertheless, Carr and Madan (1999) show that damping exponentially the call value leads to an integrable function — that is, it holds that:

$$c_t(S_t, T, K) := e^{\alpha k} C_t(S_t, T, K) \implies \int_{\mathbb{R}} |c_t(S_t, T, K)|\, dk < \infty \tag{4.24}$$

for an appropriate choice of $\alpha \in \mathbb{R}_+$. A visualisation of the effect of the damping parameter on the asymptotic properties of $C_t(S_t, T, K)$ is shown in Figure 4.1. It is clear that the damped option price becomes integrable on the negative half axis ($k < 0$) but increases exponentially for $k > 0$. Nevertheless, $c_t(S_t, T, K)$ will be square integrable (i.e., integrable on the entire k-axis) as long as its Fourier transform is finite at $z = 0$ (Carr and Madan, 1999). In turn, the characteristic function of the damped option price is given by:

$$\begin{aligned}
\psi_{c_T}(z) := \psi_T(z) &= \int_{\mathbb{R}} e^{izk} c_t(S_T, T, k)\, dk \\
&= \int_{\mathbb{R}} e^{izk} \int_{\mathbb{R}} e^{\alpha k} e^{-r\tau} \left[e^{X_T} - e^{k}\right]^+ \mathbb{Q}(X_T \mid \mathcal{F}_t)\, dX_T\, dk \\
&= \int_{\mathbb{R}} e^{izk} \int_{k}^{\infty} e^{\alpha k} e^{-r\tau} (e^{X_T} - e^{k}) \mathbb{Q}(X_T \mid \mathcal{F}_t)\, dX_T\, dk \\
&= \int_{\mathbb{R}} e^{-r\tau} \mathbb{Q}(X_T \mid \mathcal{F}_t) \underbrace{\left[\int_{-\infty}^{X_T} e^{izk} e^{\alpha k} (e^{X_T} - e^{k})\, dk\right]}_{A} dX_T
\end{aligned} \tag{4.25}$$

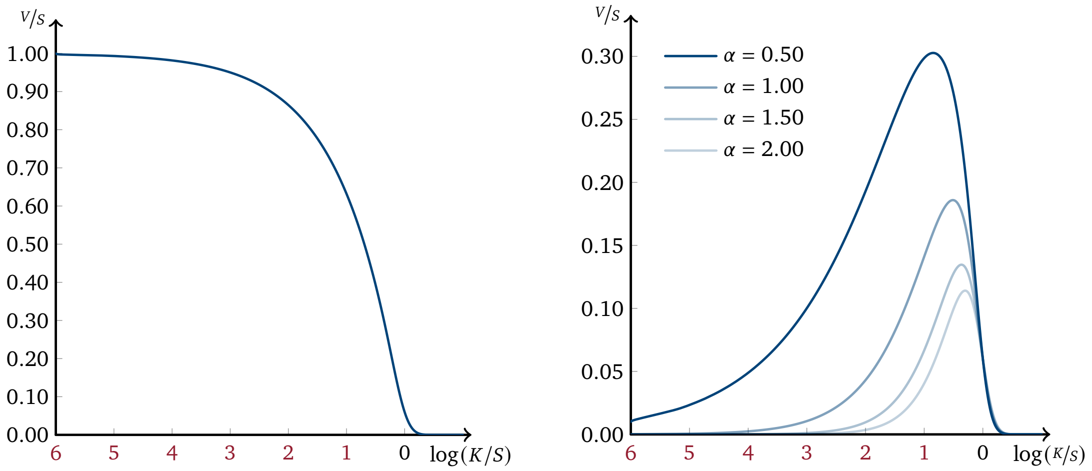


Figure 4.1 Left: Sample path of a variance-gamma process over an interval of ten years, with $\mu = 0.05$, $\sigma = 0.20$, $\alpha = 0.00$, $\beta = 0.50$, and $\lambda = 1.5$. The last parameter entails that one can expect three jumps every two years. Right: Logarithmic returns associated to the realisation.

The inner integral (A) in square brackets corresponds to the call payoff and can be computed on the interval $(-\infty, X_T]$ as follows:

$$
\begin{aligned}
A &= e^{X_T} \int_{-\infty}^{X_T} e^{(iz+\alpha)k}\, dk - \int_{-\infty}^{X_T} e^{(iz+\alpha+1)k}\, dk \\
&= \frac{e^{X_T}}{iz+\alpha} \left[e^{(iz+\alpha)k} \right]_{-\infty}^{X_t} - \frac{1}{iz+\alpha+1} \left[e^{(iz+\alpha+1)k} \right]_{-\infty}^{X_t} = \frac{e^{(iz+\alpha+1)X_T}}{iz+\alpha} - \frac{e^{(iz+\alpha+1)X_T}}{iz+\alpha+1}
\end{aligned} \tag{4.26}
$$

where the last equality holds since $\lim_{k\to-\infty} e^{(iz+\alpha)k} = \lim_{k\to-\infty} e^{(iz+\alpha+1)k} = 0$. At this point, plugging (4.26) into the integral for ψ_T, one obtains:

$$
\begin{aligned}
\psi_T(z) &= \int_{\mathbb{R}} e^{-r\tau} \mathbb{Q}(X_T \mid \mathcal{F}_t) \left[\frac{e^{(iz+\alpha+1)X_T}}{iz+\alpha} - \frac{e^{(iz+\alpha+1)X_T}}{iz+\alpha+1} \right] dX_T \\
&= \int_{\mathbb{R}} e^{-r\tau} \mathbb{Q}(X_T \mid \mathcal{F}_t) \left[\frac{e^{(iz+\alpha+1)X_T}}{(iz+\alpha)(iz+\alpha+1)} \right] dX_T \\
&= \frac{e^{-r\tau}}{(iz+\alpha)(iz+\alpha+1)} \underbrace{\int_{\mathbb{R}} e^{(iz+\alpha+1)X_T} \mathbb{Q}(X_T \mid \mathcal{F}_t)\, dX_T}_{B}
\end{aligned} \tag{4.27}
$$

Finally, notice that B is the characteristic function of the underlying asset at maturity under the equivalent martingale measure. In particular, $B = \Phi_T(z)$ with $z = z - (\alpha+1)i$ so that:

$$
\psi_T(z) = \frac{e^{-r\tau} \Phi_T(z - i(\alpha+1))}{(iz+\alpha)(iz+\alpha+1)} \tag{4.28}
$$

To recover the undamped call premia, it is sufficient to apply an inverse Fourier transform and undo the exponential damping applied to make the function integrable — that is, the value of a European call option at time t is given by:

$$
\begin{aligned}
C_t(S_t, T, k) &= e^{-\alpha k} \mathfrak{F}^{-1}[\psi_T(z)] \\
&= \frac{e^{-\alpha k}}{2\pi} \int_{\mathbb{R}} e^{-izk} \psi_T(z)\, dz = \frac{e^{-\alpha k}}{\pi} \int_0^{\infty} \Re[e^{-izk} \psi_T(z)]\, dz
\end{aligned} \tag{4.29}
$$

where the last equality holds because $\psi_T(z)$ is even in its real part (Carr and Madan, 1999). Going back to integrability requirements, it is clear that $\psi_T(0)$ is finite as long as $\Phi_T(-(\alpha+1)i) < \infty$ — that is, for $c_t(S_t, T, k)$ to be (square) integrable, it must be that:

$$
\Phi_T(-(\alpha+1)i) < \infty \iff \mathbb{E}[\exp(X_T)^{\alpha+1}] = \mathbb{E}[S_T^{\alpha+1}] < \infty \tag{4.30}
$$

In other words, the damped call option will stay integrable on the positive half axis ($k > 0$) as long as the underlying price process exhibits finite and well-defined moments up to the $(1+\alpha)$-th order. As a result, the existence of the n-th moment of the underlying sets an upper bound for choosing the damping parameter α.

Unfortunately, as noted in Carr and Madan (1999), the premium charged for a call option converges to its intrinsic value (i.e., $[S_T - K]^+$) as one approaches the contract's expiration, thereby leading to a highly oscillatory integrand. In particular, for $k > X_t$ ($k < X_t$), let $z_T(k)$ be the price of a call (put) option maturing in $\tau \coloneqq T - t$. Without any loss of generality, let $S_t = 1$; then, $z_T(k)$ is given by:

$$z_T(k) = e^{-r\tau} \int_{\mathbb{R}} \left[(e^{X_t} - e^k) \mathbb{1}_{X_t<k,k<0} + (e^{X_t} - e^k) \mathbb{1}_{X_t>k,k>0} \right] \mathbb{Q}(X_T \mid \mathcal{F}_t)\, dX_T \tag{4.31}$$

Furthermore, $z_T(k)$ may be recovered through an inverse Fourier transform:

$$\xi_T(z) = \int_{\mathbb{R}} e^{izk} z_T(k)\, dk \iff z_T(k) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-izk} \zeta_T(u)\, du \tag{4.32}$$

Plugging $z_T(k)$ into the Fourier transform of the latter, Carr and Madan (1999) show that:

$$\xi_T(z) = e^{r\tau} \left(\frac{1}{1+iz} - \frac{e^{r\tau}}{iz} - \frac{\phi_T(z-i)}{z^2 - iz} \right) \tag{4.33}$$

The resulting inverse Fourier transform in (4.32) does not present invertibility problems at $\pm\infty$; for especially low values of τ, however, the integrand of $z_T(k)$ exhibits wide oscillations in the region around $k = 0$. Here, $z_T(0)$ resembles a Dirac delta function $\delta_k(0)$ as $T \to t$, as shown in Figure 4.2. To cope with such an issue, Carr and Madan (1999) propose to damp $z_T(k)$ employing the hyperbolic sine function $\sinh(x)$ evaluated at $x = \alpha k$. In this case, the value of α allows one to control the magnitude of the integrand in the region around zero; the option's premium then equals the following:

$$z_T(k) = \frac{1}{2\sinh(\alpha k)\pi} \int_{\mathbb{R}} e^{-izk} \left[\frac{\zeta_T(z - i\alpha) - \zeta_T(z + i\alpha)}{2} \right] dz \tag{4.34}$$

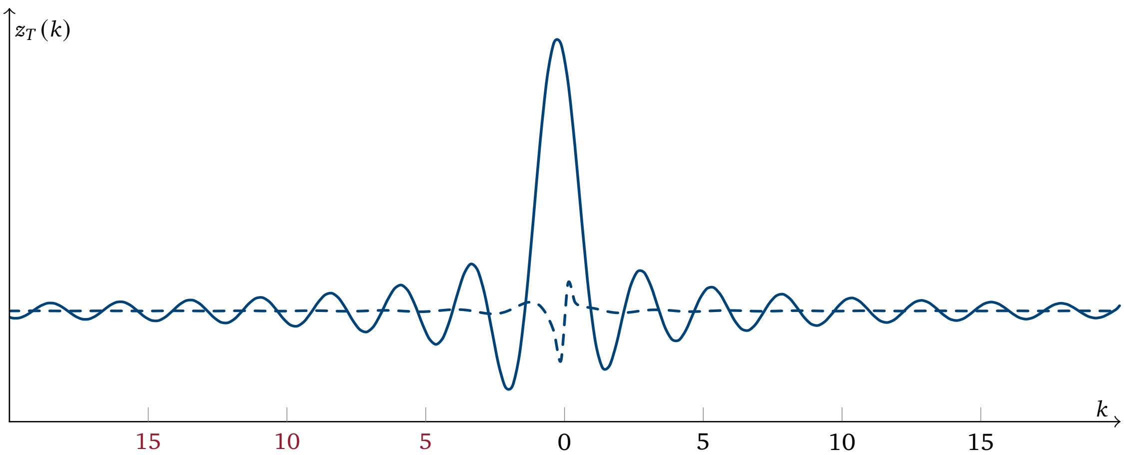


Figure 4.2 Time value $z_T(k)$ of a call expiring in five trading days (i.e., $T \approx 0.02$). The premium charged approaches the option's intrinsic value: with no hyperbolic sine damping (solid line), this leads to wide oscillations in the neighborhood of $k = 0$ and poor numerical integration results.

The damped time value proposed in (4.34) is much more well-behaved than that represented by (4.31), as highlighted by Figure 4.2.

Two years after the publication of Carr and Madan's work, Lewis (2001) generalised the authors' results by showing that integrating a damped function is equivalent to evaluating a contour integral in the complex plane. Underlying Lewis's work is the idea that the payoff function of an option admits a generalised Fourier transform $\Phi_t(z) = \mathbb{E}[e^{izX_t}]$ with $z \in \mathbb{C}$ in some strip of regularity S_X parallel to the $\mathbb{R}$-axis in the complex plane. In particular, $w(X_T) := [e^{X_T} - K]^+$ diverges as $X_T \to \infty$, but $w(X_T)e^{-\Im[z]X_T}$ becomes integrable for an appropriate choice of $z \in \mathbb{C}$. In turn, the latter expression is equivalent to the integration of the actual payoff over a contour parallel to the real axis:

$$\begin{aligned}\hat{w}(z) := \mathfrak{F}[w(X_T)] &= \int_{\mathbb{R}} e^{izX_T} w(X_T)\, dX_T \\ &= \int_{\mathbb{R}} e^{izX_T}[e^{X_T} - K]^+\, dX_T = \int_{\log(K)}^{\infty} e^{izX_T}(e^{X_T} - K)\, dX_T \\ &= \left[\frac{e^{(iz+1)X_T}}{iz+1} - K\frac{e^{izX_T}}{iz}\right]_{\log(K)}^{\infty} = -\left[\frac{K^{iz+1}}{iz+1} - K\frac{K^{iz}}{iz}\right] = -\frac{K^{iz+1}}{z^2 - iz}\end{aligned} \tag{4.35}$$

For the upper limit (i.e., $X_T \to \infty$) to exist, however, one needs $\Im[z] > 1$; in other words, the generalised Fourier transform of the call payoff is only defined in a strip of regularity S_w characterised by $\Im[z] > 1$ and presented in Figure 4.3. Put options share the same Fourier transform (4.35) of calls, but similar reasoning shows that S_w must be shifted to the region where $\Im[z] < 0$ for the transformed payoff $\hat{w}(z)$ to be well-behaved. In turn, given $\hat{w}(z)$, one can then recover the original payoff through an inverse Fourier transform:

$$w(X_T) = \mathfrak{F}^{-1}[\hat{w}(z)] = \frac{1}{2\pi}\int_{i\Im[z]-\infty}^{i\Im[z]+\infty} e^{-izX_T}\hat{w}(z)\, dz \tag{4.36}$$

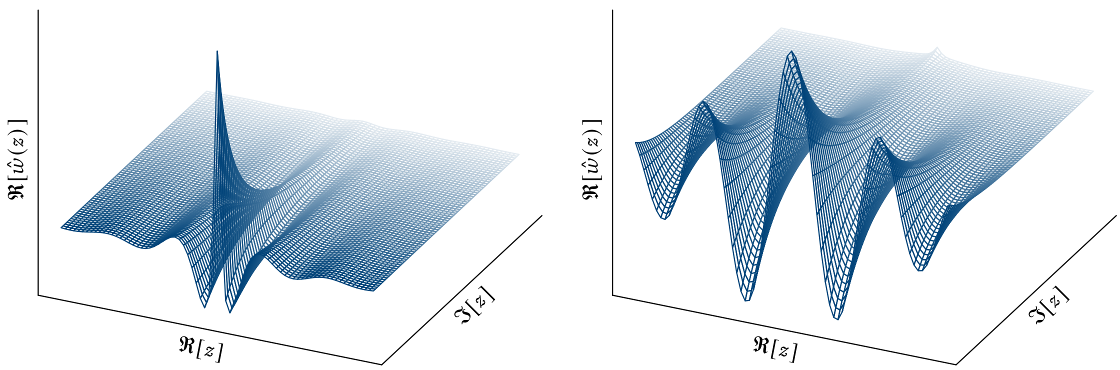


Figure 4.3 Regularity strips S_w for the call (left) and put (right) options payoffs in Fourier space. The functional form of $\hat{w}(z)$ is unchanged between calls and puts, but for the limits of the integral to be defined it must be that $\Im[z] > 1$ ($\Im[z] < 0$) for call (put) options.

