Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE
Corso di Laurea Magistrale in Matematica

Stochastic Volatility Jump Models for Cryptocurrency Option Pricing

Relatore:
Chiar.mo Prof.
Andrea Pascucci

Presentata da:
Marilena Palomba

Sessione VI
Anno Accademico 2019/2020

Sommario

Le criptovalute stanno acquisendo straordinaria importanza come asset finanziari nell'economia contemporanea. Le loro caratteristiche rivoluzionarie, l'innovativa struttura sottostante e la crescente capitalizzazione incentivano lo studio del mercato delle criptovalute. Anche se la ricerca è ancora limitata, l'effettiva disponibilità di opzioni e futures negoziati su piattaforme di scambio indipendenti incoraggia la costruzione e la formalizzazione di un mercato apposito. Vista la mancanza di una regolamentazione centrale e l'assenza di opzioni ufficialmente negoziate sul mercato, comprendere la peculiare evoluzione del prezzo delle criptovalute è essenziale per lo sviluppo di un mercato per gli strumenti derivati. In questa tesi, proponiamo un meccanismo basato sul modello a volatilità stocastica (SVCJ) proposto da Duffie, Pan e Singleton (2000) per la simulazione dell'andamento dei prezzi delle criptovalute. Viene proposta una calibrazione del modello tramite l'utilizzo dei dati storici relativi al prezzo dei Bitcoin (BTC). I risultati delle analisi svolte corroborano l'ipotesi iniziale, ossia che il modello SVCJ risulta essere una scelta ragionevole per simulare le dinamiche di prezzi e ritorni dei BTC.

Abstract

Cryptocurrencies have become extraordinarily important financial assets in contemporary economy. Their revolutionary characteristics, innovative underlying structure and the growing market capitalization are stimulating the study of cryptocurrency market. Even if research on cryptocurrency are still limited, the actual availability of options and futures traded on independent exchange platform encourages the creation and formalization of a derivatives market for cryptocurrencies. Given the lack of central regulation and the absence of fundamentals, understanding the peculiar evolution of cryptocurrencies price is essential to the development of a contingent claims market. In this thesis, we propose a mechanism based on the stochastic volatility correlated jump (SVCJ) model proposed by Duffie, Pan and Singleton (2000) for simulating cryptocurrencies price paths. We perform model calibration using Bitcoin (BTC) historical data and verify that the goodness of fit reached makes the SVCJ model a reasonable choice for simulating BTC price and return evolution.

Contents

List of Figures

Introduction

In the research on option pricing, the dynamics of the asset price is usually represented by a geometric Brownian motion. Brownian motion paths are characterized by continuity and scale invariance, then using such dynamics to simulate and reproduce stocks evolution lead to strong assumption over options underlying trajectory that are generally inconsistent with their actual behaviour.

Removing the hypotheses of continuity and scale invariance is our first objective. In this perspective, we need to modify the underlying framework to preserve the robustness of important results that may be threatened by the presence of jumps in prices and non-continuous paths. Therefore, we move from diffusion models to jump-diffusion ones, where the evolution of price is driven by a diffusion process interrupted by randomly occurring jumps representing rare crashes and abrupt changes in price dynamics. Such an evolution can be represented by modeling the price as a Lévy process.

A jump-diffusion model is substantially build up from a Brownian motion (for the diffusion part) and a Poisson process (for the jump part). In the first chapter we present and discuss several aspects related to building jump-diffusion models and the main theorems needed. The final scope of this section is to introduce the Stochastic Volatility Correlated Jump (SVCJ) model formalized by Duffie, Pan and Singleton, [1], that will be used in Chapter 3 to estimate Bitcoin prices and returns dynamics.

Once the model of interest for our dissertation is introduced, it is essential to identify a suitable method to estimate the diffusion process to real data. For stochastic volatility models, a particularly well suited method is represented by the Markov chain Monte Carlo (MCMC), as suggested by Jacquier, Polson, and Rossi (1994) [2]. The MCMC approach is, among all estimation models, one of the most computationally efficient and flexible, and it gives an accurate estimations of latent volatility, jump sizes and jump times.

Chapter 2 is entirely devoted to the description of these methods, that provide the theoretical background for extracting coherent information about latent state variables, structural parameters and market prices from observed data.

The practical part of our work is presented in Chapter 3. Here we apply our studies to a new digital asset class that is gaining attention and challenging contemporary financial markets. The focus is then on the so-called cryptocurrencies, i.e., digital currencies that rely on cryptographic proofs for confirmation of transactions. The reason why these kind of currencies are having an incredible diffusion is that they achieve a unique combination of three features: ensuring limited anonymity, independence from central authority and double spending attack protection. As the interest in cryptocurrencies grow, it becomes more and more vital to have the right tools to study and model the dynamics of this asset.

Due to the strong presence of sudden and frequent jumps in cryptocurrencies prices, the classical methods for simulating possible price paths cannot be applied. Therefore, we inspect cryptocurrencies dynamics from a double-jump-diffusion point of view and reproduce price evolution through the stochastic volatility correlated jump model of Duffie Pan and Singleton.

1 Jump-Diffusion Models

At the very beginning of financial modelling studies, the price S_t of an asset was proposed to be modeled as:

$$S_t = S_0 + \sigma W_t \tag{1.1}$$

where W_t is a Brownian motion, i.e., a random process with independent and stationary increments that follow a Gaussian distribution. The multiplicative version of (1.1) leads to the commonly used Black-Scholes model [3] where the log-price $\ln S_t$ follows a Brownian motion:

$$S_t = S_0 \exp[\mu t + \sigma W_t]. \tag{1.2}$$

The dynamics of a stock price is usually represented by its relative change, $\frac{dS_t}{S_t}$, i.e., the local form of (1.2):

$$\frac{dS_t}{S_t} = \sigma dW_t + (\mu + \frac{\sigma^2}{2})dt. \tag{1.3}$$

The process S is called geometric Brownian motion (GBM). The first part of the right hand side of (1.3), σdW_t, simulates the random change in the asset price in response to external effects, such as uncertain events. While, the second part reflects a predictable, deterministic and anticipated return which is similar to the return of the investment in banks.

Even if, at a first sight, a Brownian motion path seems to resemble a stock price evolution, the two curves are in fact pretty different. The continuity of Brownian motion's sample paths represents the first main difference: a typical path $t \mapsto W_t$ is a continuous function of time while the stock price undergoes several abrupt downward jumps, which appear as discontinuities in the price trajectory. Another property of Brownian motion that does not match with the evolution of a stock price is its scale invariance: the statistical properties of Brownian motion are the same at all time resolutions. Going down on an intraday scale, the Brownian model still retains its continuous behaviour while the stock price moves essentially through jumps.

The Black-Scholes model is not the only continuous time model built on Brownian motion. The stochastic volatility model, introduced by Heston [4], is an additional continuous time model that relies on Brownian motion: the price S_t is the first component of a bivariate diffusion (S_t, σ_t) process driven by a two-dimensional Brownian motion (W_t^1, W_t^2):

$$\begin{cases} \frac{dS_t}{S_t} & = \sigma_t dW_t^1, \\ \sigma_t & = f(T_t), \\ dY_t & = \alpha_t dt + \gamma_t dW_t^2. \end{cases} \tag{1.4}$$

While Heston model has more flexible statistical properties, it still shares with the Brownian motion the property of continuity, in contrast with the real prices behavior over time scales. In the diffusion-based models presented, the continuity of paths plays a fundamental role. Removing the hypothesis of continuity in such models, many results obtained are not robust to the presence of jumps in prices and this

leads to the necessity of changing, at least in part, the underlying framework.

From an investor's point of view, describing the price trajectories is not as important as predicting the increments of log-prices. Analysing such quantities for a Brownian motion, it emerges that the returns generated have roughly the same amplitude, while real returns can manifest frequent peaks, corresponding to jumps in the price, and vary in amplitude. This huge variability is historically experienced in financial asset returns and, statistically, results in heavy tails in the empirical distribution of returns. This is the so-called asymmetric leptokurtic features: the return distribution is skewed to the left, and has a higher peak and two heavier tails than those of the normal distribution.

However, nonlinear diffusion processes such as (1.4) are not Gaussian processes, even though the driving noise is Gaussian, and an appropriate choice of coefficients can generate arbitrary heavy tails. The point is that, due the lack of degrees of freedom for tuning the local behaviour, producing heavy tail translates automatically into obtaining non-stationarity or unrealistic high values of volatility in diffusion-based stochastic volatility models. The strongest argument that leads to the necessity of switching from diffusion-based and continuous models to discontinuous ones is the presence of jumps in the stock price. In fact, diffusion models cannot generate such discontinuous and abrupt movements in prices. In order to reproduce or capture the state of the options market at a given time, the model's parameters must be chosen to fit the market prices of options or, at least, reproduce the main features of the prices. The second shortcoming of models based on Brownian motion is that, even when they can reproduce fairy well the profile of implied volatility, they give rise to non-intuitive profile of local volatility (in the case of diffusion models) and cannot yield a realistic term structure of implied volatilities (in the case of diffusion-based stochastic volatility models).

For Black-Scholes model, even if in some cases it is straightforward to calculate the value of an option (e.g. a vanilla option) and fit the observed market price, the resulting value for volatility parameter cannot lead to a realistic profile of the implied volatility surface. In fact, the flat profile, predicted by Black-Scholes model, for implied volatility surface contradicts empirically verified properties such as:

- The dependence with respect to the strike price, that may be decreasing ("skew") or U-shaped ("smile");
- The flattening of the smile/skew profile as maturity increases;
- The floating smiles, which read as implied volatility patterns vary less in time than when expressed as a function of the strike.

By contrast, jump models lead to a variety of smile/skew patterns and allow to explain the distinction between skew and smile in terms of asymmetry of jumps anticipated by the market.

The need for changing the underlying framework comes from the awareness that diffusion models have the wrong qualitative properties and therefore can convey erroneous intuitions about price fluctuations and the risk resulting from them. In some sense, when viewed as a subset of the larger family of jump-diffusion models, diffusion models should be considered as singularities. For jump-diffusion models,

the normal evolution of price is given by a diffusion process, involving randomly occurring discontinuous jumps. Here the jumps represent rare crashes and large drawdowns. Such an evolution can be represented by modeling the price as a Lévy process, i.e., a process with stationary independent increments which is continuous in probability. The two basic building blocks of every jump-diffusion model are the Brownian motion (the diffusion part) and the Poisson process (the jump part).

This chapter aims at laying the foundation for understanding and building jump-diffusion models, that can then be implemented to perform option pricing. First of all, the most important features of Poisson and Lévy processes are presented in order to stress the differences between Brownian motion and jump processes. Subsequently, we define and inspect different jump-diffusion models focusing on their property and limitations. At the end of the chapter, we present the double-jump-diffusion model proposed by Duffie, Pan and Singleton, [1], that represents a milestone for the discussion in the rest of this thesis.

1.1 The Poisson Process

The first element needed for building a jump process is the Poisson process, i.e., a stochastic process with discontinuous paths. The Poisson process can be defined in a constructive way from a sequence $(\tau_n)_{n\geq 1}$ of independent random variables with exponential distribution with parameter $\lambda > 0$.

Definition 1.1. (Poisson Process) *Let $(\tau_i)_{i\geq 1}$ be a sequence of independent random variable such that $\tau_i \sim Exp_\lambda$, for $\lambda > 0$ and $i \geq 1$. Let $T_n := \sum_{k=1}^{n} \tau_k$, then a* Poisson process *is the process $(N_t)_{t\geq 0}$ defined by*

$$N_t = \sum_{n\geq 1} \mathbf{1}_{t\geq T_n}.$$

It is worth noticing that

$$E[T_n - T_{n-1}] = E\tau_n = \frac{1}{\lambda}, \quad n \in \mathbb{N},$$

and $\frac{1}{\lambda}$ is the average distance among subsequent jumps, i.e., λ jumps are expected in a unit time interval. For this reason, λ is also called the *intensity parameter*.

The Poisson process N_t counts the number of jumps that occur at or before time t. In particular N_t takes only non-negative integer values.

Proposition 1.2. *Let $(N_t)_{t\geq 0}$ be a Poisson process. Then:*

1. *For any $t > 0$, N_t is finite a.s.;*
2. *For any ω, the sample path $t \to N_t(\omega)$ is piecewise constant and increases by jumps of size 1;*
3. *The sample paths $t \mapsto N_t$ are càdlàg* [1]*;*

[1] *Càdlàg* is the French acronym for "continue à droite, limite à gauche" and refers to a function that is right continuous with left limits.

4. *For any $t > 0$, $N_{t-} = N_t$ with probability 1;*

5. *(N_t) is continuous in probability;*

6. *For any $t > 0$, N_t follows a Poisson distribution with parameter λt:*

$$\forall n \in \mathbb{N}, \quad P(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}.$$

7. *The characteristic function of N_t is given by*

$$E[e^{iuN_t}] = \exp\{\lambda t(e^{iu} - 1)\}, \quad \forall u \in \mathbb{R}.$$

8. *(N_t) has independent increments: for any $t_1 < \cdots < t_n$,*

$$N_{t_n} - N_{t_{n-1}}, \ldots, N_{t_2} - N_{t_1}, N_{t_1}$$

are independent random variables.