Provided the risk-neutral price process has a well-defined characteristic function $\Phi_T(z)$ for some $z \in \mathcal{S}_x$, and the modified payoff $\hat{w}(z)$ is regular with $z \in \mathbb{C}$ in a given strip of regularity $\mathcal{S}_w$, one can recover the price of an option relying on the martingale condition:

$$\begin{aligned} V_t(w(X_T)) &= e^{-r\tau}\mathbb{E}_{\mathbb{Q}}[w(X_T)] = e^{-r\tau}\mathbb{E}_{\mathbb{Q}}\left[\int_{i\Im[z]-\infty}^{i\Im[z]+\infty} e^{-izX_T}\hat{w}(z)\,dz\right] \\ &= \frac{e^{-r\tau}}{2\pi}\int_{i\Im[z]-\infty}^{i\Im[z]+\infty} \mathbb{E}_{\mathbb{Q}}[e^{-izX_T}]\hat{w}(z)\,dz = \frac{e^{-r\tau}}{2\pi}\int_{i\Im[z]-\infty}^{i\Im[z]+\infty} \Phi_T(-z)\hat{w}(z)\,dz \end{aligned} \tag{4.37}$$

As mentioned at the beginning of the previous section, $\overline{\Phi_T(z)} = \Phi_T(-z)$; therefore, if $\Phi_T(z)$ is well-defined in $\mathcal{S}_x$, its conjugate appearing in (4.37) will share the same behaviour in the strip of regularity given by $\overline{\mathcal{S}_x}$. Furthermore, for the integral in (4.37) to be tractable, one must choose a strip $\mathcal{S}_v = \mathcal{S}_w \cap \overline{\mathcal{S}_x}$ where both $\Phi_T(-z)$ and $\hat{w}(z)$ are well-behaved. In fact, computing the integral along a contour in $\mathcal{S}_v$ guarantees the former's convergence and allows to bring the expectation inside the integral by Fubini's Theorem (Lewis, 2001). Notice that (4.37) is a straightforward application of the Parseval identity in (4.13) with $f = w(X_T)$ and $\overline{g} = \mathbb{Q}(X_T \mid \mathcal{F}_t)$ — that is, Lewis (2001) showed that:

$$e^{-r\tau}\int_{\mathbb{R}} w(X_T)\mathbb{Q}(X_T \mid \mathcal{F}_t)\,dX_T = \frac{e^{-r\tau}}{2\pi}\int_{i\Im[z]-\infty}^{i\Im[z]+\infty} \Phi_T(-z)\hat{w}(z)\,dz \tag{4.38}$$

It is worth emphasising that the valuation formula holds for several types of non-path-dependent European options; some of the transformed payoffs $\hat{w}(z)$ are reported in Table 4.1. For a call option, the fair premium to be charged is obtained by plugging (4.35)

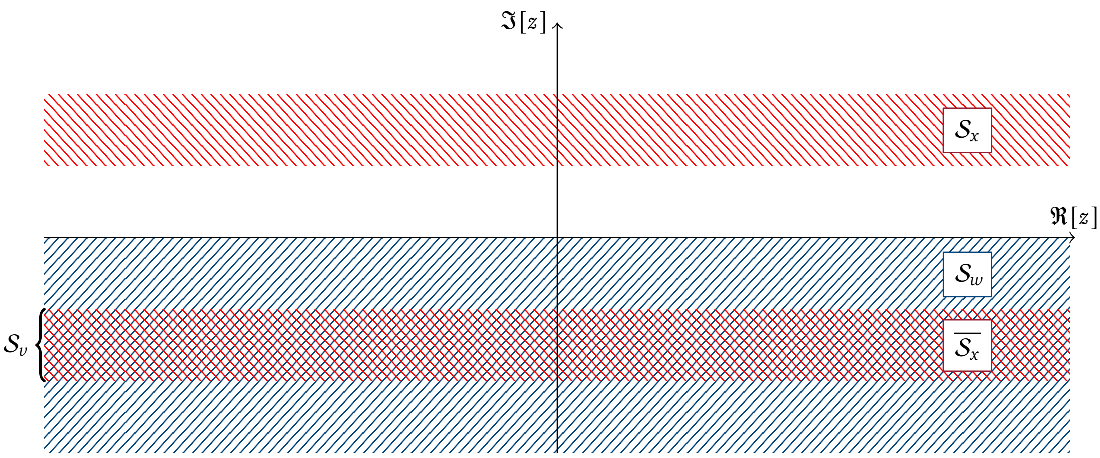


Figure 4.4 Regularity strip $\mathcal{S}_v$ for the premium charged on a European put option. The payoff function is analytic in the strip $\mathcal{S}_w$ defined by $\Im[z] < 0$; assuming the (generalised) characteristic function $\Phi_T(z)$ of the risk-neutral price process is well-behaved in some strip $\mathcal{S}_x$, then the regularity strip for the put price is found at the intersection $\mathcal{S}_v = \mathcal{S}_w \cap \overline{\mathcal{S}_x}$.

into (4.37) as to reach:

$$C_t(S_t, T, K) = -\frac{Ke^{-r\tau}}{2\pi} \int_{i\Im[z]-\infty}^{i\Im[z]+\infty} \frac{e^{-izk}\phi_T(-z)}{z^2 - iz}\, dz \tag{4.39}$$

where $k := \log(S_t/K) + r\tau$ represents a "dimensionless" measure of moneyness of the option. At this stage, Lewis (2001) relies on Cauchy's Residue Theorem to rewrite (4.39) as:

$$C_t(S_t, T, K) = S_t - \frac{Ke^{-r\tau}}{2\pi} \int_{i\Im[z]-\infty}^{i\Im[z]+\infty} \frac{e^{-izk}\phi_T(-z)}{z^2 - iz}\, dz \tag{4.40}$$

Shifting the contour by $\Im[z] = 0.5$ leads to an integration path which is equidistant from the two poles of the integrand — namely, $z = 0$ and $z = i$. In particular, setting $z = u + i/2$, (4.40) becomes:

$$\begin{aligned} C_t(S_t, T, K) &= S_t - \frac{\sqrt{S_t K}e^{r\tau/2}}{\pi} \int_0^\infty \Re\left[e^{-iuk}\phi_T\left(-u - \frac{i}{2}\right)\right] \frac{du}{u^2 + 1/4} \\ &= S_t - \frac{\sqrt{S_t K}e^{r\tau/2}}{\pi} \int_0^\infty \Re\left[e^{iuk}\phi_T\left(u - \frac{i}{2}\right)\right] \frac{du}{u^2 + 1/4} \end{aligned} \tag{4.41}$$

Finally, Lewis (2001) shows how moving the path of integration to the two poles given by $\Im[z] = 0$ and $\Im[z] = 1$ leads to the BSM-style formula (4.22) discussed by Bakshi and Madan (2000) and presented at the beginning of this section.

Contract	$w(X_T)$	$\hat{w}(z)$	S_w
Call	$[e^{X_T} - K]^+$	$-\frac{K^{iz+1}}{z^2 - iz}$	$\Im[z] > 1$
Put	$[K - e^{X_T}]^+$	$-\frac{K^{iz+1}}{z^2 - iz}$	$\Im[z] > 1$
Cash-or-Nothing Call	$\mathbb{1}_{e^{X_T} \geq K}$	$-\frac{K^{iz}}{iz}$	$\Im[z] > 0$
Cash-or-Nothing Put	$\mathbb{1}_{e^{X_T} \leq K}$	$\frac{K^{iz}}{iz}$	$\Im[z] < 0$
Asset-or-Nothing Call	$e^{X_T}\mathbb{1}_{e^{X_T} \geq K}$	$-\frac{K^{iz+1}}{iz + 1}$	$\Im[z] > 1$
Asset-or-Nothing Put	$e^{X_T}\mathbb{1}_{e^{X_T} \leq K}$	$\frac{K^{iz+1}}{iz + 1}$	$\Im[z] < 0$

Table 4.1 Generalised Fourier transforms $\hat{w}(z)$ for the payoff functions and strips of regularity S_w in the complex plane for some well-known non-path-dependent European options. Clearly, the strip of regularity S_v where the option value is well-behaved will depend on that for the generalised characteristic function of the risk-neutral price process.

4.3 | Calibration

Calibrating one of the valuation models seen so far amounts to finding the set of parameters Θ that minimises the deviation of the prices predicted by the model from those quoted on the market. In incomplete market models, the time series of the underlying is not sufficient to reach an estimated price surface for the options contracts; moreover, the equivalent martingale measure $\mathbb{Q}$ only shares qualitative characteristics of the true measure $\mathbb{P}$, such as the presence of jumps and, for instance, the finiteness of the Lévy measure (Cont and Tankov, 2003). Therefore, one must rely on an "implied" modelling technique where $\mathbb{Q}$ is found by calibrating the pricing model to the (most recent) surface of quoted option prices. Once the model parameters have been estimated, one can use the model to price exotic instruments and devise hedging strategies. In this sense, model calibration is the inverse of the pricing problem — that is, while the latter is concerned with valuing one or more options contracts given a set Θ of model parameters, the former is about finding Θ such that the model outputs a given set of prices. Specifically, given N options contracts with prices C_i $(i = 1, \ldots, N)$, the problem is finding the set of parameters under which the price process $\{\tilde{S}_t\}_{t\in[0,T]}$ defines a $\mathbb{Q}_\Theta$-martingale and

$$C_i = e^{-r\tau_i}\mathbb{E}_\Theta[[S_{T_i} - K_i]^+] \qquad \forall i \in \{1, \ldots, N\} \tag{4.42}$$

where $\mathbb{E}_\Theta[[S_{T_i} - K_i]^+]$ represents the expected intrinsic value of the option under the model identified by Θ. Clearly, a perfect calibration is not required as market frictions such as bid-ask spreads are common even in liquid options markets; hence, the main aim of the inverse pricing problem is obtaining the "best" approximation of market prices with a predetermined model. In turn, the quality of such approximation is usually intended in a (nonlinear) least-squares sense (Cont and Tankov, 2003; Schoutens et al., 2004), namely:

$$\Theta^* = \underset{\mathbb{Q}_\Theta \in \mathcal{Q}}{\arg\min} \sum_{i=1}^{N} w_i \left[C_\Theta(T_i, K_i) - C_i\right]^2 \tag{4.43}$$

where $C_\Theta(T_i, K_i)$ is the price of option C_i with strike K_i and maturity T_i under the given model with parameter set Θ and measure $\mathbb{Q}_\Theta$. As discussed in the first chapter, the presence of jumps or stochastic volatility in the price process leads to incomplete markets: the second fundamental theorem of asset pricing does not hold, and several risk-neutral measures $\mathbb{Q}_\Theta \in \mathcal{Q}$ do not introduce arbitrage into the market. While the minimisation problem in (4.43) is guaranteed to have solution(s), the quadratic pricing error is non-convex, may have several local minima, and features regions insensitive to changes in the model parameters. Furthermore, the optimum Θ^* is significantly dependent on the input parameters, prompting the need for some regularisation technique; for instance, Cont and Tankov (2003) incorporate a prior $\mathbb{P}_0$ by relying on *relative entropy* — a measure of divergence between two probability measures. This section implements a somewhat naive calibration and is instead devoted to describing the problem and the numerical procedure to solve it. The weights w_i in (4.43) should be related to the reliability of the option prices available; assuming liquid options have the most reliable data, such aim may be achieved by looking at the bid-ask spread for each contract and setting:

$$w_i = \frac{1}{C_i^{\text{Ask}} - C_i^{\text{Bid}}}$$

so that highly illiquid options will have little influence on the minimisation problem. The numerical implementation of the calibration problem relies on an influential algorithm known as *fast Fourier transform* (FFT), discussed below.

Suppose one is dealing with a given function f with values $\boldsymbol{f_k} = [f_0 \quad \dots \quad f_{N-1}]^T$ sampled at N discrete points $\boldsymbol{x} = [x_0 \quad \dots \quad x_{N-1}]^T$. The Fourier transform of such sample is well-defined and known as *discrete Fourier transform* (DFT)*; here, the representation in Fourier space of the k-th data point is given by:

$$\mathfrak{F}[f_k] = \sum_{j=0}^{N-1} f_j e^{i2jk\pi/N} \tag{4.44}$$

For each f_k, (4.44) sums integer multiples of $\omega := e^{2\pi i/N}$; hence, one can rely on a $N \times N$ square matrix to write the vector of Fourier coefficients $\mathfrak{F}[\boldsymbol{f_k}]$ as:

$$\begin{bmatrix} \mathfrak{F}[f_0] \\ \mathfrak{F}[f_1] \\ \mathfrak{F}[f_2] \\ \vdots \\ \mathfrak{F}[f_{N-1}] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)^2} \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \end{bmatrix} \tag{4.45}$$

Each of the N entries of $\mathfrak{F}[\boldsymbol{f_k}]$ requires N multiplications to be found, leading to a *computational complexity* of $O(N^2)$; luckily, a decomposition developed by Cooley and Tukey (1965), applied recursively, allows to reduce the number of operations to be performed to the order of $O(N\log(N))$; for this reason, Cooley and Tukey's algorithm is known as FFT.

In the context of option pricing, the FFT allows one to value options contracts for several strikes in a single run. Applying the transform to the value of an option requires a numerical approximation for the relevant integral; for instance, since the real part of

*The transform actually returns the weights in the Fourier *series* representation of the function. Letting the period go to ∞, one moves from a Fourier series to a Fourier transform.

a complex number is linear, the call price proposed by Lewis (2001) can be numerically approximated through the trapezoid rule as follows:

$$
\begin{aligned}
C_t(S_t,T,K) &= S_t - \frac{\sqrt{S_t K}e^{r\tau/2}}{\pi}\int_0^{\infty}\Re\left[e^{iuk}\phi_T\left(u-\frac{i}{2}\right)\right]\frac{du}{u^2+1/4} \\
&= S_t - \frac{\sqrt{S_t K}e^{r\tau/2}}{\pi}\Re\left[\int_0^{\infty}e^{iuk}\phi_T\left(u-\frac{i}{2}\right)\frac{du}{u^2+1/4}\right] \\
&\approx S_t - \frac{\sqrt{S_t K}e^{r\tau/2}}{\pi}\Re\left[\sum_{n=0}^{N-1}e^{iu_n k}\phi_T\left(u_n-\frac{i}{2}\right)\frac{\eta}{u_n^2+1/4}\right] \qquad (4.46)
\end{aligned}
$$

where η is the discrete increment of the Fourier variable $u_n = \eta n$, so that the integral is truncated at $N\eta$. It is worth mentioning that this approximation will induce both a *sampling error*, for the continuous Fourier variable u is sampled at N discrete points, and a *truncation error*, for the infinite integral is truncated at a finite value $N\eta$. If one's aim is to price in-the-money options, it is sufficient to focus on the region where $k \approx 0$. Assuming an interval $[-b, b]$ for the (log-)moneyness k with a regular step size equal to λ, one has:

$$k_j = -b + \lambda j \qquad (4.47)$$

where $b = N\lambda/2$. Substituting for (4.47) into (4.46) leads to:

$$C_t(S_t,T,K) \approx S_t - \frac{\sqrt{S_t K}e^{r\tau/2}}{\pi}\Re\left[\sum_{n=0}^{N-1}e^{iu_n(-b+\lambda k)}\phi_T\left(u_n-\frac{i}{2}\right)\frac{\eta}{u_n^2+1/4}\right] \qquad (4.48)$$

Moreover, since $u_n = \eta n$, one can rewrite (4.48) as:

$$
\begin{aligned}
C_t(S_t,T,K) &\approx S_t - \frac{\sqrt{S_t K}e^{r\tau/2}}{\pi}\Re\left[\sum_{n=0}^{N-1}e^{-ib\eta n + i\eta n\lambda k}\phi_T\left(u_n-\frac{i}{2}\right)\frac{\eta}{u_n^2+1/4}\right] \\
&= S_t - \frac{\sqrt{S_t K}e^{r\tau/2}}{\pi}\Re\left[\sum_{n=0}^{N-1}e^{-ibu_n}e^{i\eta n\lambda k}\phi_T\left(u_n-\frac{i}{2}\right)\frac{\eta}{u_n^2+1/4}\right] \qquad (4.49)
\end{aligned}
$$

For (4.49) to be a suitable input as input of a FFT routine, the (log-)moneyness spacing λ and the Fourier variable step size η must verify the identity:

$$\lambda\eta = \frac{2\pi}{N} \qquad (4.50)$$

It follows that the (log-)moneyness lies in the interval bounded by $\pm\pi/\eta$. Substituting for λ into (4.49), one reaches a DFT representation of the call price proposed by Lewis:

$$C_t(S_t,T,K) \approx S_t - \frac{\sqrt{S_t K}e^{r\tau/2}}{\pi}\Re\left[\sum_{n=0}^{N-1}e^{-ibu_n}e^{i2jn\pi/N}\phi_T\left(u_n-\frac{i}{2}\right)\frac{\eta}{u_n^2+1/4}\right] \qquad (4.51)$$