9. *The increments of N are homogeneous: for any $t > s$,*

$$N_t - N_s \sim N_{t-s}.$$

10. *(N_t) has the Markov property:*

$$\forall t > s, \quad E[f(N_t)|N_u,\ u \leq s] = E[f(N_t)|N_s].$$

Proof. See [5] Chapter 2.5. □

In addition, notice that from the preceding proposition we have

$$E[N_{t+1} - N_t] = E[N_1] = \sum_{n\geq 1} nP(N_1 = n) = e^{-\lambda}\sum_{n\geq 1}\frac{\lambda^n}{(n-1)!} = \lambda,$$

which confirms that the number of jumps expected in a unit time interval is represented by the intensity λ.

Remark 1.3. Any counting process with stationary independent increments is a Poisson process.

Remark 1.4. Let $(\mathcal{F}_t)$ be the filtration generated by a Poisson process N. By the independence of increments, for any $t > s \geq 0$ we have

$$E[N_t|\mathcal{F}_s] = E[N_t - N_s] + N_s = \lambda(t - s) + N_s.$$

As a consequence, the process $N_t - \lambda t$ is a martingale and is usually called *compensated Poisson process*. Compensated Poisson process is no longer integer valued and it is not a counting process.

Definition 1.5. (Compound Poisson process) *Let N be a Poisson process with intensity λ and let $Y = (Y_n)$ be a sequence of i.i.d. random variables in $\mathbb{R}^n$ with distribution η, i.e., $Y_n \sim \eta$ for $n \geq 1$, and which are independent of N. The* compound Poisson process *is defined as*

$$X_t = \sum_{n=1}^{N_t} Y_n, \quad t \geq 0. \tag{1.5}$$

The jumps of the compound Poisson process X in (1.5) occur at the same times as the jumps of N and X is a càdlàg process. However, the jumps in X are not of fixed size, they present random sizes with distribution η. A compound Poisson process has independent and stationary increments. Moreover, if we set $m = E[Z_1] \in \mathbb{R}^d$, then we have

$$E[X_t] = m\lambda t.$$

Definition 1.6. (Compensated compound Poisson process) *Let X be a compound Poisson process with intensity λ and distribution of jumps η. The process*

$$\tilde{X}_t = X_t - E[X_t] = X_t - m\lambda t,$$

where

$$m = \int_{\mathbb{R}^d} x\eta(dx) = E[Y_1],$$

is called compensated compound Poisson process*.*

A compensated compound Poisson process is a martingale with respect to the filtration generated by N and Y.

1.2 Lévy processes

The Brownian motion and the Poisson process are fundamental examples of Lévy processes. In fact, they represent the starting point for building a Lévy process: it is a superposition of a Brownian motion and a number of independent Poisson processes. The Lévy processes retains the property of the independence and stationarity of the increments.

An important consequence is the infinite divisibility of distributions, which implies that X_t at a fixed time, say $t = 1$, can be expressed as the sum of a finite number of i.i.d. random variables: this provides a motivation for modeling price changes as resulting from a large number of shocks in economy. The Brownian motion is a very special example, since it is the only Lévy process with continuous trajectories; on the other hand, the presence of jumps is one main motivation that has led to consider Lévy processes in finance. Hereafter, we assume we are given a filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$ satisfying the *usual hypotheses* [2].

[2]Following [6], we say that $(\mathcal{F}_t)$ satisfies the *usual hypotheses* with respect to P if:

- $\mathcal{F}_0$ contains the family of P-negligible events;
- the filtration is right-continuous, i.e., for every $t \geq 0$, $\mathcal{F}_t = \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$.

Definition 1.7. (Lévy process) *A* Lévy process *is a càdlàg adapted stochastic process* $X = (X_t)_{t\geq 0}$ *defined on the space* $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$ *with values in* $\mathbb{R}^d$*, such that*

i) $X_0 = 0$ *a.s.;*

ii) X *has increments independent of the past, that is,* $X_t - X_s$ *is independent of* $\mathcal{F}_s$ *for* $0 \leq s < t$*;*

iii) X *has stationary increments, that is,* $X_t - X_s$ *has the same distribution of* X_{t-s} *for* $0 \leq s < t$*;*

iv) X *is stochastically continuous, that is, for all* $\varepsilon > 0$ *and for all* $t \geq 0$*, we have*

$$\lim_{h\to 0} P(|X_{t+h} - X_t| \geq \varepsilon) = 0.$$

The following lemma, proved in [6], leads to the specification of an important property of Lévy processes, i.e., the fact that such processes can have infinitely (countably) many small jumps but only a finite number of large ones.

Lemma 1.8. *Let* $I = [0, T]$ *be a compact interval and let* f *be a càdlàg function defined on* I*. Then, for any* $n \in N$*, the number of jumps of* f*,* Δf*, of size greater than* $\frac{1}{n}$ *is finite:*

$$\#\{t \in]0, T] \mid |\Delta f(t)| \geq \frac{1}{n}\} < \infty.$$

In particular, f *has at most a countable number of jumps.*

Proof. See [6]. □

As a consequence of Lemma 1.8, given a Lévy process X, $T > 0$ and $H \in \mathcal{B}(\mathbb{R}^d)$ such that $0 \notin \bar{H}$ so that

$$dist(H, 0) = \inf\{|x| \mid x \in H\} > 0,$$

we have that, with probability one, $(X_t)_{t\in[0,T]}$ has only a finite number of jumps of size belonging to H.

Definition 1.9. (Finite activity Lévy process) *Let* X *be a Lévy process having only a finite number of jumps in any bounded time interval, then we say that* X *is a* finite activity *Lévy process.*

Definition 1.10. (Infinite activity Lévy process) *Let* X *be a Lévy process, if* X *is not a finite activity Lévy process, then we say that it has an* infinite activity*.*

We observe that the natural filtration of a Lévy process X completed by the negligible events, $(\mathcal{F}_t^X)$, is right-continuous and therefore it satisfies the usual hypotheses. In addition, X is a Lévy process with respect to $\mathcal{F}^X$.

1.2.1 Characteristic function of Lévy processes

A huge part of option pricing theory under Lévy processes builds on Fourier transform methods. Below we recall some useful facts that are strictly linked to this approach.

The characteristic function of a Lévy process X is given by

$$\varphi_{X_t}(\xi) = E[e^{i\xi \cdot X_t}],$$

where $\xi \in \mathbb{R}^d$ and $t \geq 0$.

A notable property of the characteristic function of a Lévy process is expressed by the following

Theorem 1.11. *Let X be a Lévy process, then there exists and it is unique a function $\psi \in C(\mathbb{R}^d, \mathbb{C})$ such that $\psi(0) = 0$ and*

$$\varphi_{X_t}(\xi) = e^{t\psi(\xi)}, \quad t \geq 0, \xi \in \mathbb{R}^d.$$

Proof. See [6]. □

Definition 1.12. (Characteristic exponent) *Let X be a Lèvy process. The function ψ, whose existence and uniqueness are given by Theorem 1.11, is called the* characteristic (or Lévy) exponent *of X.*

Since the distribution of a random variable is determined by its characteristic function, Theorem 1.11 suggests that the law of X_t is only determined by the characteristic exponent ψ or equivalently by the law of X_1. Therefore the distribution of a Lévy process X can be completely specified through the distribution of X_t for a single time.

1.2.2 Jump measures of compound Poisson process

Let us now consider a jump-diffusion process X in $\mathbb{R}^d$ of the form

$$X_t = \mu t + W_t + \sum_{n=1}^{N_t} Y_n \tag{1.6}$$

where $\mu \in \mathbb{R}^d$, W is a d-dimensional correlated Brownian motion with correlation matrix $\mathcal{C}$, N is a Poisson process with intensity λ and $(Y_n)_{n\geq 1}$ are i.i.d. random variables in $\mathbb{R}^d$ with distribution η. The Brownian and Poisson components are considered to be independent.

For any $I \times H \in \mathcal{B}([0, +\infty[\times\mathbb{R}^d)$, we put

$$J(I \times H) := \sum_{n\geq 1} \delta_{T_n}(I)\delta_{Y_n}(H) \tag{1.7}$$

where δ denotes the Dirac delta and (T_n) is the increasing sequence of jump times.

Note that for $I = [0,t]$ (1.7) reduces to

$$J([0,t] \times H) = \sum_{n=1}^{N_t} \delta_{Y_n}(H),$$

and $J([0,t] \times H)$ counts the number of jumps occurring in the time interval $[0,t]$ such that their size is in H. The first sum has a finite number of terms and J is well-defined since, with probability one, only a finite number of jumps occurs on any bounded time interval. Moreover, J is a finite sum of Dirac deltas and therefore it is a σ-finite measure on $\mathcal{B}([0,+\infty[\times\mathbb{R}^d)$ taking values in the set of non-negative integers $\mathbb{N}_0$: notice that J is a measure taking random values, i.e. it is a random measure.

Definition 1.13. (Jump measure) *The random measure J in* (1.7) *is called* jump measure *of X.*

It is possible to calculate the expectation of J:

$$E[J([0,t] \times H)] = t\lambda\eta(H),$$

and it follows that

$$E[J([0,t] \times H)] = tE[J([0,1] \times H)].$$

Defining then

$$\nu(H) := E[J([0,1] \times H)] \tag{1.8}$$

for $H \in \mathcal{B}(\mathbb{R}^d)$, we find that $\nu(H) = \eta(H)$ and ν defines a finite measure on $\mathcal{B}(\mathbb{R}^d)$ such that $\nu(\mathbb{R}^d) = \lambda$.

Definition 1.14. (Intensity measure) *The measure ν in* (1.8) *is called the* intensity (or Lévy) measure *of X.*

The intensity measure $\nu(H)$ determines the expected number, per unit time, of jumps of X whose amplitude belongs to H. The definition of ν implies that this measure cannot be integer-valued.

Remark 1.15. The characteristic exponent of the jump part of X in (1.6) (i.e. the compound Poisson process $\sum_{n=1}^{N_t} Z_n$), can be expressed in terms of the Lévy measure as follows:

$$\psi(\xi) = \int_{\mathbb{R}^d} (e^{i\xi\cdot x} - 1)\nu(dx), \quad \xi \in \mathbb{R}^d.$$

The Lévy measure thus characterizes the jump part of X.

More generally, the Lévy process X in (1.6) is completely identified by the triplet $(\mu, \mathcal{C}, \nu)$ where:

- μ is the coefficient of the drift part;
- $\mathcal{C}$ is the covariance matrix of the diffusion part;
- ν is the intensity measure of the jump part.

A similar characterization holds for the general class of Lévy processes.

It is possible to obtain a representation of X in terms of its jump measure J since it holds:

Theorem 1.16. *Let X be a jump-diffusion process of the form (1.6) with jump measure J and Lévy measure ν. For any function $f = f(t,x)$ we have*

$$\sum_{0<s\leq t,\ \Delta X_s\neq 0} f(s,\Delta X_s) = \int_0^t \int_{\mathbb{R}^d} f(s,x)J(ds,dx). \tag{1.9}$$

Let us assume, in addition, that $f \in L^1([0,+\infty[\times\mathbb{R}^d, ds\otimes\nu)$ and defining

$$M_t = \int_0^t \int_{\mathbb{R}^d} f(s,x)\tilde{J}(ds,dx)$$

where

$$\tilde{J}(dt,dx) := J(dt,dx) - dt\nu(dx)$$

is called the compensated jump measure *of X. Then M is a martingale and $E[M_t] = 0$. Moreover, if $f \in L^2([0,+\infty[\times\mathbb{R}^d, ds\otimes\nu)$, then we have*

$$var(M_t) = \int_0^t \int_{\mathbb{R}^d} f^2(s,x)\nu(dx)ds.$$

Proof. See [6] □

Choosing $f(x) = x$, it is then possible to represent the process X in (1.6) as

$$X_t = \mu t + W_t + \int_0^t \int_{\mathbb{R}^d} xJ(ds,dx).$$

Moreover, if $f(x) = x$ is η-integrable (and therefore ν-integrable) then we have

$$E[X_t] = t\big(\mu + \int_{\mathbb{R}} x\nu(dx)\big).$$

1.2.3 Lévy-Itô decomposition

Given a jump-diffusion process X and its jump measure J, it is then possible to express the process as

$$X_t = \mu t + W_t + \int_0^t \int_{\mathbb{R}^d} xJ(ds,dx). \tag{1.10}$$

In the case of compound Poisson process, X has a finite number of jumps in any bounded time interval and this leads to a measure $J([0,t]\times\mathbb{R}^d)$ finite a.s.; however, in general, Lévy processes do not necessarily share this property. It is worth, then, investigate if every Lévy process X admits a representation of the form (1.10). First of all we need to inspect the definition of the jump measure J. When analysing

finite activity processes (compound Poisson processes) J is well defined since the sum in (1.6) has only a finite number of terms. On the other hand, when infinitely many jumps are expected in finite time, J may become infinite. However, a Lévy process X can only have a finite number of "large" jumps and this make it possible to define the jump measure for a generic Lévy process. More precisely, if $H \in \mathcal{B}(\mathbb{R}^d)$ and $0 \notin \bar{H}$ then X has only a finite number of jumps with size in H. This allows to define $J(I \times H)$ for any $I \times H \in \mathcal{B}([0, +\infty[\times\mathbb{R}^d)$ with I bounded and H such that $0 \notin \bar{H}$:

$$J(I) := \#\{t \in I \mid \Delta X_t \in H\}$$

Then, J can be extended to a σ-finite random measure on $\mathcal{B}([0, +\infty[\times\mathbb{R}^d \setminus \{0\})$. In this context, the Lévy measure of X,

$$\nu(H) := E[J([0,1] \times H)], \quad H \in \mathcal{B}(\mathbb{R}^d),$$

gives the expected number, per unit time, of jumps of X whose amplitude belongs to H. Even if ν is a measure on $\mathcal{B}(\mathbb{R}^d)$, it is not a probability measure nor it is necessarily finite.