The identity shown in (4.50) entails an inverse relationship between the step size of the (log-)moneyness and the Fourier variable. Therefore, opting for a finer integration scheme by selecting a small η will lead to a sparser values of k which may not lie in the region of interest of the problem (i.e., around $k = 0$). Hence, it is desirable to incorporate some weighting scheme w_j into the discretised integral above; doing so, the prices of call options for several strikes can be efficiently computed by evaluating:

$$C_t(S_t, T, K) \approx S_t - \frac{\sqrt{S_t K} e^{r\tau/2}}{\pi} \Re \left[\sum_{n=0}^{N-1} w_j e^{-ibu_n} e^{i2jn\pi/N} \phi_T \left(u_n - \frac{i}{2} \right) \frac{\eta}{u_n^2 + 1/4} \right] \tag{4.52}$$

Two popular choices of integration rules are:

$$\text{Trapezoidal:} \quad w_j = \frac{\mathbb{1}_{j\in\{0,N-1\}}(j)}{2} + \mathbb{1}_{j\notin\{0,N-1\}}(j) \tag{4.53}$$

$$\text{Simpson's:} \quad w_j = \frac{3 + (-1)^j - \delta_j}{3} \tag{4.54}$$

where δ_j is the *Kronecker delta* function, defined as $\delta_j = \mathbb{1}_{j=0}(j)$. In other words, Simpson's integration rule assigns a weight $w_j = 1/3$ to the first and last points (i.e., $j = 0$ or $j = N-1$), a weight $w_j = 2/3$ if $j \equiv 0 \pmod 2$ (i.e., j is even), and a weight $w_j = 4/3$ if $j \equiv 1 \pmod 2$ (i.e., j is odd). A visualisation of the trapezoidal and Simpson's integration schemes is shown in Figure 4.5 for an arbitrary real-valued function. Finally, as the k_j-s lie on an equally spaced grid, one can rely on some interpolation technique to obtain prices for a continuum of strikes. In this context, some authors claim that the error introduced by a simple linear interpolation exceeds the sum of sampling and truncation error due to the discretisation of the evaluation formula (Chourdakis, 2005; McCulloch, 2003). In contrast, a cubic spline introduces much more negligible error. It is therefore advisable to choose a finer grid for the discretisation of the Fourier variable u as to reach more accurate results;

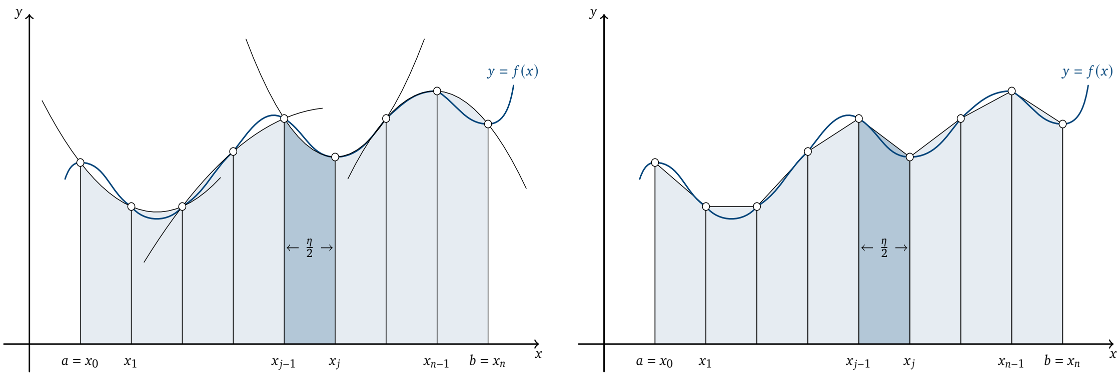


Figure 4.5 Left: Numerical approximation of the area under an arbitrary function $f(x)$ according to the Simpson's rule based on quadratic equations for each sub-interval. Right: approximated definite integral of $f(x)$ as computed through the trapezoidal rule based on trapezoids described by vertices at each bound of the sub-intervals in which the function is partitioned.

from (4.50) one knows that a smaller u-spacing will lead to few values of k in the region of interest, which can, however, be found by cubic interpolation.

One can then assess the performance of the pricing formulas relying on Fourier inversion discussed so far; in particular, Table 4.2 presents the results concerning computational time and pricing accuracy of the three methods for a number of call options ranging from 10 to 1000. Pricing accuracy is proxied by the *mean absolute percentage error* (MAPE) — that is, a lower MAPE is associated by construction to a larger pricing accuracy; for a data set of N observations X_i and respective forecasts $\hat{X}_i$, the MAPE is simply given by:

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^{N} \left| \frac{X_i - \hat{X}_i}{X_i} \right| \tag{4.55}$$

The choice of MAPE comes from its superior interpretability when compared to other accuracy measures such as *root mean squared error* (RMSE); indeed, a MAPE of x implies the predictions are, on average, off by x% with respect to the outstanding market prices.

	Number of options contracts					
	10	50	100	250	500	1000
Closed formula						
CPU Time	0.04927	0.21900	0.55028	1.17697	2.36880	4.58952
MAPE	—	—	—	—	—	—
Speed-up	—	—	—	—	—	—
Bakshi and Madan						
CPU Time	0.32445	0.06420	0.65664	1.61980	3.27012	6.49194
MAPE	3.51e-12	3.07e-12	3.01e-12	3.45e-12	3.02e-12	2.20e-09
Speed-up	0.15185x	3.41120x	0.83802x	0.72661x	0.72438x	0.70696x
Lewis						
CPU Time	0.05627	0.29824	0.59589	2.88720	2.96523	5.78536
MAPE	6.71e-12	6.74e-12	6.75e-12	6.82e-12	6.92e-12	7.78e-09
Speed-up	0.87550x	0.73429x	0.92346x	0.40765x	0.79886x	0.79330x
Fast Fourier transform						
CPU Time	0.03549	0.03665	0.03418	0.09848	0.03321	0.03370
MAPE	2.44e-07	2.72e-07	2.78e-07	3.16e-07	3.19e-07	3.19e-07
Speed-up	1.38832x	5.97582x	16.1822x	11.9515x	71.3259x	136.178x

Table 4.2 Computational time (in seconds) and mean absolute percentage error of different pricing methods based on Fourier inversion in the jump-diffusion model of Merton (1976). The FFT is implemented with $N = 2^{13} = 8192$ points and a discretisation step $\eta = 0.1$ for the Fourier variable u. Hence, the integral is truncated at $0.1 \times 8192 = 819.2$ and the k_j's are $\lambda = 0.007669$ apart.

Setting the closed-formula values as a benchmark, the last row of Table 4.2 highlights how relying on the algorithm proposed by Cooley and Tukey (1965) becomes increasingly beneficial as the number of contracts gets larger. Furthermore, all the three methods presented come with a significant pricing accuracy, irrespectively of the sample size; still, the reduction in computational time makes a FFT the desirable approach when pricing more than one contract (Cont and Tankov, 2003).

The actual calibration of the pricing models discussed in the previous chapters is conducted relying on the outstanding call options chain written on the *Apple* stock as of the 17$^{\text{th}}$ of August, 2022. The analysis could easily be conducted by referring to put contracts as well, but the put-call parity arbitrage relation ensures that the magnitude of mispricings with respect to the BSM framework are similar across the two contract speifications; as a result, the analysis which follows focuses on call options. Furthermore, even though the derivations of the models assumed a non-dividend paying underlying, the generalisation is straightforward: it is indeed sufficient to impose a correction in the drift of the spot (log-)price by letting it equal $r - q$ in the respective risk-neutral characteristic functions; at the time of the analysis, the *Apple* stock closed at a price of \$174.55, thereby implying a dividend yield of $q = 0.90/174.55 \approx 0.52\%$.

The contract with the earliest expiration matures on the 26/08/2022, while the last available option in terms of maturity refers to the 21/06/2024. The exercise prices range from a minimum of \$30 for the 20/01/2023 maturity up to a maximum of \$320 for the last expiration available. Even though the options prices are proxied by the bid-ask mid-point, it is worth noting that even on such a liquid instrument — that is, options written on the *Apple* stock — the bid-ask spread goes from \$0.01 to \$0.55. Overall, the data set contains 342 calls sharing 19 common strikes with a minimum of \$130 and a maximum of \$240. To render the calibration procedure (slightly) faster, a filter considering strikes lying in an interval bounded by ±25% of the spot price is applied, thereby leading to a price surface made up of 15 strikes for 18 different expirations.

The risk-free rate is assumed to be proxied by the par yield published daily by the US Treasury. Unfortunately, the Treasury only provides the yield for a predetermined set of maturities. Hence, in order to obtain the relevant rates for the options expirations in the data set, it is necessary to resort to yield curve models. Within the latter, a popular and proven efficient choice is the well-known framework of Nelson and Siegel (1987) and in particular its six-parameter extension proposed by Svensson (1994). Here, the risk-free rate (yield) for a given time $t \geq 0$ is given by:

$$r_t = \beta_1 + \beta_2 \left(\frac{1 - e^{-t/\lambda_1}}{t/\lambda_1} \right) + \beta_3 \left(\frac{1 - e^{-t/\lambda_1}}{t/\lambda_1} - e^{-t/\lambda_1} \right) + \beta_4 \left(\frac{1 - e^{-t/\lambda_2}}{t/\lambda_2} - e^{-t/\lambda_2} \right) \tag{4.56}$$

		All Contracts				Short-Term Contracts			
		JD	VG	SV	SVJD	JD	VG	SV	SVJD
Merton	σ	0.1613 (0.0041)	—	—	—	0.1985 (0.0037)	—	—	—
	λ	2.5746 (0.1683)	—	—	—	7.0621 (0.1829)	—	—	—
	α	0.1369 (0.0058)	—	—	—	0.1586 (0.0049)	—	—	—
	β	0.1620 (0.0022)	—	—	—	0.0936 (0.0017)	—	—	—
Madan et al.	θ	—	0.3678 (0.0194)	—	—	—	0.7911 (0.0236)	—	—
	σ	—	0.2768 (0.0018)	—	—	—	0.2664 (0.0023)	—	—
	ν	—	0.3031 (0.0034)	—	—	—	0.0783 (0.0037)	—	—
Heston	ρ	—	—	0.5519 (0.0070)	—	—	—	0.6301 (0.0074)	—
	σ	—	—	0.7863 (0.0239)	—	—	—	1.1498 (0.0314)	—
	θ	—	—	0.0520 (0.0002)	—	—	—	0.0107 (0.0005)	—
	κ	—	—	1.1646 (0.1632)	—	—	—	3.3791 (0.1743)	—
	v_0	—	—	0.1399 (0.0006)	—	—	—	0.0959 (0.0005)	—
Bates	ρ	—	—	—	0.4469 (0.0159)	—	—	—	0.5953 (0.0231)
	σ	—	—	—	0.6298 (0.0722)	—	—	—	0.7737 (0.0536)
	θ	—	—	—	0.0739 (0.0042)	—	—	—	0.1165 (0.0073)
	κ	—	—	—	2.8937 (0.3986)	—	—	—	4.3243 (0.5143)
	λ	—	—	—	0.8160 (0.1477)	—	—	—	0.8463 (0.1318)
	α	—	—	—	0.0624 (0.0056)	—	—	—	0.1107 (0.0053)
	β	—	—	—	0.0726 (0.0026)	—	—	—	0.0904 (0.0031)
	v_0	—	—	—	0.0672 (0.0046)	—	—	—	0.0977 (0.0047)
MAPE (%)		5.94	6.82	2.23	1.81	12.24	14.01	4.39	3.51

Table 4.3 Results of the calibration procedure for the four pricing models discussed so far. When it comes to pricing accuracy, all the frameworks provide a substantial improvement on the BSM framework, associated with 36.38% and 65.67% MAPEs for all and short-term contracts, respectively. The most significant contribution is undoubtedly due to the introduction of stochastic volatility paths, and allowing for discontinuities improves the fit even further.

The six parameters are then estimated by minimising the sum of squared errors between the model predictions and the yields published by the US Treasury. The yield curve obtained through such procedure is presented in Figure 4.6. Even though the fit is not perfect, especially for short tenors the Nelson-Siegel-Svensson interpolation technique provides an almost exact representation of the yields. Luckily, this portion of the curve is the most crucial for the calibration problem at hand, as most of the options considered in the data are set to expire within a year and all the contracts will be worthless in at most two years.

Given the data described above, considering (4.43) for each of the four models leads to the parameter sets and respective standard errors are presented in Table 4.3. In particular, the calibration procedure is carried out once for all the options contracts available and then focusing on short-term calls only; a contract is assumed to be short-term if it is supposed to expire within 100 trading days, or about 0.397 years.

Starting from the jump-diffusion model of Merton (1976), the results highlight how the options considered come with a significant implied jump intensity for the Poisson process in (2.40), with around five jumps every two years. Together with the estimate for the average log-jump size α, the jump-diffusion parameters seem to imply a large kurtosis and a substantial asymmetry for the risk-neutral density of log-returns on the *Apple* stock. Coherently, the results for the variance-gamma model proposed by Madan et al. (1998) show how the implied risk-neutral density is both significantly left-skewed and leptokurtic; in particular, the implied kurtosis is about 30% greater than it would be

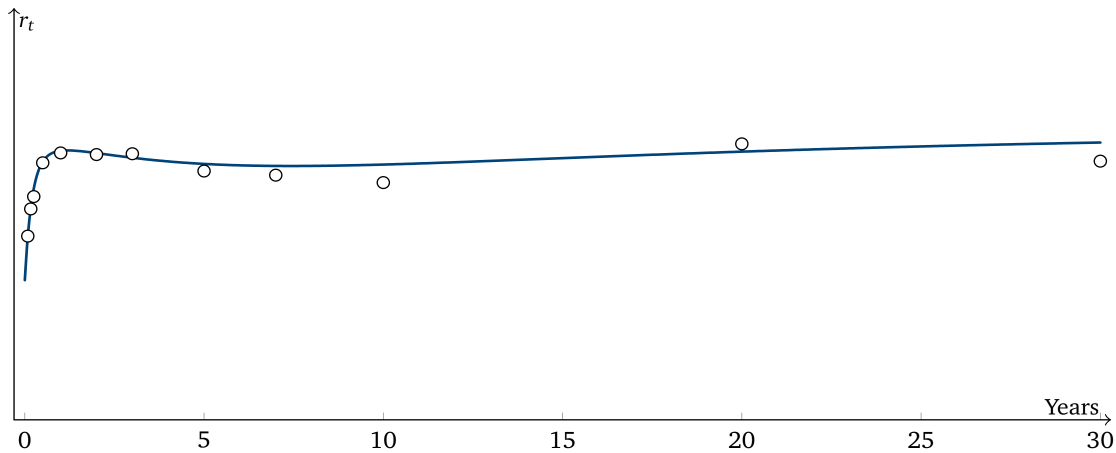


Figure 4.6 Yield curve obtained by applying the six-parameter Nelson-Siegel-Svennson model. The best fit is characterised by $\beta_1 = 0.0326$, $\beta_2 = -0.1939$, $\beta_3 = 0.0083$, $\beta_4 = -0.0174$, $\lambda_1 = 0.1707$, and $\lambda_2 = 4.4075$. Although overestimating the par yield at longer maturities, the model provides a significantly accurate representation of the curve in the time span of interest.

for a Normal distribution of returns, or 3.9. The parameter set Θ^* minimising the weighted sum of pricing errors for the stochastic volatility model of Heston (1993) confirm the presence of the leverage effect discussed at the beginning of Chapter 3, wherein returns and volatility are known to share a negative correlation. It is also known that the correlation ρ between volatility v_t and spot price S_t and the volatility of volatility σ are responsible for the skewness and kurtosis of the returns distribution, respectively. It is then clear that the implied distribution is asymmetrical and leptokurtic, as the Heston (1993) model requires a substantially high estimate of σ to justify the pronounced volatility skew observed in the options data for short maturities. Finally, turning to the stochastic volatility jump-diffusion model described by Bates (1996) one notes that the leverage effect is still present but with a lower magnitude, while the estimated speed of mean reversion κ is slightly higher. At the same time, the estimate for the volatility of volatility σ is smaller than that of the stochastic volatility relying on continuous paths only. Concerning the discontinuous component of the price process, the intensity of jumps, average log-jump size, and variance of the latter are all associated to estimates lower than those obtained for the jump-diffusion model assuming a constant volatility. Interestingly, the results for the Bates (1996) model seem to suggest that the jump-diffusion model justifies the implied volatility skew with greater estimates for the parameters of the compound Poisson process which may be instead the result of an additional source of randomness in the volatility process; similarly, the estimated Θ^* for the framework proposed by Heston (1993), coupled with the results observed for the Bates (2006) model, show that the high variability attached to v_t may actually be the result of a discontinuous component in the price process.

Focusing on short-term call contracts, one immediately notes how the estimate for the volatility coefficient — either for the spot price or its volatility — is inevitably higher for all the models considered. As a result, one can infer that a significantly more erratic path for the underlying's price or volatility is implied in the prices of short-term options contracts. Furthermore, both the stochastic volatility and stochastic volatility jump-diffusion models predict a (slightly) stronger leverage effect which, in the Bates (1996) framework, is coupled with more frequent jumps characterised by a significantly more negative average (log-)jump size, down by about 5% from the estimate obtained with the calibration to all the available contracts. A similar observation holds for the models allowing for discontinuity points in the price paths discussed in, wherein the more significant asymmetry proxied by a lower correlation coefficient ρ in the Heston (1993) and Bates (1996) models is translated to a much more negative estimate for α and θ in the jump-diffusion and (asymmetric) variance-gamma representations. Similarly, the greater volatility also leads to a larger kurtosis, as shown by the increase in the estimates for λ and ν. Hence, it follows that for the considered short-term contracts to be priced consistently by models allowing for discontinuous price paths, the magnitude of market crashes and their frequency must be greater. On the other hand, frameworks involving stochastic volatility attribute the greater variation potentially generated by jumps in the spot price to the presence of a stronger leverage effect and an even more erratic, or *rough*, volatility process.

Overall, the results of Table 4.3 show that the biggest improvement over the BSM model in terms of pricing accuracy is achieved through the explicit inclusion of a stochastic process dictating the dynamics of the underlying volatility. While providing a good fit for short-term options, models only allowing for discontinuities in the price path such as exponential Lévy processes suffer from their independent increments requirement, eventually leading to implied volatility surfaces which flatten out too quickly as the contracts expirations increase. Still, a generalisation of the class of jump-diffusion models accounting for stochastic volatility or, equivalently, an expansion of the stochastic volatility framework allowing for discontinuities in the price path — such as the model described by Bates (1996) — can indeed improve the fit to the volatility surface even further.

Conclusion

This thesis aimed to explore option pricing techniques that explicitly account for some of the most common properties observed in empirical data on asset prices and volatility. These observations are reflected in different data sets so frequently that they are now recognised as actual stylised facts about asset returns. As far as the arguments of this thesis are concerned, the most attractive property is the excess kurtosis and the left-asymmetry of the returns distribution; a leptokurtic distribution comes with heavy tails, meaning that "extreme" events do occur with a greater frequency than in the Gaussian specification. Such configuration for the densities implies that deep in-the-money and (especially) deep out-of-the-money options contracts are underpriced in the BSM framework, leading to the volatility smiles and skews implied in the market prices of options contracts.

Starting from one of the most influential contributions in quantitative finance — the fundamental partial differential equation developed by Black and Scholes (1973) and Merton (1973) — derived in the first chapter, the second and third parts are designed to provide the theoretical background underlying four of the most common pricing models entailing a nongaussian distribution of returns. An intuitive technique to increase the probability of extreme returns consists in allowing the price path of the underlying on which the option contract is written to exhibit discontinuities — that is, jumps. The latter may occur for various reasons: the announcement of an M&A transaction, the outbreak of a war (or pandemic), the discovery of a new drug by some biotech company, and similar. As a result, the stochastic differential equations describing the dynamics of the price paths are not only artificially associated with leptokurtic distributions but also reflect the significant changes observed in stock prices over a short time interval. One may argue that the key here would be to define how long a short time interval lasts. However, the argument would only favour the discontinuity hypothesis as jumps essentially dictate intraday movements. As described in the second chapter, càdlàg stochastic processes allowing for discontinuity points share most of the properties verified by the Brownian motion on which the BSM framework is built. In particular, the class of Lévy processes provides a generalisation wherein Brownian motion is the only continuous specification and may or may not be included in the final stochastic differential equation, as in the variance-gamma process proposed by Madan et al. (1998). In fact, the independence of increments features among the properties of Brownian motion present in the more general class of Lévy processes; as a result, while providing a good fit for options contracts expiring in the short term, the accuracy worsens and approximates that of BSM as the maturity of the contract moves further out in time. Such observation prompts the need to relax the assumption of independence, which is indeed absent in the second subset of pricing frameworks — based on stochastic volatility — presented in the third chapter. This class of models rejects the assumption of independent increments and constant volatility inherent to the BSM representation while addressing two additional stylised facts of asset returns, specifically about their variation. First, volatility tends to cluster — significant returns follow wide price swings while more minor variations often predict negligible returns — and second, volatility and returns share a negative correlation, an observation known as the leverage effect. Among the earliest and most significant contributions in this field is the Heston (1993) model, which proposes a stochastic representation of volatility that follows the same square root process used for interest rates by Cox et al. (1985). Heston's model has several advantages, including the presence of a closed-form solution for European option pricing. However, the assumption of continuous paths for the asset price and its volatility makes the model incapable of reproducing the pronounced smile or skew observed for short-term options contracts. In particular, the bivariate diffusion does not generate enough variation over short time intervals unless one assumes an unreasonably large value for the volatility of volatility. At the same time, however, the volatility surfaces at longer maturities match those typically observed on the market. An ideal specification for the price paths would still model volatility as stochastic while allowing for discontinuities, either in the price path or the volatility one (or both). Bates (1996) follows the first approach, introducing jumps in the spot price and combining the benefits of the Heston (1993) and Merton (1976) models while providing the fourth pricing framework discussed in this thesis. In such representation, the volatility surfaces implied in the artificially generated prices accurately reflects their empirical counterparts with reasonable parameter estimates.

Unfortunately, specifying a more realistic stochastic process to model the behaviour of a given underlying asset impact the analytical tractability of the model. The essential results obtained by Harrison and Pliska in 1981 and 1983 — namely, the two fundamental theorems of asset pricing — allow to price options without introducing arbitrage opportunities by computing an expectation under an equivalent martingale measure. However, as soon as (say) jumps enter the equation, the density needed to compute the expectation becomes much more involved by including infinite sums or special functions defined ad-hoc for the problem at hand. Nevertheless, as discussed in the fourth chapter, the approach pionereed by Stein and Stein (1991) and popularised by Heston (1993) to price via expectation without specifying the density was later generalised by the works of Bakshi and Madan (2000), Carr and Madan (1999), and Lewis (2001). In particular, each probability density function is entirely characterised by its characteristic function — a standard terminology in probability theory to denote the Fourier transform of a density. The relevant characteristic function can then be inverted to recover the original density, a result due to Gil-Pelaez (1951). While such a procedure may be seen as excess work to obtain the same result, working with characteristic functions allows specifying complex dynamics for the model and still price options without the need for Monte Carlo simulations. Most importantly, the Fourier transforms of the distribution induced by the log-price in the models discussed in this thesis have a simple closed-form representation due to the Lévy-Itô decomposition or the affine nature of the framework. Furthermore, posing the pricing problem in terms of characteristic functions allows one to rely on the fast Fourier transform — a recursive algorithm able to price thousands of options contracts in a matter of seconds. The results show how, when it comes to pricing accuracy, the most significant improvement over the BSM framework is obtained by introducing stochastic volatility in the model. Still, allowing for discontinuities, as in Bates (2006), permits achieving an even better performance.

Concerning the calibration procedure, a research limit of this thesis lies in the absence of an out-of-sample assessment of the results obtained — that is, it would be interesting to look at the pricing accuracy obtained with the parameters calibrated on a different, more extensive, options data set. Equally interestingly, one may assess the fit of an increasingly popular class of models relying on fractional Brownian motion — a process relaxing the continuity assumption — known as fractional stochastic volatility frameworks. Here, exploring variations of the rough fractional stochastic volatility model, based on a fractional Brownian motion with Hurst exponent below one-half, would be particularly engaging. Although present in a more elaborate representation, the ubiquitous need to abstract from the assumption of independent increments within these models validates the building blocks on which the models presented in this thesis are built.

References

Attari, Mukarram. Option Pricing Using Fourier Transforms: A Numerically Efficient Simplification. *SSRN Electronic Journal*, 2004.

Avramov, Doron, Chordia, Tarun, and Goyal, Amit. The Impact of Trades on Daily Volatility, 2004.

Bachelier, Louis. *Théorie de la Spéculation*. PhD thesis, Université de Paris, 1900.

Bakshi, Gurdip and Madan, Dilip. Spanning and derivative-security valuation. *Journal of Financial Economics*, 55(2):205–238, 2000.

Bates, David S. Jumps and Stochastic volatility: Exchange Rate Processes Implicity in Deutsche Mark Options. *The Review of Financial Studies*, 9(1):69–107, 1996.

Bates, David S. Maximum Likelihood Estimation of Latent Affine Processes. *Review of Financial Studies*, 19:909–965, 2006.

Bertrand, Joseph. *Calcul des probabilités*. Gauthier-Villars et fils, Paris, 1889.

Black, Fischer. Studies of Stock Price Volatility Changes. In *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, pages 177–181, Washington DC, 1976.

Black, Fischer and Scholes, Myron. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

Boness, James A. Elements of a Theory of Stock-Option Value. *Journal of Political Economy*, 72(2): 163–175, 1964.

Borel, Émile. *Leçons sur la théorie des fonctions*. Gauthier-Villars et fils, Paris, 1898.

Brown, Robert. A brief account of microscopical observations made on the particles contained in the pollen of plants. *Philosophical Magazine*, 4:161–173, 1828.

Calderone, Gregory. Options Being Used More for Speculation Than Traditional Hedging. *Bloomberg*, 2020.

Carr, Peter and Madan, Dilip. Option valuation using the fast Fourier transform. *Journal of computational finance*, 2(4):61–73, 1999.

Carr, Peter, Geman, Hélyette, Madan, Dilip, and Yor, Marc. The Fine Structure of Asset Returns: An Empirical Investigation. *The Journal of Business*, 75(2):305–332, 2002.

Cherubini, Umberto, Lunga, Giovanni Della, Mulinacci, Sabrina, and Rossi, Pietro. *Fourier Transform Methods in Finance*. John Wiley & Sons, 2010.

Chourdakis, Kyriakos. Option Pricing Using the Fractional FFT. SSRN Scholarly Paper 707566, Social Science Research Network, Rochester, NY, 2005.

Clark, Peter K. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica*, 41(1):135–155, 1973.

Cont, Rama and Tankov, Peter. *Financial Modelling with Jump Processes*. Chapman and Hall/CRC, 2003.

Cooley, James W. and Tukey, John W. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

Cox, John C. and Ross, Stephen A. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3(1):145–166, 1976.

Cox, John C., Ingersoll, Jonathan E., and Ross, Stephen. A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407, 1985.

Donsker, Monroe David. An invariance principle for certain probability limit theorems. *Memoirs of the American Mathematical Society*, 6:12, 1951.

Du Bois-Reymond, Paul. *Die allgemeine Funktionentheorie*. Tübingen, 1882.

Dupire, Bruno. Pricing with a Smile. *Risk Magazine*, pages 18–20, 1994.

Einstein, Albert. On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. *Annalen der Physik*, 17:549–560, 1905.

Eisenberg, Larry and Jarrow, Robert. Option pricing with random volatilities in complete markets. FRB Atlanta Working Paper 91, Federal Reserve Bank of Atlanta, 1991.

Fama, Eugene Francis. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2):383–417, 1970.

Feller, William. Two Singular Diffusion Problems. *Annals of Mathematics*, 54(1):173–182, 1951.

Fristedt, Bert and Gray, Lawrence. *A Modern Approach to Probability Theory*. Birkhäuser, Boston, 1997.

Gatheral, Jim. *The Volatility Surface: A Practitioner's Guide*. John Wiley & Sons, 2011.

Georg Cantor. Ueber unendliche, lineare Punktmannichfaltigkeiten. *Mathematische Annalen*, 21 (4):545–591, 1883.

Gil-Pelaez, J. Note on the Inversion Theorem. *Biometrika*, 38(3):481–482, 1951.

Hankel, Hermann. Untersuchungen über die unendlich oft oscillirenden und unstetigen Functionen. *Mathematische Annalen*, 20:63–112, 1882.

Harrison, J. Michael and Pliska, Stanley R. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260, 1981.

Harrison, J. Michael and Pliska, Stanley R. A stochastic calculus model of continuous trading: Complete markets. *Stochastic Processes and their Applications*, 15(3):313–316, 1983.

Haug, Espen Gaarder. The History of Option Pricing and Hedging. In Hafner, Wolfgang and Zimmermann, Heinz, editors, *Vinzenz Bronzin's Option Pricing Models: Exposition and Appraisal*, pages 471–486. Springer, Berlin, Heidelberg, 2009.

Heston, Steven L. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies*, 6(2):327–43, 1993.

Hull, John and White, Alan. The Pricing of Options on Assets with Stochastic Volatilities. *The Journal of Finance*, 42(2):281–300, 1987.

Hull, John C. *Options, Futures, and Other Derivatives*. Pearson, 11th edition, 2021.

Itô, Kiyosi. On stochastic processes (I). *Japanese Journal of Mathematics*, 18:261–301, 1941.

Itô, Kiyosi. On a formula concerning stochastic differentials. *Nagoya Mathematical Journal*, 3: 55–65, 1951.

Jordan, Camille. Remarques sur les intégrales définies. *Journal de Mathématiques Pures et Appliquées*, 8:69–100, 1892.

Kallenberg, Olav. *Random Measures, Theory and Applications*. Probability Theory and Stochastic Modelling. Springer International Publishing, 2017.

Kolmogorov, Andrej Nikolaevič. *On Analytical Methods In Probability Theory*, pages 62–108. Springer Netherlands (1992), Dordrecht, 1931.

Kolmogorov, Andrej Nikolaevič. *Foundations of the Theory of Probability*. Chelsea Publishing Company (1950), 1933.

Lebesgue, Henri. *Integrale, Longueur, Aire*. PhD thesis, Université de Paris, 1902.

Lewis, Alan L. A Simple Option Formula for General Jump-Diffusion and Other Exponential Levy Processes. Technical report, Social Science Research Network, Rochester, NY, 2001.

Lévy, Paul. *Calcul des Probabilités*. Gauthier-Villars et fils, Paris, 1925.

Lévy, Paul. Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze*, 3-4:337–366, 1934.

Madan, Dilip and Milne, Frank. Option pricing with VG martingale components. *Mathematical Finance*, 1(4):39–55, 1991.

Madan, Dilip and Seneta, Eugene. The Variance Gamma Model for Share Market Returns. *The Journal of Business*, 63(4):511–524, 1990.

Madan, Dilip, Carr, Peter, and Chang, Eric. The Variance Gamma Process and Option Pricing. *European Finance Review*, 2:79–105, 1998.

Mandelbrot, Benoit. The Variation of Certain Speculative Prices. *The Journal of Business*, 36(4): 394–419, 1963.

McCulloch, J. Huston. The Risk-Neutral Measure and Option Pricing under Log-Stable Uncertainty. Working Papers 03-07, Ohio State University, Department of Economics, 2003.

Merton, Robert Cox. Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

Merton, Robert Cox. Option Prices When Underlying Stock Returns Are Discontinuous. *Journal of Financial Economics*, 3:125–144, 1976.

Nelson, Charles and Siegel, Andrew F. Parsimonious modeling of yield curves. *The Journal of Business*, 60(4):473–489, 1987.

Peano, Giuseppe. *Applicazioni geometriche del calcolo infinitesimale*. Cornell University Library Historical Math Monographs. Bocca, Torino, 1887.

Riemann, Bernard. Über die Darstellbarkeit einer Function durch eine trigonometrische Reihe. *Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 13:87–132, 1868.

Rogers, Chris L. G. and Satchell, Steve E. Estimating Variance From High, Low and Closing Prices. *The Annals of Applied Probability*, 1(4):504–512, 1991.

Samuelson, Paul Anthony. Rational Theory of Warrant Pricing. *Industrial Management Review*, 6 (2), 1965.

Sato, Ken-iti. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge, United Kingdom, 1st edition, 1999.

Schilling, René L. *Measures, Integrals and Martingales*. Cambridge University Press, 2017.

Schoutens, Wim, Simons, Erwin, and Tistaert, Jurgen. A Perfect Calibration! Now What? *Wilmott*, 2004.

Schroder, Mark. Computing the Constant Elasticity of Variance Option Pricing Formula. *The Journal of Finance*, 44(1):211–219, 1989.

Schwert, William G. Why Does Stock Market Volatility Change Over Time? Working Paper 2798, National Bureau of Economic Research, 1988.

Scott, Louis O. Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application. *Journal of Financial and Quantitative Analysis*, 22(4):419–438, 1987.

Shapiro, Samuel Sanford and Wilk, Martin. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4):591–611, 1965.