Lemma 1.17. *Let X be a Lévy process with jump measure J and Lévy measure ν. Then*

i) if $H \in \mathcal{B}(\mathbb{R}^d)$, $0 \notin \bar{H}$, then the process

$$t \mapsto J_t(H) := J([0,t] \times H) = \#\{s \in]0,t] \mid \Delta X_s \in H\}$$

is a Poisson process with intensity $\nu(H)$ and the compensated process

$$t \mapsto \tilde{J}_t(H) = J_t(H) - t\nu(H)$$

is a martingale;

ii) if $H \in \mathcal{B}(\mathbb{R}^d)$, $0 \notin \bar{H}$, and f is a measurable function, then the process

$$t \mapsto J_t(H, f) := \int_0^t \int_H f(s,x) J(ds, dx) = \sum_{0<s\leq t} f(s, \Delta X_s) \mathbf{1}_H(\Delta X_s) \quad (1.11)$$

is a compound Poisson process;

iii) if f, g are measurable functions and H, K are disjoint Borel sets such that $0 \notin \bar{H} \cup \bar{K}$, then the processes $J_t(H, f), J_t(K, g)$ are independent.

Proof. See [6]. □

Let X be a Lévy process with jump measure J and Lévy measure ν. For any Borel function $f = f(t,x)$ on $\mathbb{R}_{\geq 0} \times \mathbb{R}^d$ one can construct the integral

$$\int_0^t \int_{\mathbb{R}^d} f(s,x) J(ds, dx)$$

with respect to the random measure J proceeding as in the deterministic case. In case the number of jumps is finite, this definition coincides with that given by (1.9) or (1.11).

Given these remarks, the results of Theorem 1.16 can be generalized.

Theorem 1.18. *Let $(X_t)_{t\geq 0}$ be a d-dimensional Lévy process with Lévy measure ν and jump measure J. For any measurable function f such that*

$$\int_0^t \int_{|x|\leq \varepsilon} |f(s,x)|\nu(dx)ds < \infty$$

for some $\varepsilon > 0$, we have

$$\int_0^t \int_{\mathbb{R}^d} f(s,x)J(ds,dx) = \sum_{0<s\leq t,\ \Delta X_s \neq 0} f(s,\Delta X_s) < \infty \quad a.s.$$

If $f \in L^1([0,+\infty[\times\mathbb{R}^d, ds\otimes\nu)$, then the process

$$M_t = \int_0^t \int_{\mathbb{R}^d} f(s,x)\tilde{J}(ds,dx)$$

is a martingale and $E[M_t] = 0$. Moreover, if $f \in L^2([0,+\infty[\times\mathbb{R}^d, ds\otimes\nu)$ then $M_t \in L^2$ and we have

$$var(M_t) = \int_0^t \int_{\mathbb{R}^d} f^2(s,x)\nu(dx)ds.$$

Proof. See [6]. □

The structure of the paths of a Lévy process can be characterized by the following

Theorem 1.19. (Lévy-Itô decomposition) *Let $(X_t)_{t\geq 0}$ be a d-dimensional Lévy process with jump measure J_t and Lévy measure ν. Then the Lévy measure ν satisfies*

$$\int_{|x|\geq 1} \nu(dx) < \infty,$$

$$\int_{|x|<1} |x|^2\nu(dx) < \infty.$$

Moreover, there exists a d-dimensional correlated Brownian motion W and, for any $R > 0$, there exists $\mu_R \in \mathbb{R}^d$ such that

$$X_t = \mu_R t + W_t + X_t^R + M_t^R \tag{1.12}$$

where

$$X_t^R = \int_0^t \int_{|x|\geq R} xJ(ds,dx),$$

$$M_t^R = \int_0^t \int_{|x|<R} x\tilde{J}(ds,dx),$$

and $\tilde{J}$ denotes the compensated jump measure. The terms in (1.12) *are independent.*

Proof. See [7] Section 2.4. □

Given the decomposition (1.12), a continuous and a jump part can be isolated: the first two terms in (1.12) correspond to a Brownian motion with drift while the remaining terms are discontinuous processes incorporating the jumps of X and only depend on the jump measure. In particular, X_t^R can be expressed as

$$X_t^R = \sum_{0<s\leq t} \Delta X_s \mathbf{1}_{\{|\Delta X_s|\geq R\}},$$

and by lemma 1.17, X^R is a compound Poisson process responsible for the large jumps of X (X^R has a finite number of jumps in $[0,t]$, which correspond to the jumps of X with absolute value larger than R). While M_R is a L^2-martingale which is responsible for the small jumps:

$$M_t^R = \lim_{\varepsilon\to 0^+} \tilde{X}_t^{\varepsilon,R}$$

where

$$\tilde{X}_t^{\varepsilon,R} = \sum_{0<s\leq t,\ \varepsilon\leq|\Delta X_s|<R} \Delta X_s - tE[\Delta X_1 \mathbf{1}_{\{\varepsilon\leq|\Delta X_1|<R\}}]$$

is the compensated compound Poisson process of the jumps of X with size between ε and R. From the results above, it is straightforward to deduce, [[8], Chapter 13, Section 3.3] that for $0 < S \leq R$ it holds

$$\mu_S = \mu_R - \int_{S<|x|\leq R} x\nu(dx), \tag{1.13}$$

when the drift coefficient μ is modified.

By the Lévy-Itô decomposition, every Lévy process is determined by the triplet $(\mu_R, \mathcal{C}, \nu)$ where μ_R is the drift coefficient in (1.12), $\mathcal{C}$ is the covariance matrix of the Brownian motion and ν is the Lévy measure.

Definition 1.20. (Characteristic triplet) *We call $(\mu_R, \mathcal{C}, \nu)$ the* characteristic R-triplet *of X.*

There are two special triplets that allow to separate the continuous from the jump part and the martingale from the drift part of the process X. In particular,

- If the jump part of X has bounded variation, equivalently

$$\int_{|x|\leq 1} |x|\nu(dx) < \infty,$$

such as for the compound Poisson process, we get the following Lévy-Itô

decomposition

$$\begin{aligned} X_t &= \mu_0 t + W_t + \int_0^t \int_{\mathbb{R}^d} x J(ds, dx) \\ &= \mu_0 t + W_t + \sum_{0<s\leq t} \Delta X_s, \end{aligned}$$

where the last term is a pure jump process containing all the jumps of X and has its own drift and martingale parts;

- if

$$\int_{|x|\geq 1} |x| \nu(dx) < \infty,$$

i.e., X has finite expectation, we can let R go to infinity in (1.13) and we get

$$\mu_\infty := \lim_{R\to\infty} \mu_R = \mu_S + \int_{|x|>S} x\nu(dx).$$

Then X has ∞-triplet $(\mu_\infty, \mathcal{C}, \nu)$ and the alternative Lévy-Itô decomposition

$$\begin{aligned} X_t &= \mu_\infty t + W_t + \int_0^t \int_{\mathbb{R}^d} x \tilde{J}(ds, dx) \\ &= \mu_\infty t + W_t + \big(\sum_{0<s\leq t} \Delta X_s - tE[\Delta X_1] \big). \end{aligned}$$

The last term is a martingale (it is a process with a jump part that is compensated by a continuous part) and therefore the drift of X is entirely contained in the term $\mu_\infty t$.

It is always possible to split a Lévy process into the sum of a martingale with bounded jumps and a process with bounded variation.

Corollary 1.21. *Let X be a Lévy process. Then $X = M + Z$ where M and Z are Lévy processes, M is a martingale such that $M_t \in L^p(\Omega)$ for any $p \geq 1$ and Z has (locally in time) bounded variation.*

Proof. See [6]. □

In practical terms, the second and fourth terms in (1.12) (i.e. Brownian motion and compensated small jumps) form the martingale part of X, while the first and third terms (i.e. drift term and large jumps) govern the drift of the process.

It is also possible to show that any local martingale X can be written $X = M + Z$ where M is a local martingale with bounded jumps and Z has (locally in time) bounded variation.

The Lévy-Itô decomposition has also fundamental practical implications since every Lévy process can be approximated with arbitrary precision by a jump-diffusion process which is an independent sum of a Brownian motion with drift and a compound Poisson process. This fact lets the simulation of Lévy processes be more straightforward.

1.2.4 Lévy-Khintchine representation

From Theorem 1.19 it is possible to deduce the most general form for the characteristic exponent of a Lévy process: this is provided by the *Lévy-Khintchine formula*.

Theorem 1.22. (Lévy-Khintchine representation) *Let X be a Lévy process in $\mathbb{R}^d$ with characteristic triplet $(\mu_1, \mathcal{C}, \nu)$. Then we have*

$$\varphi_{X_t}(\xi) = E[e^{i\xi \cdot X_t}] = e^{t\psi_X(\xi)}$$

where the characteristic exponent ψ_X is equal to

$$\psi_X(\xi) = i\mu_1 \cdot -\frac{1}{2}\langle \mathcal{C}\xi, \xi \rangle + \int_{\mathbb{R}^d} (e^{i\xi \cdot x} - 1 - i\xi \cdot x \mathbf{1}_{\{|x|<1\}})\nu(dx).$$

Proof. See [6]. □

An equivalent Lévy-Khintchine representation may be obtained by using the Lévy-Itô decomposition with a generic $R > 0$:

$$\begin{aligned}\psi_X(\xi) = i\mu_R \cdot \xi - \frac{1}{2}\langle \mathcal{C}\xi, \xi \rangle &+ \int_{|x|\geq R} (e^{i\xi \cdot x} - 1)\nu(dx) \\ &+ \int_{|x|<R} (e^{i\xi \cdot x} - 1 - i\xi \cdot x)\nu(dx)\end{aligned}$$

where

$$\mu_R = \mu_1 + \int_{\mathbb{R}^d} x(\mathbf{1}_{\{|x|\leq R\}} - \mathbf{1}_{\{|x|\leq 1\}})\nu(dx).$$

Corollary 1.23. *Let X be a Lévy process with characteristic triplet $(\mu_1, \mathcal{C}, \nu)$ and Lévy measure ν such that*

$$\nu(\mathbb{R}^d) < \infty,$$

then X is a jump-diffusion process with intensity $\lambda = \nu(\mathbb{R}^d)$ and distribution of jumps $\eta = \lambda^1 \nu$.

Proof. See [6]. □

The integrability condition, i.e.,

$$\int_{|x|\leq 1} |x|\nu(dx) < \infty,$$

characterizes the Lévy processes that (up the Brownian term) have the trajectories with bounded variation. Indeed, we have:

Proposition 1.24. *Let X be a Lévy process with triplet $(\mu_1, \mathcal{C}, \nu)$. Then X has (locally in time) bounded variation if and only if*

$$\mathcal{C} = 0 \quad \textit{and} \quad \int_{|x|\leq 1} |x|\nu(dx) < \infty.$$

Proof. See [6]. □

Let us collect the main results on Lévy processes with bounded variation.

Corollary 1.25. *Let X be a Lévy process with (locally in time) bounded variation and characteristic triplet $(\mu_1, \mathcal{C}, \nu)$. Then we have the Lévy-Itô decomposition*

$$X_t = \mu_0 t + \int_{\mathbb{R}^d} xJ(ds, dx),$$

where

$$\mu_0 = \mu_1 - \int_{|x|\leq 1} x\nu(dx).$$

Moreover the characteristic exponent takes the form

$$\psi_X(\xi) = i\mu_0 \cdot \xi + \int_{\mathbb{R}^d} (e^{i\xi \cdot x} - 1)\nu(dx).$$

Finite activity models, by Corollary 1.23, are based on jump-diffusion processes that are independent sums of a Brownian motion with drift and a compound Poisson process: the jumps are "rare" events and the evolution of the process is similar to that of a diffusion. On the contrary, for an infinite activity process, we have $\nu(\mathbb{R}^d) = \infty$, where ν is the Lévy measure of the process, and it is known that the set of jump times of every trajectory is countable and dense in $\mathbb{R}_{\geq 0}$: in this case, jumps arrive infinitely often and the dynamics of jumps can be considered rich enough to avoid the introduction of the Brownian component. Concerning the construction of Lévy processes, the simplest way to define a Lévy process is via the Lévy-Khintchine representation, that is by giving the characteristic triplet of the process.

1.3 Lévy models with stochastic volatility

In the following, we will focus on models that can be constructed starting from Lévy processes.

Merton (1976) was the first who actually introduced jumps in stock distribution. In his jump-diffusion model, the evolution of the price is driven by a Lévy process with a nonzero Gaussian component and a compound Poisson process with finitely many jumps in every time interval. The Merton jump-diffusion model [9] with Gaussian jumps extends the Black-Scholes model to a model that attempts to capture the negative skewness and excess kurtosis of the log stock price density by a simple addition of a compound Poisson jump process. The Poisson process and the jumps are assumed to be independent of the Brownian motion.

Another extension of the Black-Scholes model based on Lévy processes, is the double exponential jump-diffusion model, proposed by Kou [10]. Merton and Kou models generalize Black-Scholes model introducing jumps but maintaining the independence of log-return. This enables to increase the flexibility of the model in reproducing tail behaviour at various time scales and generates implied volatility smiles and skew that resemble the ones observed in market prices. Unfortunately, the hypothesis of independence of increments cannot be properly deduced from historical

time series of returns. In fact, the amplitude of returns seems to be positively autocorrelated in time, due to the volatility clustering property suggested by the data. From a risk-neutral modelling point of view, independent increments allow to flexibly calibrate implied volatility patterns across strike and maturity but time dependent parameters are still required to include observed term structure of implied volatility. In addition, when the risk-neutral dynamics of the log-price derives from a Lévy process, the implied volatility surface has a deterministic evolution.