Sprenkle, Case. Warrant prices as indicators of expectations and preferences. *Yale economic essays*, 1(2):179–232, 1961.

Stein, Elias M. and Stein, Jeremy C. Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *Review of Financial Studies*, 4:727–752, 1991.

Svensson, Lars. Estimating and Interpreting Forward Interest Rates: Sweden 1992 - 1994. NBER Working Papers 4871, National Bureau of Economic Research, 1994.

Taleb, Nassim Nicholas. *Antifragile: Things That Gain From Disorder (Incerto)*, chapter 12, pages 352–355. Random House Publishing Group, 1st edition, 2012.

Thiele, Thorvald Nicolai. *Sur la compensation de quelques erreurs quasi-systématiques par la méthode des moindres carrés*. Reitzel, Copenhague, 1880.

Thorp, Edward Oakley. Optimal Gambling Systems for Favorable Games. *Revue de l'Institut International de Statistique*, 37(3):273–293, 1969.

Wiener, Norbert. Differential-Space. *Journal of Mathematics and Physics*, 2(1):131–174, 1923.

Appendix A

A.1 | Measure Theory

The discovery by Hankel (1882) of sets of zero extent (or *null sets*), as well as their close relationship (Du Bois-Reymond, 1882) to the *integrability* conditions of a function proposed by Riemann (1868), led the German mathematician and pioneer of *set theory* Georg Cantor (1883) to provide the first definition of the *measure* (*Inhalt*) of an arbitrary set. Intuitively, a null set can be *covered* by a countable union of intervals of arbitrarily small total length. A formal definition can be found in Null Sets. The notion of measure proposed by Georg Cantor (1883) was substantially generalised by Peano (1887) and Jordan (1892). Hence, such a measure is often called the Peano-Jordan (PJ) measure. Even though the definition of PJ measure is suitable for any dimension, the definition below refers to a plane — that is, $\mathbb{R}^2$, as in Jordan (1892).

Definition A.1. *Let S be the sum of the areas of those (closed) squares of the grid located inside the set E. Let S' be the sum of areas of those squares which contain at least one boundary point of the set E. The sum $S + S'$ is the total area of the squares containing points of the closure $E + E'$. As the grid diameter tends to zero, the numbers S and $S + S'$ tend to limits; the first of the limits is called the inner measure of the set E, and the second is known as the outer measure of the set E. When these values agree, the set is called measurable in the Peano-Jordan sense (or PJ-measurable), and the shared value of the inner and outer measures is called the Peano-Jordan measure of the set E.*

A substantial contribution to the field of measure theory is due to the French mathematician Borel (1898), who provided a series of descriptive postulates that became the standard for defining measures of sets.

Definition A.2. *The postulates contained in Borel (1898) to define a measure are as follows.*

1. *A measure is always nonnegative.*

2. *The measure of a sum of a finite number of (non-overlapping) sets equals the sum of their measures.*

3. *The measure of the difference of two sets (a set and a subset) equals the difference of their measures.*

4. *Every set whose measure is not zero is uncountable.*

Note that the second and fourth conditions, taken together, imply that a *singleton* — that is, a set with a single element — has Borel measure zero.

Borel, however, did not manage to show that the first postulate was always satisfied; these deficiencies were later made up by Lebesgue (1902), who provided a rigorous algorithm to construct the class of sets obtained by successive application of addition and subtraction on open sets. Lebesgue proposed to call the members of this class *Borel sets*.

Definition A.3. *The σ-algebra $\sigma(\mathcal{O}^n)$ generated by the open sets $\mathcal{O}^n$ of $\mathbb{R}^n$ is the Borel σ-algebra, and its members are the Borel (measurable) sets; the notation $\mathcal{B}(\mathbb{R}^n)$ indicates the (set of) Borel sets in $\mathbb{R}^n$.*

An intuitive example of a measure is the *Lebesgue measure* $\lambda(\cdot)$ on $\mathbb{R}^n$, defined on $\mathcal{B}(\mathbb{R}^n)$ and corresponding to the n-dimensional notion of volume: $\lambda(A) = \int_A dx = \int \mathbb{1}_A \, dx$. In the one-dimensional case, the notion of Lebesgue measure of a measurable set $A \subset \mathcal{B}(\mathbb{R})$ reduces to its "length" — that is, $\lambda([a,b]) = \lambda((a,b)) = b - a$. As a result, the Lebesgue measure can be used to construct the uniform probability measure as long as one restricts the domain to $[0,1]$. Another well-known measure is the *Dirac measure* $\delta_x(\cdot)$ associated to a point $x \in \mathbb{R}$ and defined to be $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise.

A measure μ_0 assigning a value of zero to any single point is said to be *diffuse* (or *atomless*). The Lebesgue measure defined above is a clear example of diffuse measure; on the contrary, Dirac measures are by definition positive only in a finite number of points. It is therefore natural to introduce the notion of *Radon measure* $\mu(\cdot)$, defined on a measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and such that for every bounded closed subset $B \in \mathcal{B}(\mathbb{R}^n)$ the Radon measure is finite, $\mu(B) < \infty$. Any Radon measure μ can be decomposed as the sum of a diffuse measure μ_0 and a linear combination of Dirac measures (Kallenberg, 2017):

$$\exists\, x_j \in \mathbb{R}, b_j \in \mathbb{R}_+ : \mu = \mu_0 + \sum_{j=1}^{\infty} b_j \delta_{x_j}$$

If μ_0 happens to equal zero, then the Radon measure μ reduces to a linear combination of Dirac measures and is called a *purely atomic* measure.

Definition A.3 mentions a fundamental concept for probability theory applied to continuous-time financial models, where collections of σ-algebras (or σ-fields) are used to model the information flowing to an agent; Fama (1970) is just one example showing the importance of modelling information in finance applications.

Definition A.4. *A σ-algebra $\mathcal{A}$ on a non-empty set X is a family of subsets of X such that:*

1. *The non-empty set X belongs to $\mathcal{A}$*
$$X \in \mathcal{A} \tag{A.1}$$

2. *If a set A belongs to $\mathcal{A}$, its complement $A^c := X \setminus A$ is also in $\mathcal{A}$*
$$A \in \mathcal{A} \implies A^c \in \mathcal{A} \tag{A.2}$$

3. *If a sequence of sets $A_1, A_2, \ldots$ belongs to $\mathcal{A}$, their union is also in $\mathcal{A}$*
$$\{A_j\}_{j\in\mathbb{N}_+} \subseteq \mathcal{A} \implies \bigcup_{j\in\mathbb{N}_+} A_j \in \mathcal{A} \tag{A.3}$$

A set $X \in \mathcal{A}$ is said to be ($\mathcal{A}$-)measurable.

Three properties of σ-algebras follow from Definition A.4 and are listed below.

1. The empty set $\emptyset$ belongs to $\mathcal{A}$: $\emptyset = X^c \in \mathcal{A}$ by (A.1) and (A.2)

2. If A and B are two sets in $\mathcal{A}$, then $A \cup B \in \mathcal{A}$: set $A_1 = A$, $A_2 = B$, and $A_3 = A_4 = \cdots = \emptyset$. Then, $A \cup B = \bigcup_{j\in\mathbb{N}_+} A_j \in \mathcal{A}$ by (A.3)

3. If a sequence of sets $A_1, A_2, \ldots$ belongs to $\mathcal{A}$, their intersection is also in $\mathcal{A}$; indeed, if $A_j \in \mathcal{A}$ then $A_j^c \in \mathcal{A}$ by (A.2), hence $\bigcup_{j\in\mathbb{N}_+} A_j^c \in \mathcal{A}$ by (A.3) and, again by (A.2) we conclude that $\bigcap_{j\in\mathbb{N}_+} A_j = \left(\bigcup_{j\in\mathbb{N}_+} A_j^c\right)^c \in \mathcal{A}$.

Clearly, the *power set* (i.e., the set of all subsets) 2^X of X represents the maximal σ-field, whereas the minimal σ-algebra on X is always given by $\mathcal{A} = \{\emptyset, X\}$.

Since every subset of $\mathbb{R}^n$ encountered in this thesis belongs to $\mathcal{B}(\mathbb{R}^n)$ — that is, it is Borel-measurable — it is worth introducing the notion of σ-algebra generated by a set.

Definition A.5. *Let X be a non-empty set. For every system of sets $\mathcal{M} \subseteq 2^X$ there exists a unique minimal σ-algebra containing $\mathcal{M}$. This is known as the σ-field generated by $\mathcal{M}$, denoted by $\sigma(\mathcal{M})$, and $\mathcal{M}$ is called its generator.*

In light of Definition A.5, the Borel σ-algebra on $\mathbb{R}^n$ shall be interpreted as the minimal σ-field generated by the class $\mathcal{O}^n$ of open sets of $\mathbb{R}^n$.

To make the reasoning less abstract, consider $X = \{a, b, c, d\}$ and $\mathcal{M} = \{\{a\}, \{b\}\} \subset X$. The subset $\mathcal{M}$ is clearly not a σ-algebra (e.g., it does not contain neither $\emptyset$ nor X); in fact, the "smallest" σ-field generated by $\mathcal{M}$ is given by:

$$\sigma(\mathcal{M}) = \Big\{\emptyset\,,\ X\,,\ \underbrace{\{a\}\,,\ \{b\}}_{\mathcal{M}}\,,\ \underbrace{\{a,b\}}_{\{a\}\cup\{b\}}\,,\ \underbrace{\{c,d\}}_{[\{a\}\cup\{b\}]^c}\,,\ \underbrace{\{b,c,d\}}_{\{a\}^c}\,,\ \underbrace{\{a,c,d\}}_{\{b\}^c}\Big\}$$

To conclude this brief section on measure theory it is essential to mention one last notion, which, once we restrict our focus to measure spaces with total *mass* equal to one (i.e., probability spaces), defines the concept of *random variables*.

Definition A.6. *Let $(X, \mathcal{A})$ and $(X', \mathcal{A}')$ be two measurable spaces. A map $T : X \to X'$ is said to be ($\mathcal{A}/\mathcal{A}'$-)measurable if the pre-image of every measurable set is itself a measurable set:*

$$X^{-1}(A') := \{x : X(x) \in A'\} \in \mathcal{A} \qquad \forall A' \in \mathcal{A}' \tag{A.4}$$

As presented in Section 1.1, the concepts defined so far have broad applications in probability theory: an *event* is just a subset of the σ-algebra $\mathcal{F}$ defined on the *sample space* Ω, *random variables* are measurable maps from a *probability space* to a measurable space, the *expectation* of a random variable requires its (Lebesgue) integral over the sample space with respect to the probability measure, and so forth.

A.2 | Null Sets

A formal definition of null set requires the notion of *cover* of a set, presented below.

Definition A.7. *A cover of a set X represents a collection of sets whose union includes X as a subset. If $C = \{U_\alpha : \alpha \in A\}$ is an indexed family of sets U_α, then C is a cover of X if*

$$X \subseteq \bigcup_{\alpha \in A} U_\alpha \tag{A.5}$$

We say that C is an open cover if each of its members is an open set.

Definition A.7 allows us to introduce the *null* set, which does not necessarily coincide with the *empty* set and whose description requires the existence of a sequence of open covers of a given set A for which the limit of the lengths of its covers is zero.

Definition A.8. *The set A is called a set of extent zero if for every $\varepsilon > 0$ there exists a finite system of intervals $\{U_n\}_n$ of total length smaller than ε covering the set A. Analytically, suppose A is a subset of the real line $\mathbb{R}$ such that*

$$\forall \varepsilon > 0, \exists \{U_n\}_n : U_n = (a_n, b_n) \subset \mathbb{R} : A \subset \bigcup_{n=1}^{\infty} U_n \wedge \sum_{n=1}^{\infty} |U_n| < \varepsilon \tag{A.6}$$

where the U_n are intervals and $|U|$ is the "length" of U, then A is a null set.

Sets of zero extent are heavily related to the integrability conditions of a function set forth by Riemann (1868). In particular, let us define the set $E_\alpha := E_x \{\omega(f, x) > \alpha\}$, where $\omega(f, E) := \sup_E f(x) - \inf_E f(x)$ and thus the *oscillation* $\omega(f, x_0)$ of f around a point x_0 as:

$$\lim_{\varepsilon \to 0} \omega(f, [x_0 - \varepsilon, x_0 + \varepsilon]) \tag{A.7}$$

In other words, E_α defines the set of the point(s) x for which the oscillation of $f(\cdot)$ is greater than α. In 1882, Du Bois-Reymond showed that if a function is such that, for each $\alpha > 0$, the set E_α can be included in a finite system of intervals of an arbitrarily small total length, then Riemann's integrability criterion is satisfied and vice versa. The results of Du Bois-Reymond (1882) shed light on the relationship between integration and the concept of measure: a bounded function is integrable if the set E_α is null for all $\alpha > 0$.

A.3 | Itô Isometry

Following Itô (1951), given a Brownian motion $W_t(\omega)$ we define:

$$I(f) := \int_0^T f_t(\omega)\, dW_s \tag{A.8}$$

The statement that Brownian motion is nowhere differentiable almost surely amounts to saying that its paths exhibit *unbounded variation* almost surely on any interval. Itô was aware that not all continuous processes could be integrated, and he thus focused on stochastic processes that are adapted to the *natural filtration* of the Brownian Motion. Let us then characterise the class $M_2[0,T]$ of functions f that admit an Itô integral.

Definition A.9. *A function $f\colon [0,\infty)\times\Omega \mapsto \mathbb{R}$ belongs to the $M_2[0,T]$ class if:*

- *$f_t(\omega)$ is $\mathcal{B}\times\mathcal{F}$-measurable*
- *$f_t(\omega)$ is non-anticipative*
- $\mathbb{E}\left[\int_0^T f_t^2(\omega)\, dt\right] < \infty$

Definition A.10. *A function $\phi \in M_2[0,T]$ is called a simple process if*

$$\phi_t(\omega) = \sum_{j=0}^{n-1} \alpha_j(\omega)\mathbb{1}_{[t_j,t_{j+1}]}(t) \tag{A.9}$$

for some partition $\Pi = \Pi(t_0, t_1, \ldots, t_n)$ with $0 = t_0 < t_1 < \cdots < t_n = T$. If $\phi \in M_2[0,T]$ is a simple process, then its Itô integral is given by

$$I(f) := \int_0^T \phi_t(\omega)\, dW_t = \lim_{n\to\infty} \sum_{j=0}^{n-1} \alpha_j(\omega)\left[W_{t_{j+1}} - W_{t_j}\right] \tag{A.10}$$

Let ϕ_1 and ϕ_2 be two simple processes in $M_2[0,T]$. Then:

1. $\int_0^T (\phi_{1,t} + \phi_{2,t})\, dW_t = \int_0^T \phi_{1,t}\, dW_t + \int_0^T \phi_{2,t}\, dW_t$

2. $\int_0^T c\phi_{1,t}\, dW_t = c\int_0^T \phi_{1,t}\, dW_t$ for all $c \in \mathbb{R}$
3. $\mathbb{E}\left[\int_0^T \phi_{1,t}\, dW_t\right] = 0$
4. $\mathbb{E}\left[\int_0^T \phi_{1,t}\, dW_t \cdot \int_0^T \phi_{2,t}\, dW_t\right] = \mathbb{E}\left[\int_0^T (\phi_{1,t}\cdot\phi_{2,t})\, dt\right]$

A proof of the last property, known as Itô isometry, for the case of simple processes follows.

Proof. We start with the case $\phi_1 = \phi_2 = \phi$, that is

$$\begin{aligned}
\mathbb{E}\left[\int_0^T \phi_t\, dW_t \cdot \int_0^T \phi_{2,t}\, dW_t\right] &= \mathbb{E}\left[\left(\int_0^T \phi_t\, dW_t\right)^2\right] \\
&= \mathbb{E}\Bigg[\sum_{j=0}^{n-1} \alpha_j^2(\omega)\left[B_{t_{j+1}} - B_{t_j}\right]^2 \\
&\quad \underbrace{+\ 2\sum_{0\le j<k\le n-1} \alpha_j(\omega)\left[B_{t_{j+1}} - B_{t_j}\right]\alpha_k(\omega)\left[B_{t_{k+1}} - B_{t_k}\right]}_{C}\Bigg]
\end{aligned}$$

Notice that:

- $\mathbb{E}\left[(B_{t_{j+1}} - B_{t_j})^2\right] = t_{j+1} - t_j$, since $(B_{t_{j+1}} - B_{t_j}) \sim \mathcal{N}(0, t_{j+1} - t_j)$
- For $k > j$, we have $\alpha_j(\omega)\left[B_{t_{j+1}} - B_{t_j}\right]\alpha_k(\omega) \perp\!\!\!\perp \left[B_{t_{k+1}} - B_{t_k}\right]$. Hence, $C = 0$.

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T \phi_t\, dW_t\right)^2\right] &= \mathbb{E}\left[\sum_{j=0}^{n-1} \alpha_j^2(\omega)\left[B_{t_{j+1}} - B_{t_j}\right]^2\right] \\
&= \sum_{j=0}^{n-1} \mathbb{E}\left[\alpha_j^2(\omega)\left[B_{t_{j+1}} - B_{t_j}\right]^2\right] \\
&= \sum_{j=0}^{n-1} \mathbb{E}\left[\alpha_j^2(\omega)\right]\mathbb{E}\left[\left[B_{t_{j+1}} - B_{t_j}\right]^2\right] \\
&= \mathbb{E}\left[\sum_{j=0}^{n-1} \alpha_j^2(\omega)(t_{j+1} - t_j)\right] \\
&= \mathbb{E}\left[\int_0^T \phi_t^2(\omega)\, dt\right] \qquad \text{(A.11)}
\end{aligned}$$

For the general case $\phi_1 \neq \phi_2$, we start by defining ϕ_3 as the linear combination $\phi_1 + \phi_2$. Hence,

$$\mathbb{E}\left[\left(\int_0^T \phi_{3,t}\, dW_t\right)^2\right] = \mathbb{E}\left[\left(\int_0^T (\phi_{1,t} + \phi_{2,t})\, dW_t\right)^2\right] = \mathbb{E}\left[\left(\int_0^T \phi_{1,t}\, dW_t + \int_0^T \phi_{2,t}\, dW_t\right)^2\right]$$

Expanding the square inside the expectation, we reach

$$\mathbb{E}\left[\left(\int_0^T \phi_{3,t}\, dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T \phi_{1,t}^2\, dt\right] + 2\mathbb{E}\left[\int_0^T \phi_{1,t}\, dW_t \int_0^T \phi_{2,t}\, dW_t\right] + \mathbb{E}\left[\int_0^T \phi_{2,t}^2\, dt\right]$$

At the same time, however, we also know the following:

$$\mathbb{E}\left[\left(\int_0^T \phi_{3,t}\, dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T \phi_{3,t}^2\, dt\right] = \mathbb{E}\left[\int_0^T (\phi_{1,t}^2 + 2\phi_{1,t}\phi_{2,t} + \phi_{2,t}^2)\, dt\right] \tag{A.12}$$

where the first equality holds by Itô Isometry for a (single) simple process — that is, (A.11). Again, by exploiting the linearity of the expectation operator we have

$$\mathbb{E}\left[\left(\int_0^T \phi_{3,t}\, dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T \phi_{1,t}^2\, dt\right] + 2\mathbb{E}\left[\int_0^T \left(\phi_{1,t} \cdot \phi_{2,t}\right) dt\right] + \mathbb{E}\left[\int_0^T \phi_{2,t}^2\, dt\right] \tag{A.13}$$

Finally, equating (A.12) and (A.13) we obtain:

$$\mathbb{E}\left[\int_0^T \phi_{1,t}\, dW_t \cdot \int_0^T \phi_{2,t}\, dW_t\right] = \mathbb{E}\left[\int_0^T (\phi_{1,t} \cdot \phi_{2,t})\, dt\right]$$

□

A.4 | Properties of Fourier Transform

Differentiation is conveniently translated to multiplication in the Fourier space; consider the Fourier transform of the derivative of an integrable function $f(x)$:

$$\mathfrak{F}\left[f(x)\right] = \int_{\mathbb{R}} e^{iux} f(x)\, dx \implies \mathfrak{F}\left[\frac{d}{dx} f(x)\right] = \int_{\mathbb{R}} e^{iux} \frac{d}{dx} f(x)\, dx \tag{A.14}$$

The integral above can be solved by integration by parts — that is, $\int f dg = fg - \int g df$. Here, letting $f(x) = e^{iux}$ and $dg(x)\, dx = df(x)$ so that the transform can be rewritten as:

$$\mathfrak{F}\left[\frac{d}{dx} f(x)\right] = \int_{\mathbb{R}} e^{iux} \frac{d}{dx} f(x)\, dx = \underbrace{\left[f(x)e^{iux}\right]_{-\infty}^{+\infty}}_{A} - \underbrace{\int_{\mathbb{R}} f(x)(iu)e^{iux}\, dx}_{B} \tag{A.15}$$

Concerning A, the norm of e^{iux} can be at most one for any $x, u \in \mathbb{R}$; at the same time, since $f(x)$ is absolutely integrable we have $\int_{\mathbb{R}} |f(x)|\, dx < \infty$ which entails that $f(x) \to 0$ as $x \to \pm\infty$. In other words, $A = 0$. The second term, B, features an integral with respect to x. This allows us to bring iu outside, thereby obtaining:

$$\mathfrak{F}\left[\frac{d}{dx} f(x)\right] = -\int_{\mathbb{R}} f(x)(iu)e^{iux}\, dx = -(iu)\int_{\mathbb{R}} e^{iux} f(x)\, dx = -(iu)\mathfrak{F}\left[f(x)\right] \tag{A.16}$$

The convolution of two (probability density) functions $f(x)$ and $g(x)$ is denoted $(f * g)(x)$ and defined by the following integral:

$$(f * g)(x) = \int_{\mathbb{R}} f(x-s)g(s)\,ds \tag{A.17}$$

In the Fourier space, convolutions are mapped to products — that is, $\mathfrak{F}[f * g] = \mathfrak{F}[f]\,\mathfrak{F}[g]$:

$$\begin{aligned}
\mathfrak{F}^{-1}[\mathfrak{F}[f]\,\mathfrak{F}[g]] &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iux}\left(\mathfrak{F}[f]\,\mathfrak{F}[g]\right)du = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iux}\mathfrak{F}[f]\left(\int_{\mathbb{R}} e^{ius}g(s)\,ds\right)du \\
&= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{iu(s-x)}\mathfrak{F}[f]\,g(s)\,ds\,du = \int_{\mathbb{R}} g(s)\underbrace{\frac{1}{2\pi}\int_{\mathbb{R}} e^{iu(s-x)}\mathfrak{F}[f]\,du}_{\mathfrak{F}^{-1}[\mathfrak{F}[f]]=f}\,ds \\
&= \int_{\mathbb{R}} f(x-s)g(s)\,ds
\end{aligned} \tag{A.18}$$

□

A.5 | Gil-Pelaez CDF

The representation of a cumulative distribution function proposed by Gil-Pelaez (1951) and given by (4.7) can be obtained through a convolution of the density $f(x)$ with the *signum* function $sgn(x)$, defined as $sgn(x) = -\mathbb{1}_{x<0}(x) + \mathbb{1}_{x>0}(x)$; moreover, it is known that $\mathfrak{F}[sgn(x)] = 2(iu)^{-1}$. Consider now the convolution of the density $f_X(x)$ with $sgn(x)$:

$$\begin{aligned}
(f_X * sgn)(x) &= \int_{\mathbb{R}} f(x+s)sgn(s)\,ds \\
&= \int_{-\infty}^{0} -f_X(x+s)\,ds + \int_{0}^{\infty} f_X(x+s)\,ds = 1 - 2F_X(x)
\end{aligned} \tag{A.19}$$

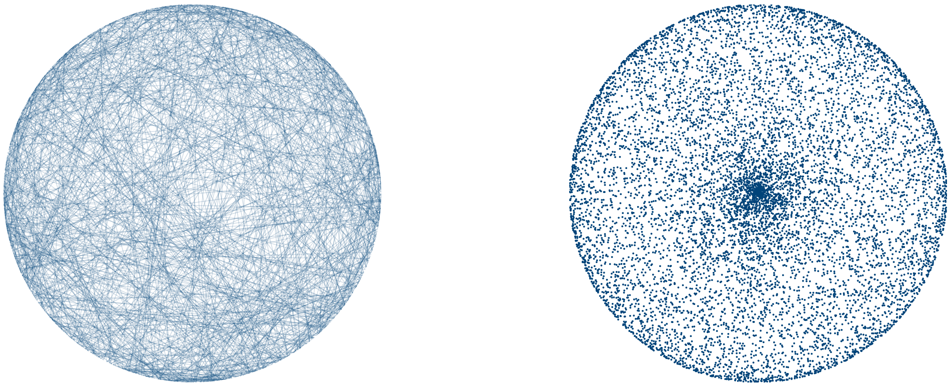
At the same time, we also know that convolutions are mapped onto products in the Fourier space — that is, $(f_X * sgn)(x)$ is also given by:

$$(f_X * sgn)(x) = \mathfrak{F}^{-1}\left[\mathfrak{F}[f(x)]\mathfrak{F}[sgn(x)]\right] = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{2e^{-izx}}{iz}\Phi_X(z)\,dz \tag{A.20}$$

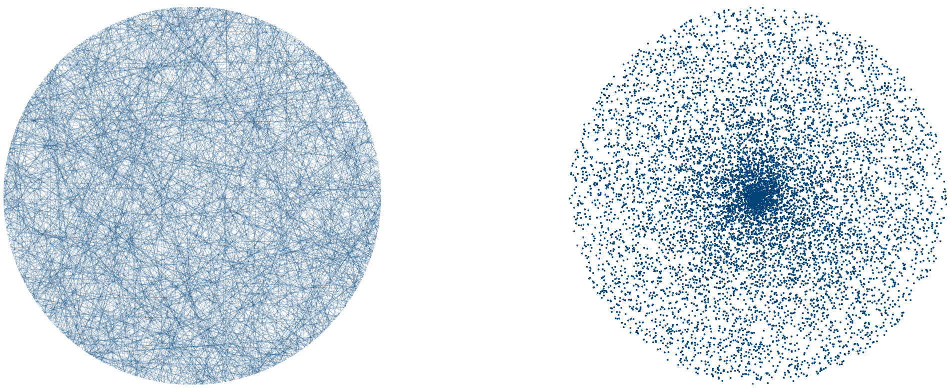
Hence, equating (A.19) and (A.20), one obtains (4.7):

$$1 - 2F_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{2e^{-izx}}{iz}\Phi_X(z)\,dz \iff F_X(x) = \frac{1}{2} - \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{-izx}\Phi_X(z)}{iz}\,dz \tag{A.21}$$

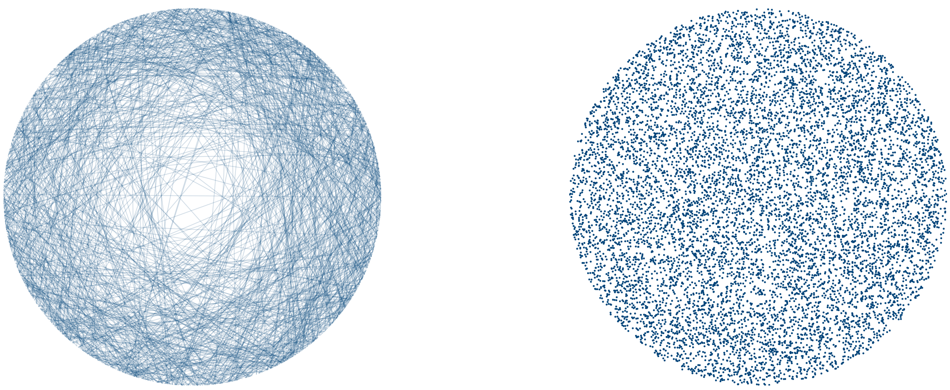
□



(a) Chords (left) and midpoints (right) chosen according to the "random endpoints" method: choose two random points on the circumference of the circle and draw the chord joining them



(b) Chords (left) and midpoints (right) chosen according to the "random radial point" method: choose a radius, choose a point on the radius and construct the chord perpendicular to it



(c) Chords (left) and midpoints (right) chosen according to the "random midpoint" method: choose a point anywhere within the circle and construct a chord with the former as a midpoint

Figure A.1 Consider an equilateral triangle inscribed in a circle, and suppose a chord is chosen at random; Bertrand (1889) asks what is the probability p of picking a chord longer than a side of the triangle. One can show that three different criteria, all in line with the *principle of indifference*, produce three different results: $p = 1/3$ in (a), $p = 1/2$ in (b), and $p = 1/4$ in (c).

Beyond Black Scholes: Fourier Transform Methods for European Option Pricing

Thesis Summary

Marco Paroli

Supervised by Professor Marco Pirra

Co-supervised by Professor Nicola Borri

Department of Economics and Management

Academic Year 2021/22

Chair of Asset Pricing

LUISS Guido Carli

October 2022

A thesis submitted in partial fulfilment of the requirements for the degree of Corporate Finance.

Summary

In quantitative finance, a *derivative* represents a contract whose value is *derived* from other underlying variables (Hull, 2021). Such instruments date back to ancient times: the first derivative was likely dealt with by the Greek philosopher Thales of Miletus around 600 BC. Generally engaged in philosophising, philosophers certainly did not shine for their wealth. An anecdote in Aristotle's *Politics* recounts how Thales, to prove the fallacy of this commonplace, acquired the seasonal use of every nearby olive press at a discount. The subsequent harvest, which proved exceptionally fruitful, allowed him to dispose of the olive presses on his own terms, accumulating significant wealth in the process. Hence, Thales was the first to set up a contingent contract depending upon the realisation of an underlying state variable (here, the abundance of the harvest).

According to Aristotle, "*from his knowledge of astronomy he had observed while it was still winter that there was going to be a large crop of olives...*". Thales would therefore have used the derivative contract for speculative purposes, which is predominant in the use of such financial instruments today. Another attractive perspective, provided by Taleb (2012), is instrumental in introducing the other primary reason for derivative contracts: *hedging*. From this perspective, Thales would have positioned himself to take advantage of his lack of knowledge rather than exploit his superior astronomy understanding: "*he did not need to understand too much the messages from the stars.*"

S.1 | Risk-Neutral Valuation

The insight backing the success of Thales' strategy lies in the asymmetry of his position, characterised by a finite but significant upside potential against a fixed cash outflow. Indeed, the most significant contribution provided by options is undoubtedly the introduction of asymmetry in the payoff, bounded at zero from below; clearly, the intrinsically contingent nature of options contracts comes at a cost. The existence of options may be rationalised through the two primary scopes such contracts are adopted in — speculation and hedging.

The former is quite evident with options contracts, generally "controlling" for one hundred shares of the underlying — for instance, the holder of a (say) call will buy 100 shares of asset S for a price K, as long as $S_T - K > 0$. Such built-in leverage makes options contracts especially suitable for speculative purposes. Risk management represents another field featuring widespread use of options, which may be employed to hedge a position and comply with regulatory requirements set forth by banking authorities (e.g., *EBA*), such as *Value at Risk* (VaR). Determining the value provided by asymmetry, known as the option's premium, requires an accurate description of asset prices. In turn, such modelling assumptions are based on a compendium of formal definitions from the theory initiated by Kolmogorov in 1933 for probability fields.

Definition S.1. *Let Ω be a nonempty set, and let $\mathcal{F}$ be a σ-algebra on subsets of Ω. Following Kolmogorov (1933), a probability measure $\mathbb{P}$ is a function that, to every set $A \in \mathcal{F}$, assigns a value in $[0, 1]$ called the probability of A and denoted $\mathbb{P}(A)$. We require $\mathbb{P}(\Omega) = 1$ and σ-additivity — that is, when $\{A_j\}_{j \in \mathbb{N}_+}$ is a sequence of pairwise disjoint events in $\mathcal{F}$ we have:*

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}\left(A_j\right) \tag{S.1}$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.

Definition S.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a measurable space. A random variable X is a measurable map $X \colon \Omega \to \mathbb{R}$. Equivalently, X is a random variable if for every Borel-measurable set $B \in \mathcal{B}(\mathbb{R})$ the pre-image of X is $\mathcal{F}$-measurable:*

$$\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F} \tag{S.2}$$

Definition S.3. *Let $X \colon \Omega \to \mathbb{R}$ be a real-valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution measure of X is defined as the map $\mu_X \colon \mathcal{B}(\mathbb{R}) \to [0, 1]$ which assigns to each Borel set $B \subset \mathcal{B}(\mathbb{R})$ a mass $\mu_X(B)$, where:s*

$$\mu_X(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B) \tag{S.3}$$

Definition S.4. *The characteristic function $\Phi_X \colon \mathbb{R} \to \mathbb{C}$ of the $\mathbb{R}$-valued random variable X is defined as the Fourier transform of the distribution μ_X:*

$$\Phi_X(z) = \mathbb{E}\left[\exp(izX)\right] = \int_{\mathbb{R}} e^{izx}\, d\mu_X(x) \tag{S.4}$$

for all $z \in \mathbb{R}$.