These difficulties can be overcome adding another degree of randomness, i.e., introducing a second random process, (σ_t), driving the instantaneous volatility of the underlying:

$$\frac{dS_t}{S_t} = \mu dt + \sigma_t dW_t.$$

$(\sigma_t)_{t\geq 0}$ is a positive, mean-reverting stochastic process and the volatility process σ_t specifies the stochastic volatility model.

If σ_t is driven by a Brownian motion, possibly correlated with W_t, the volatility model is said to be stochastic-based. Stochastic-based volatility models cannot generate jumps nor can give realistic implied volatility patterns in the short-term but do account for volatility clustering, dependence in increments and long term smiles and skews. In this case, the diffusion must be nonlinear in order to have the positiveness of σ_t.

Nevertheless, it is possible to add jumps, both in returns and in volatility, in stochastic volatility models. Adding an independent jump component to the returns improves short-maturity behaviour of implied volatility, preserving long-term smiles (Bates model [11]). On the other hand, using a (positive) Lévy process to drive the volatility σ_t allows to build a positive, mean-reverting volatility process with realistic dynamics maintaining the linearity of the model. These models are analytically tractable but computationally quite involving as soon as "leverage" effects are included. Building models with dependence in increments is possible time changing a Lévy process by a positive increasing process with dependent increments.

The scope of stochastic volatility models is to replicate the erratic behaviour of market volatility and to introduce the dependence in the increments, making flexible modelling of term structure of various quantities possible. In this context, the evolution of the price process is determined both by its value and by the level of volatility. It is then straightforward to see that the price S_t cannot be a Markov process alone. However, increasing in dimension and considering the two-dimensional process (S_t, σ_t), it is possible to regain Markov property. On the one hand, having a Markov process can help in modelling the process while, on the other hand, the addition of an extra source of randomness leads to incomplete market models and non-uniqueness of option prices derived from returns' behaviour.

1.3.1 Stochastic volatility models without jumps

We start introducing bivariate diffusion stochastic volatility models. In such models, the dynamic of the asset price $(S_t)_{t\geq 0}$ satisfies the following SDE:

$$dS_t = \mu S_t dt + \sigma_t S_t dW_t.$$

Here $(\sigma_t)_{t\geq 0}$ is said to be the *instantaneous volatility process.* This framework allows to compute option prices via PDEs and finite difference methods, and the correlation coefficient of the two Wiener processes helps describing in a simple way the correlation between movements in volatility and returns.

The instantaneous volatility process needs to be positive and mean-reverting. The positiveness can be gained defining $\sigma_t = f(y_t)$ with f positive function and y_t random driving process. Mean-reversion is obtained introducing a mean-reverting drift in the dynamics of (y_t):

$$dy_t = \lambda(\eta - y_t)dt + \cdots + d\hat{Z}_t,$$

where $(\hat{Z}_t)_{t\geq 0}$ is a Wiener process correlated with (W_t), λ is the *rate of mean-reversion* and η is the long-run average level of y_t.

In addition, $(\hat{Z}_t)_{t\geq 0}$ and W can be correlated, if the correlation is driven by the *instantaneous correlation coefficient* $\rho \in [-1, 1]$, one can write

$$\hat{Z}_t = \rho W_t + \sqrt{1-\rho^2} Z_t,$$

where (Z_t) is a standard Brownian motion independent of (W_t).

The driving process (y_t) can be chosen to be of the type

- Geometric Brownian motion

$$dy_t = c_1 y_t dt + c_2 y_t d\hat{Z}_t,$$

- Cox-Ingersoll-Ross (CIR)

$$dy_t = \kappa(\eta - y_t)dt + v\sqrt{y_t}d\hat{Z}_t.$$

1.3.2 The square root process

The square root process, or CIR process after Cox, Ingersoll and Ross, is both mean-reverting and positive, hence it is a perfect choice for the driving process. It is defined as the solution of the stochastic differential equation:

$$y_t = y_0 + \lambda \int_0^t (\eta - y_s)ds + \theta \int_0^t \sqrt{y_s} dW_s. \tag{1.14}$$

where λ, η and θ are positive constants.

From this definition, it is straightforward to see that the process is continuous and positive. The value of the parameters drives the behavior of the process near zero:

- if $\theta^2 \leq 2\lambda\eta$, the process cannot reach zero;
- if $\theta > 2\lambda\eta$, the process can touch zero and, in these cases, it reflects.

Expectation can be computed starting from equation (1.14), we obtain

$$E[y_t] = y_0 + \lambda \int_0^t (\eta - E[y_s])ds,$$

whose solution is given by

$$E[y_t] = \eta + (y_0 - \eta)e^{-\lambda t}.$$

Here, η represents the long-term mean of the process and λ is responsible of the rate of mean-reversion.

Analogously, variance is given by

$$var(y_t) = \frac{\theta^2\eta}{2\lambda} + \frac{\theta^2(y_0 - \eta)}{\lambda}e^{-\lambda t} + \frac{\theta^2(\eta - 2y_0)}{2\lambda}e^{-2\lambda t}.$$

In the limit $t \to \infty$, or in the stationary case, the variance reduces to $\frac{\theta^2\eta}{2\lambda}$ and it is driven by θ, i.e. the *volatility of volatility*.

Integrating the volatility process, we obtain

$$Y_t = \int_0^t y_s ds.$$

Given the positiveness of y_t, Y_t is an increasing process and its mean is given by

$$E[Y_t] = \int_0^t E[y_s]ds = \eta t + \frac{(y_0 - \eta)(1 - e^{-\lambda t})}{\lambda}.$$

The Laplace transform of Y_t is known in closed form:

$$E[e^{-uY_t}] = \frac{exp\big(\frac{\lambda^2\eta t}{\theta^2}\big)}{\big(\cosh\frac{\gamma t}{2} + \frac{\lambda}{\gamma}\sinh\frac{\gamma t}{2}\big)^{\frac{2\lambda\eta}{\theta^2}}} \exp\Big(-\frac{2y_0 u}{\lambda + \gamma\coth\frac{\gamma t}{2}}\Big),$$

where $\gamma = \sqrt{\lambda^2 + 2\theta^2 u}$.