Just as the value of an asset at a given point in time $t \geq 0$ is represented by a RV, the dynamics of such value are proxied by a *stochastic process* — that is, a collection $\{X_t\}_{t \geq 0}$ of random variables indexed by time t. Most importantly, a stochastic process may be seen as a function $X\colon [0,T] \times \Omega \to \mathbb{R}$ of both time $t \in [0,T]$ and the randomness $\omega \in \Omega$.

Definition S.5. *Let Ω be a non-empty set. Let T be a fixed positive number, and assume that for each $t \in [0,T]$ there exists a σ-algebra $\mathcal{F}_t$. Assume that if $s < t$, then every set in $\mathcal{F}(s)$ is also found in $\mathcal{F}_t$ — that is, $s < t \implies \mathcal{F}(s) \subseteq \mathcal{F}_t$. Then, for $t \in [0,T]$, we call the collection $\{\mathcal{F}_t\}_{t\in[0,T]}$ of non-decreasing σ-algebras a filtration.*

Definition S.6. *A stochastic process $\{X_t\}_{t\in[0,T]}$ is said to be a martingale if*

1. *The process $\{X_t\}_{t\in[0,T]}$ is $\mathcal{F}_t$-adapted*
2. *The expectation $\mathbb{E}[|X_t|]$ is finite for all $t \in [0,T]$*
3. *For all $s < t$, $\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$*

Definition S.6 makes sense only with respect to a filtration and once the probability measure $\mathbb{P}$ has been specified. When more than one mapping $\mathcal{F} \to [0,1]$ is considered, the term $\mathbb{P}$-martingale clarifies the measure under which the process is a martingale.

Brownian motion (BM) is named after the Scottish botanist Robert Brown, who observed the jittery behaviour of pollen particles suspended in water, and may be considered the continuous counterpart of a (rescaled) *random walk* with the following properties:

1. It is continuous.
2. It starts, without loss of generality, at zero.
3. It has independent and stationary increments.
4. At time t it behaves as a Normal distribution $\mathrm{N}(0,t)$.

Unfortunately, specifying a BM for the spot price leads to some undesirable outcomes, such as negative equity values. Assuming such a process for the instantaneous return gives much more realistic results and coincides with the approach followed by Black and Scholes (1973) and Merton (1973). Applying the Itô-Doeblin formula to a geometric BM reveals that the spot price at t is log-normally distributed, ruling out negative values. The breakthrough introduced by Black-Scholes-Merton (BSM) was not the pricing formula itself but rather its derivation. The extension of Δ-hedging from a discrete to a continuous framework, coupled with the no-arbitrage principle, allowed the authors to exclude the underlying's drift from the formula and thus apply *risk-neutral valuation*. A hedged portfolio long an option valued at $f(S_t,t)$ and short $\partial f/\partial S$ units of the underlying must earn the risk-free rate r, leading to the fundamental pricing *partial differential equation* (PDE):

$$\frac{\partial f(S_t,t)}{\partial t} + \frac{\partial f(S_t,t)}{\partial S_t} r S_t + \frac{1}{2}\frac{\partial^2 f(S_t,t)}{\partial S_t^2}\sigma^2 S_t^2 = r f(S_t,t) \tag{S.5}$$

which can be solved once the terminal condition — $f(S_T,T) = [S_T - K]^+$ for a call — is specified. The absence of arbitrage opportunities is closely linked to the existence of a probability measure $\mathbb{Q}$, *equivalent* to $\mathbb{P}$, under which the process followed by the discounted underlying's price defines a martingale. Two measures $\mathbb{P}$ and $\mathbb{Q}$ are said equivalent (denoted $\mathbb{P} \sim \mathbb{Q}$) if they share the same impossible events:

$$\mathbb{P} \sim \mathbb{Q} \iff [\mathbb{P}(A) = 0 \iff \mathbb{Q}(A) = 0] \quad \forall A \in \mathcal{F} \tag{S.6}$$

The absence of arbitrage and the existence of $\mathbb{Q}$ are related by the *first fundamental theorem of asset pricing* due to Harrison and Pliska (1981).

Theorem S.1. *Consider a market model defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathcal{F}_t$. Let $\{S_t^i\}_{t\in[0,T]}$ represent the price process followed by asset S^i. Then, the market does not admit arbitrage opportunities if and only if there exists a probability measure $\mathbb{Q} \sim \mathbb{P}$ such that the discounted price process(es) $\{\tilde{S}_t^i\}_{t\in[0,T]}$ define $\mathbb{Q}$-martingales.*

Once the dynamics of the spot price under $\mathbb{Q}$ are known, the premium charged for an option with terminal payoff H is the expected payoff under $\mathbb{Q}$: $V_t(H) = e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[H \mid \mathcal{F}_t]$. Computing such expectation amounts to specifying the risk-neutral dynamics dictating the underlying's price movements. In the BSM framework, the (log-)return on the underlying at any time $t \in [0,T]$ is normally distributed:

$$X_t \sim \mathrm{N}\left(X_0 + \left(r - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right) \tag{S.7}$$

It is well-known, however, that stock returns do not obey a Gaussian probability law, as Mandelbrot (1963) pointed out.

S.2 | Lévy Processes

Solving the fundamental PDE with the terminal condition $f(S_T,T) = [S_T - K]^+$, the resulting price of a European call option maturing in $\tau := T - t$ is given by:

$$V_t^{BS} = S_t \Phi\left[\frac{\log(S_t/K) + (r+\sigma^2/2)\tau}{\sigma\sqrt{\tau}}\right] - Ke^{-r\tau}\Phi\left[\frac{\log(S_t/K) + (r-\sigma^2/2)\tau}{\sigma\sqrt{\tau}}\right] \tag{S.8}$$

This equation, caeteris paribus, defines a monotonically increasing function of the underlying's volatility σ, mapping $(0,+\infty)$ into $([S_t - Ke^{-r\tau}]^+, S_t)$. Hence, given a market price V_t^M

one can invert the pricing formula and recover the unique value of Σ_t reconciling the BSM and market prices:

$$\exists!\ \Sigma_t(K,T) > 0 : V_t^{BS}(S_t, K, \tau, r, \Sigma_t(K,T)) = V_t^M(K,T) \tag{S.9}$$

The mapping $\Sigma_t\colon (K,T) \to \Sigma_t(K,T)$ is known as the *implied volatility surface* at time t. If the market followed the assumptions set forth by Black and Scholes (1973) and Merton (1973), one would expect a flat surface across strikes and maturities. However, this is not the case once one looks at the empirically extracted volatility surface. Given the above, it is natural to move on to a generalisation of the BSM framework — the class of Lévy processes.

Definition S.7. *Let $X\colon [0,T]\times\Omega \to \mathbb{R}$ be a real-valued (càdlàg) stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, $\{X_t\}_{t\in[0,T]}$ defines a Lévy process if and only if:*

1. *It starts, without loss of generality, at zero.*
2. *It has independent and stationary increments.*
3. *It is stochastically continuous — that is, $\forall \varepsilon > 0, \lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| \geq \varepsilon) = 0$.*

Letting $\{X_t\}_{t\geq 0}$ be a Lévy process, there exists a continuous map $\psi\colon \mathbb{R} \to \mathbb{C}$, known as the *characteristic exponent* of X, such that:

$$\Phi_{X_t}(z) = \mathbb{E}[\exp(izX_t)] = \exp(t\psi(z)) \quad \forall z \in \mathbb{R} \tag{S.10}$$

The assumption of stationary increments implies that the CF of X_t is linear in time t. Brownian motion is the only example of continuous Lévy process; the other "building block" needed to reach a general formulation of Lévy processes, thereby allowing for discontinuities in the resulting paths, is the *Poisson process*.

Definition S.8. *Let $\{\tau_i\}_{i\geq 1}$ be a collection of i.i.d. exponentially distributed random variables and define $T_n \coloneqq \sum_{i=1}^n \tau_i$. Then, the process $\{N_t\}_{t\geq 0}$ given by*

$$N_t = \sum_{n\geq 1} \mathbb{1}_{t\geq T_n} = \#\{n \geq 1 : T_n \in [0,t]\} \tag{S.11}$$

is known as Poisson process with parameter λ.

Given a càdlàg process $\{Y_t\}_{t\geq 0}$ and an arithmetic Brownian motion $\{\gamma t + W_t\}_{t\geq 0}$ independent of $\{Y_t\}_{t\geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the process described by $X_t = \gamma t + W_t + Y_t$ defines a Lévy process and can be thus decomposed as:

$$X_t = \gamma t + W_t + \sum\nolimits_{s\in[0,t]} (Y_s - Y_{s-}) = \gamma t + W_t + \int_0^t \int_{\mathbb{R}} y J_Y(ds \times dy) \tag{S.12}$$

Theorem S.2. *Let $\{X_t\}_{t\geq 0}$ be a real-valued Lévy process with characteristic triplet given by (γ, σ^2, ν). Then, the characteristic function of the process at time t is given by:*

$$\Phi_{X_t}(z) = \mathbb{E}[\exp(izX_t)] = \exp\left\{t\left[i\gamma z - \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}}\left(e^{izx} - 1 - izx\mathbb{1}_{|x|<1}\right)\nu(dx)\right]\right\} \tag{S.13}$$

The first author to take account of discontinuities in price paths is was Merton (1976). In his derivation, a (compound) Poisson process, independent of the diffusive component, is included into the dynamics of the underlying. Hence, the instantaneous return on the underlying is given by the following *stochastic differential equation* (SDE):

$$\frac{dS_t}{S_t} = (\mu - \lambda\kappa)\,dt + \sigma\,dW_t + \left(\prod_{j=1}^{dN_t} Y_j - 1\right) \tag{S.14}$$

with $\kappa := \mathbb{E}[Y_t - 1] = \exp\left(\alpha + \beta^2/2\right) - 1$. Applying a generalised Itô-Doeblin formul allowing for discontinuities with $X_t := \log(S_t)$ reveals that the log-price follows:

$$dX_t = \left(\mu - \lambda\kappa - \frac{\sigma^2}{2}\right)dt + \sigma\,dW_t + \sum_{j=1}^{dN_t}\log(Y_j) \tag{S.15}$$

Consequently, the distribution of returns will inevitably be characterised by heavy tails and thus be more representative of the financial reality.

One does not need to introduce a diffusive component to reach an accurate representation of the evolution of asset prices. Letting $\{Z_t\}_{t\geq 0}$ be a Brownian motion with drift θ, the variance-gamma process $\{X_t\}_{t\geq 0}$ proposed by Madan et al. (1998) is defined as:

$$X_t(\omega) = Z_{\gamma_t(1,\nu)}(\omega) = \theta\gamma_t(1,\ \nu) + \sigma W_{\gamma_t(1,\nu)} \tag{S.16}$$

where $\gamma_t(1, \nu)$ is a gamma process with unit mean rate — that is, a process with independent increments over non-overlapping intervals $[t, t+h]$ with density given by:

$$\left(\frac{1}{\nu}\right)^{h/\nu}\frac{x^{(h/\nu)-1}\exp\left(-x/\nu\right)}{\Gamma\left(h/\nu\right)} \qquad \forall x > 0 \tag{S.17}$$

where $t + h - t = h$ is the interval length and $\Gamma(\cdot)$ the Gamma function. Despite being the result of the subordination of a diffusive process, all the information regarding $\{X_t\}_{t\geq 0}$ is contained in its Lévy measure, making the variance-gamma model a pure jump process.

S.3 | Time-Varying Volatility

The previous section focused on the relaxation of the assumption of continuous price paths inherent in the geometric Brownian motion employed by Black and Scholes (1973) and

Merton (1973). The BSM model, however, builds on another fundamental assumption: namely, a constant volatility for the underlying asset. Just as the exponential Lévy processes discussed improve on the BSM framework by allowing for discontinuity points in the price process, *stochastic volatility* (SV) models are based on the idea that volatility itself follows a stochastic process. The additional randomness introduced by these class of models eventually leads to leptokurtic return distributions resembling the empirically observed ones (Gatheral, 2011). This class of models rejects the assumption of independent increments and constant volatility inherent to the BSM representation while addressing two additional stylised facts of asset returns, specifically about their variation. First, volatility tends to cluster — significant returns follow wide price swings while more minor variations often predict negligible returns — and second, volatility and returns share a negative correlation, an observation known as the leverage effect. The first framework explicitly accounting for both the leverage effect and the mean reverting nature of v_t is due to Heston (1993), who specified a bivariate diffusion for the spot price and its variance:

$$dS_t = \mu S_t\,dt + \sqrt{v_t}S_t\,dW_{1,t} \tag{S.18}$$

$$dv_t = \kappa(\theta - v_t)\,dt + \sigma\sqrt{v_t}\,dW_{2,t} \tag{S.19}$$

where $\{W_{1,t}\}_{t\geq 0}$ and $\{W_{2,t}\}_{t\geq 0}$ are two correlated BMs so that $\mathbb{E}[W_{1,t}W_{2,t}] = \rho t$, or equivalently $\mathbb{E}[dW_{1,t}\,dW_{2,t}] = \rho\,dt$. A replication argument similar to the one discussed for the derivation of the BSM fundamental PDE implies the following pricing equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}v_t S_t^2 + \frac{1}{2}\frac{\partial^2 V}{\partial v^2}\sigma^2 v_t + \frac{\partial^2 V}{\partial S\partial v}\rho\sigma S_t v_t - rV_t + rS_t\frac{\partial V}{\partial S} = -\frac{\partial V}{\partial v}[\kappa(\theta - v_t) - \xi\sigma\sqrt{v_t}] \tag{S.20}$$

where ξ stands for the market price of volatility risk. At this stage, Heston proposes a solution for the pricing PDE of the form:

$$V_t = S_t P_1 - Ke^{-r\tau}P_2 \tag{S.21}$$

where P_1 and P_2 represent the (risk-neutral) conditional probabilities of the option ending in the money at expiration under the stock and risk-free asset numéraires, respectively. Then, passing to the log-space with $X_t := \log(S_t)$ and combining (S.20) and (S.21) reveals that the two probabilities must satisfy the following PDEs:

$$\frac{\partial P_j}{\partial t} + \frac{\partial P_j}{\partial x}(r + u_j v_t) + \frac{\partial P_j}{\partial v}(\alpha_j - \beta_j v_t) + \frac{1}{2}\frac{\partial^2 P_j}{\partial x^2}v_t + \frac{1}{2}\frac{\partial^2 P_j}{\partial v^2}\sigma^2 v_t + \frac{\partial^2 P_j}{\partial x\partial v}\rho\sigma v_t = 0 \tag{S.22}$$

where $u_1 = -u_2 = 1/2$, $\alpha_1 = \alpha_2 = \kappa\theta$, $\beta_1 = \kappa - \rho\sigma$, and $\beta_2 = \kappa$. The two probabilities can then be found by inverting the characteristic function of the log-price, which is built to be affine in the state variables S_t and v_t. While returning a good fit for options contracts with expirations far ahead, the introduction of dependence in the spot price increments is not sufficient to replicate the pronounced smile for short-term contracts. In fact, the bivariate diffusion of Heston (1993) does not generate a sufficient variation in the spot price over a short tenor unless one assumes an excessive value for the volatility of volatility σ; hence, one needs to include discontinuities as to achieve such variation. Jumps may be included either in the diffusion driving the spot price or its variance, or in both. The first approach is followed by Bates (1996), who combined the benefits of the stochastic volatility model of Heston (1993) and the jump-diffusion framework of Merton (1976). In particular, the bivariate stochastic process dictating the dynamics of the spot price and its variance is described by the SDEs below:

$$dX_t = \left(r - \lambda\eta - \frac{v_t}{2}\right) dt + \sigma\, dW_{1,t}^{\mathbb{Q}_\xi} + \sum_{j=1}^{dN_t} \log(Y_j) \tag{S.23}$$

$$dv_t = [\kappa(\theta - v_t) - \xi\sigma\sqrt{v_t}]\, dt + \sigma\sqrt{v_t}\, dW_{2,t}^{\mathbb{Q}_\xi} \tag{S.24}$$

Even in the model proposed by Bates (1996), the premia charged for European options is found through the inversion of the (risk-neutral) characteristic function of the log-spot price of the underlying. As the Poisson counter $\{N_t\}_{t\geq 0}$ for jumps is assumed independent from the two BMs driving the two SDEs, the Fourier transform in the stochastic volatility jump-diffusion model can be obtained by simply multiplying that of the SV framework times the characteristic function of the JD process.

S.4 | Fourier Transform Methods

The concept of Fourier transform is named after French mathematician Joseph Fourier, who introduced it in 1822. This decomposition is widely employed in studying differential equations and signal processing. Within financial modelling, the Fourier transform was introduced by Stein and Stein (1991) and popularised by the work of Heston (1993). Before delving into the option pricing techniques based on characteristic functions, it is essential to point out that the Fourier transform of a given function f requires the latter to be (absolutely) integrable — that is, for $\mathfrak{F}[f(x)]$ to exist it must be that:

$$\int_{\mathbb{R}} |f(x)|\, dx < \infty \tag{S.25}$$

Furthermore, if $f(x)$ and $g(x)$ are both square integrable functions, then inner products $\langle f, g\rangle$ are preserved under Fourier transforms:

$$\langle f(x), g(x)\rangle = \int_{\mathbb{R}} f(x)\overline{g}(x)\, dx = \frac{1}{2\pi}\int_{\mathbb{R}} \mathfrak{F}[f]\overline{\mathfrak{F}[g]}\, dz \tag{S.26}$$

The relationship above is known as Parseval identity (or Plancherel Theorem) and provides the basis for the call valuation formula proposed by Lewis (2001). Not even a decade later after the work of Heston, Bakshi and Madan (2000) generalised several past results concerning option pricing through Fourier transforms. Letting $k = \log(K)$ and $X_T = \log(S_T)$, one can always decompose the value of a call option as follows:

$$\begin{aligned} C_t(S_t, T, K) &= e^{-r\tau} \int_K^{\infty} (S_T - K)\mathbb{Q}(S_T \mid \mathcal{F}_t)\, dS_T \\ &= S_t \Pi_1(S_T > K \mid \mathcal{F}_t) - Ke^{-r\tau}\Pi_2(S_T > K \mid \mathcal{F}_t) \end{aligned} \tag{S.27}$$

Here, both Π_1 and Π_2 represent the (risk-neutral) conditional probabilities of the option ending in the money at $t = T$, and are given by:

$$\Pi_1 = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \Re\left[\frac{e^{izk}\Phi_T(z - i)}{iz}\right] dz \tag{S.28}$$

$$\Pi_2 = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \Re\left[\frac{e^{izk}\Phi_T(z)}{iz}\right] dz \tag{S.29}$$

Another approach to the problem of valuing a European option is pursued by Carr and Madan (1999), who provide a valuation formula especially suitable for a discretised version of the Fourier transform, allowing to rely on a single inversion to price options for several exercise prices. Unfortunately, the premium of a call option is not absolutely integrable, and an appropriate transformation must be performed to make the implementation of the *fast Fourier transform* feasible. Nevertheless, Carr and Madan (1999) show that damping exponentially the call value leads to an integrable function — that is, it holds that:

$$c_t(S_t, T, K) := e^{\alpha k} C_t(S_t, T, K) \implies \int_{\mathbb{R}} |c_t(S_t, T, K)|\, dk < \infty \tag{S.30}$$

for an appropriate choice of $\alpha \in \mathbb{R}_+$. In turn, the characteristic function of the damped option price is given by the expression below:

$$\psi_{c_T}(z) := \psi_T(z) = \int_{\mathbb{R}} e^{-r\tau}\mathbb{Q}(X_T \mid \mathcal{F}_t)\left[\int_{-\infty}^{X_T} e^{izk}e^{\alpha k}(e^{X_T} - e^k)\, dk\right] dX_T \tag{S.31}$$

Computing the integral reveals that the characteristic function corresponds to:

$$\psi_T(z) = \frac{e^{-r\tau}\Phi_T(z - i(\alpha + 1))}{(iz + \alpha)(iz + \alpha + 1)} \tag{S.32}$$

To recover the undamped call premia, it is sufficient to apply an inverse Fourier transform and undo the exponential damping applied to make the function integrable — that is, the value of a European call option at time t is given by:

$$C_t(S_t, T, k) = \frac{e^{-\alpha k}}{\pi}\int_0^{\infty} \Re[e^{-izk}\psi_T(z)]\, dz \tag{S.33}$$

Two years after the publication of Carr and Madan's work, Lewis (2001) generalised the authors' results by showing that integrating a damped function is equivalent to evaluating a contour integral in the complex plane. Underlying Lewis's work is the idea that the payoff function of an option admits a generalised Fourier transform $\Phi_t(z) = \mathbb{E}[e^{izX_t}]$ with $z \in \mathbb{C}$ in some strip of regularity $\mathcal{S}_X$ parallel to the $\mathbb{R}$-axis in the complex plane. Provided the risk-neutral price process has a well-defined characteristic function $\Phi_T(z)$ for some $z \in \mathcal{S}_x$, and the modified payoff $\hat{w}(z)$ is regular with $z \in \mathbb{C}$ in a given strip of regularity $\mathcal{S}_w$, one can recover the price of an option relying on the martingale condition:

$$\begin{aligned} V_t(w(X_T)) &= e^{-r\tau}\mathbb{E}_{\mathbb{Q}}[w(X_T)] = e^{-r\tau}\mathbb{E}_{\mathbb{Q}}\left[\int_{i\Im[z]-\infty}^{i\Im[z]+\infty} e^{-izX_T}\hat{w}(z)\,dz\right] \\ &= \frac{e^{-r\tau}}{2\pi}\int_{i\Im[z]-\infty}^{i\Im[z]+\infty} \mathbb{E}_{\mathbb{Q}}[e^{-izX_T}]\hat{w}(z)\,dz = \frac{e^{-r\tau}}{2\pi}\int_{i\Im[z]-\infty}^{i\Im[z]+\infty} \Phi_T(-z)\hat{w}(z)\,dz \end{aligned} \tag{S.34}$$

For a call option, the fair premium to be charged is given by:

$$C_t(S_t,T,K) = -\frac{Ke^{-r\tau}}{2\pi}\int_{i\Im[z]-\infty}^{i\Im[z]+\infty} \frac{e^{-izk}\phi_T(-z)}{z^2-iz}\,dz \tag{S.35}$$

Shifting the contour by $\Im[z] = 0.5$ leads to an integration path which is equidistant from the two poles of the integrand — namely, $z = 0$ and $z = i$. In particular, setting $z = u + i/2$, the pricing formula above becomes:

$$\begin{aligned} C_t(S_t,T,K) &= S_t - \frac{\sqrt{S_tK}e^{r\tau/2}}{\pi}\int_0^\infty \Re\left[e^{-iuk}\phi_T\left(-u-\frac{i}{2}\right)\right]\frac{du}{u^2+1/4} \\ &= S_t - \frac{\sqrt{S_tK}e^{r\tau/2}}{\pi}\int_0^\infty \Re\left[e^{iuk}\phi_T\left(u-\frac{i}{2}\right)\right]\frac{du}{u^2+1/4} \end{aligned} \tag{S.36}$$

Finally, Lewis (2001) shows how moving the path of integration to the two poles given by $\Im[z] = 0$ and $\Im[z] = 1$ leads to the BSM-style formula discussed by Bakshi and Madan (2000) and presented at the beginning of this section.

Calibrating one of the valuation models seen so far amounts to finding the set of parameters Θ that minimises the deviation of the prices predicted by the model from those quoted on the market. In incomplete market models, the time series of the underlying is not sufficient to reach an estimated price surface for the options contracts; moreover, the equivalent martingale measure $\mathbb{Q}$ only shares qualitative characteristics of the true measure $\mathbb{P}$, such as the presence of jumps and, for instance, the finiteness of the Lévy measure (Cont and Tankov, 2003). Therefore, one must rely on an "implied" modelling technique where $\mathbb{Q}$ is found by calibrating the pricing model to the (most recent) surface of quoted option prices. Once the model parameters have been estimated, one can use the model to price exotic instruments and devise hedging strategies. In this sense, model calibration is the inverse of the pricing problem — that is, while the latter is concerned with valuing one or more options contracts given a set Θ of model parameters, the former is about finding Θ such that the model outputs a given set of prices.

The actual calibration of the pricing models discussed is conducted relying on the outstanding call options chain written on the *Apple* stock as of the 17$^{\text{th}}$ of August, 2022. The analysis could easily be conducted by referring to put contracts as well, but the put-call parity arbitrage relation ensures that the magnitude of mispricings with respect to the BSM framework are similar across the two contract speifications; as a result, the analysis which follows focuses on call options. Furthermore, even though the derivations of the models assumed a non-dividend paying underlying, the generalisation is straightforward: it is indeed sufficient to impose a correction in the drift of the spot (log-)price by letting it equal $r - q$ in the respective risk-neutral characteristic functions; at the time of the analysis, the *Apple* stock closed at a price of \$174.55, thereby implying a dividend yield of $q = 0.90/174.55 \approx 0.52\%$. The contract with the earliest expiration matures on the 26/08/2022, while the last available option in terms of maturity refers to the 21/06/2024. The exercise prices range from a minimum of \$30 for the 20/01/2023 maturity up to a maximum of \$320 for the last expiration available. Even though the options prices are proxied by the bid-ask mid-point, it is worth noting that even on such a liquid instrument — that is, options written on the *Apple* stock — the bid-ask spread goes from \$0.01 to \$0.55. Overall, the data set contains 342 calls sharing 19 common strikes with a minimum of \$130 and a maximum of \$240. To render the calibration procedure (slightly) faster, a filter considering strikes lying in an interval bounded by ±25% of the spot price is applied, thereby leading to a price surface made up of 15 strikes for 18 different expirations. The risk-free rate is assumed to be proxied by the par yield published daily by the US Treasury. Unfortunately, the Treasury only provides the yield for a predetermined set of maturities. Hence, in order to obtain the relevant rates for the options expirations in the data set, it is necessary to resort to yield curve models. Within the latter, a popular and proven efficient choice is the well-known framework of Nelson and Siegel (1987) and in particular its six-parameter extension proposed by Svensson (1994). Starting from the jump-diffusion model of Merton (1976), the results highlight how the options considered come with a significant implied jump intensity for the Poisson process, with around five jumps every two years. Together with the estimate for the average log-jump size α, the jump-diffusion parameters seem to imply a large kurtosis and a substantial asymmetry for the risk-neutral density of log-returns on the *Apple* stock. Coherently, the results for the variance-gamma model proposed by Madan et al. (1998) show how the implied risk-neutral density is both significantly left-skewed and leptokurtic; in particular, the implied kurtosis is about 30%

		All Contracts				Short-Term Contracts			
		JD	VG	SV	SVJD	JD	VG	SV	SVJD
Merton	σ	0.1613 (0.0041)	—	—	—	0.1985 (0.0037)	—	—	—
	λ	2.5746 (0.1683)	—	—	—	7.0621 (0.1829)	—	—	—
	α	0.1369 (0.0058)	—	—	—	0.1586 (0.0049)	—	—	—
	β	0.1620 (0.0022)	—	—	—	0.0936 (0.0017)	—	—	—
Madan et al.	θ	—	0.3678 (0.0194)	—	—	—	0.7911 (0.0236)	—	—
	σ	—	0.2768 (0.0018)	—	—	—	0.2664 (0.0023)	—	—
	ν	—	0.3031 (0.0034)	—	—	—	0.0783 (0.0037)	—	—
Heston	ρ	—	—	0.5519 (0.0070)	—	—	—	0.6301 (0.0074)	—
	σ	—	—	0.7863 (0.0239)	—	—	—	1.1498 (0.0314)	—
	θ	—	—	0.0520 (0.0002)	—	—	—	0.0107 (0.0005)	—
	κ	—	—	1.1646 (0.1632)	—	—	—	3.3791 (0.1743)	—
	v_0	—	—	0.1399 (0.0006)	—	—	—	0.0959 (0.0005)	—
Bates	ρ	—	—	—	0.4469 (0.0159)	—	—	—	0.5953 (0.0231)
	σ	—	—	—	0.6298 (0.0722)	—	—	—	0.7737 (0.0536)
	θ	—	—	—	0.0739 (0.0042)	—	—	—	0.1165 (0.0073)
	κ	—	—	—	2.8937 (0.3986)	—	—	—	4.3243 (0.5143)
	λ	—	—	—	0.8160 (0.1477)	—	—	—	0.8463 (0.1318)
	α	—	—	—	0.0624 (0.0056)	—	—	—	0.1107 (0.0053)
	β	—	—	—	0.0726 (0.0026)	—	—	—	0.0904 (0.0031)
	v_0	—	—	—	0.0672 (0.0046)	—	—	—	0.0977 (0.0047)
MAPE (%)		5.94	6.82	2.23	1.81	12.24	14.01	4.39	3.51

Table S.1 Results of the calibration procedure for the four pricing models discussed so far. When it comes to pricing accuracy, all the frameworks provide a substantial improvement on the BSM framework, associated with 36.38% and 65.67% MAPEs for all and short-term contracts, respectively. The most significant contribution is undoubtedly due to the introduction of stochastic volatility paths, and allowing for discontinuities improves the fit even further.

greater than it would be for a Normal distribution of returns, or 3.9. The parameter set Θ^* minimising the weighted sum of pricing errors for the stochastic volatility model of Heston (1993) confirm the presence of the leverage effect discussed in the third section, wherein returns and volatility are known to share a negative correlation. It is also known that the correlation ρ between volatility v_t and spot price S_t and the volatility of volatility σ are responsible for the skewness and kurtosis of the returns distribution, respectively. It is then clear that the implied distribution is asymmetrical and leptokurtic, as the Heston (1993) model requires a substantially high estimate of σ to justify the pronounced volatility skew observed in the options data for short maturities. Finally, turning to the stochastic volatility jump-diffusion model described by Bates (1996) one notes that the leverage effect is still present but with a lower magnitude, while the estimated speed of mean reversion κ is slightly higher. At the same time, the estimate for the volatility of volatility σ is smaller than that of the stochastic volatility relying on continuous paths only. Concerning the discontinuous component of the price process, the intensity of jumps, average log-jump size, and variance of the latter are all associated to estimates lower than those obtained for the jump-diffusion model assuming a constant volatility. Interestingly, the results for the Bates (1996) model seem to suggest that the jump-diffusion model justifies the implied volatility skew with greater estimates for the parameters of the compound Poisson process which may be instead the result of an additional source of randomness in the volatility process; similarly, the estimated Θ^* for the framework proposed by Heston (1993), coupled with the results observed for the Bates (2006) model, show that the high variability attached to v_t may actually be the result of a discontinuous component in the price process.

Focusing on short-term call contracts, one immediately notes how the estimate for the volatility coefficient — either for the spot price or its volatility — is inevitably higher for all the models considered. As a result, one can infer that a significantly more erratic path for the underlying's price or volatility is implied in the prices of short-term options contracts. Furthermore, both the stochastic volatility and stochastic volatility jump-diffusion models predict a (slightly) stronger leverage effect which, in the Bates (1996) framework, is coupled with more frequent jumps characterised by a significantly more negative average (log-)jump size, down by about 5% from the estimate obtained with the calibration to all the available contracts. A similar observation holds for the models allowing for discontinuity points in the price paths discussed in, wherein the more significant asymmetry proxied by a lower correlation coefficient ρ in the Heston (1993) and Bates (1996) models is translated to a much more negative estimate for α and θ in the jump-diffusion and (asymmetric) variance-gamma representations. Similarly, the greater volatility also leads to a larger kurtosis, as shown by the increase in the estimates for λ and ν. Hence, it follows that for the considered short-term contracts to be priced consistently by models allowing for discontinuous price paths, the magnitude of market crashes and their frequency must be greater. On the other hand, frameworks involving stochastic volatility attribute the greater variation potentially generated by jumps in the spot price to the presence of a stronger leverage effect and an even more erratic, or *rough*, volatility process.

Overall, the results of Table S.1 show that the biggest improvement over the BSM model in terms of pricing performance is achieved through the explicit inclusion of a stochastic process dictating the dynamics of the underlying volatility. While providing a good fit for short-term options, models only allowing for discontinuities in the price path such as exponential Lévy processes suffer from their independent increments requirement, eventually leading to implied volatility surfaces which flatten out too quickly as the contracts expirations increase. Still, a generalisation of the class of jump-diffusion models accounting for stochastic volatility or, equivalently, an expansion of the stochastic volatility framework allowing for discontinuities in the price path — such as the model described by Bates (1996) — can indeed improve the fit to the volatility surface even further. Concerning the calibration procedure, a research limit of this thesis lies in the absence of an out-of-sample assessment of the results obtained — that is, it would be interesting to look at the pricing accuracy obtained with the parameters calibrated on a different, more extensive, options data set. Equally interestingly, one may assess the fit of an increasingly popular class of models relying on fractional Brownian motion — a process relaxing the continuity assumption — known as fractional stochastic volatility frameworks. Here, exploring variations of the rough fractional stochastic volatility model, based on a fractional Brownian motion with Hurst exponent below one-half, would be particularly engaging. Although present in a more elaborate representation, the ubiquitous need to abstract from the assumption of independent increments within these models validates the building blocks on which the models presented in this thesis are built.