1.3.3 Stochastic volatility models with jumps: The Bates model

The Bates jump-diffusion stochastic volatility model is an improvement of diffusion-based stochastic volatility models. The Bates models adds proportional log-normal jumps to the Heston stochastic volatility model generating sufficient variability and asymmetry in short-terms returns to match implied volatility skew for short maturities. This model is driven by the following equations:

$$\begin{cases} \frac{dS_t}{S_t} & = \mu dt + \sqrt{V_t}dW_t^S + dZ_t, \\ dV_t & = \xi(\eta - V_t)dt + \theta\sqrt{V_t}dW_t^V, \end{cases} \tag{1.15}$$

with (W_t^S) and (W_t^V) correlated Brownian motions, having correlation coefficient ρ, driving price and volatility, Z_t a compound Poisson process with intensity λ and log-normal distribution of jump sizes such that if k is its jump size then $\ln(1+k) \sim \mathcal{N}(\ln(1+\bar{k}) - \frac{1}{2}\delta^2, \delta^2)$. Under the risk-neutral probability, the no-arbitrage condition gives the drift $\mu = r - \lambda\bar{k}$. The equation for the log-price

$X_t = \ln S_t$ derives from (1.15) and Itô's lemma:

$$dX_t = \Big(r - \lambda\bar{k} - \frac{1}{2}V_t\Big)dt + \sqrt{V_t}dW_t^S + d\tilde{Z}_t,$$

where $(\tilde{Z}_t)$ is a compound Poisson process with intensity λ and Gaussian distribution of jump sizes.

This model can be viewed as a generalization of both Heston stochastic volatility model, adding jumps, and Merton jump-diffusion model, allowing for stochastic volatility. As we discussed above, in the Bates model the no arbitrage condition fixes the drift of the price process while changes in other model's parameters lead to different risk-neutral measures. It is worth noticing also that in the Bates model there is an additional degree of freedom in choosing the log-price dynamics. In fact, jumps in the log-price are not deemed to be Gaussian, they can follow any distribution with computable characteristic function, without losses in tractability.

Then, in this stochastic volatility model, we assume that the characteristic function of the log-price is known in closed form. Therefore, European options can be priced using Fourier transform methods while for path-dependent options it is necessary to turn to numerical methods, see [5], [4].

Inspecting the implied volatility patterns in the Bates model, we observe that there are two ways to generate an implied volatility skew:

- introducing a (negative) correlation between returns and volatility (in analogy with diffusion-based stochastic volatility models);
- in the case of short-term options, introducing asymmetric jumps (as in exp-Lévy models).

Correlation and jumps act in the same way on the implied volatility smile. The smiles due to jumps are more pronounced for short maturities and flatten as time to maturity increases, while smiles due to correlation cannot reproduce prices of short-maturity options but make it possible to obtain skews for long maturities. Then, jumps result in having more impact in increasing the level of implied volatility than correlation. The huge improvement introduced by Bates model in terms of calibration is that the implied volatility patterns of long or short term options can be adjusted separately, leading to sufficiently reasonable results.

1.4 The impact of jumps in volatility and returns

In the previous sections we suggest that the introduction of stochastic volatility leads to realistic implied volatility patterns across long maturities without introducing strong time variation of parameters. However, stochastic volatility models cannot give realistic behavior of implied volatility for short maturities. Adding jumps in returns to stochastic volatility models allows to calibrate the implied volatility surface across different strikes and maturities using parameters without an explicit time dependence. The calibration parameters obtained are sufficiently stable through time due to the fact that increments of the price process are not completely independent; then forward smiles can be computed, with a certain degree of freedom, without being

influenced by present smile. Models having both diffusive stochastic volatility and jumps in return, such as the Bates model, cannot capture all the empirical features of option prices, as highlighted in [12], [13] and [14]. Empirically, the conditional volatility of returns increases and decreases rapidly and it is hard to reproduce such behavior using only a diffusive specification for volatility and jumps in returns. In this section, we introduce continuous-time stochastic volatility models with jumps in return and volatility. Jumps in returns have a transient effect on returns while the effect of diffusive volatility is persistent and, being driven by a Brownian motion, it increases gradually trough small and normally distributed increments. Jumps in volatility leads to rapid and persistent shocks in the conditional volatility of returns. Determining the contributions of jumps in period of market stress can help in the estimation of premium needed to hedge jump risk.

Below, we consider the model introduced by Duffie, Pan and Singleton [1] with jumps in volatility and returns. This model could also be declined in two other models: one with contemporaneous arrivals and correlated jump sizes and another with independent arrivals and sizes. From an empirical point of view, adding jumps in returns enables to explain a significant part of the total variance returns, such jumps are rare events of large size that reproduce large and infrequent drawdowns. Jumps in volatility allow volatility to increase instantaneously and then mean-revert back to its long-run level, highlighting the persistent effect of jumps in volatility on returns' behavior. Jumps in volatility and returns are more incisive than diffusive stochastic volatility models in generating crashlike movements and high variation in volatility, thus the contribution of jump components in period of market stress cannot be neglected and they must play an important role in defining risk premia. Adding jumps in returns steepens the slope of the implied volatility curve, the addition of jumps in volatility further steepens implied volatility curves and increases implied volatility for in-the-money options.

1.4.1 The "double-jump"-diffusion model

In this section we focus on the model proposed by Duffie, Pan and Singleton [1].

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\mathcal{F}_t)$ an information filtration. Given a state space $\mathcal{D} \subset \mathbb{R}^n$, let X be a Markov process solving the stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dZ_t,$$

with W standard Brownian motion in $\mathbb{R}^n$, $\mu : \mathcal{D} \to \mathbb{R}^n$, $\sigma : \mathcal{D} \to \mathbb{R}^{n\times n}$, and Z pure jump process having jump's fixed probability distribution ν on $\mathbb{R}^n$ and intensity of arrivals $\{\lambda(X_t) : t \geq 0\}$, for $\lambda : \mathcal{D} \to [0, \infty)$. We observe that Z has the jump times of a Poisson process with intensity $\{\lambda(X_s) : 0 \leq s \leq t\}$ varying over time and that the jump size has probability distribution ν and, at a give time T, results to be independent from $\{(X_s) : 0 \leq s < T\}$.

We assume that μ, $\sigma\sigma^T$, and λ, are affine on $\mathcal{D}$ and that X is well defined, with all the needed assumptions and restrictions on $(\mathcal{D}, \mu, \sigma, \lambda, \nu)$.

The Fourier transform of X_t is known in closed form up to the solution of an ordinary differential equation, as specified in [1]. Then, it is possible to recover the distribution of X_t and the prices of options by inverting the transform.

In this framework, it is possible to formulate a generalization of stochastic volatility models with jumps only in returns that can fix the misspecification in the volatility process exposed in Bakshi, Cao, and Chen (1997) [12], Bates (2000) [13], and Pan (2002) [14] and that includes rapidly moving factor driving conditional volatility.

Let us consider a 2-dimensional affine jump-diffusion model, taken a strictly positive price process S of an asset paying dividends at a constant proportional rate $\bar{\zeta}$, define $Y = \ln(S)$. Including the volatility process V, the state process is represented by $X = (Y, V)^\top$. We choose the short rate r to be constant and assume that there is an equivalent martingale measure $\mathcal{Q}$, under which

$$d\begin{pmatrix} Y_t \\ V_t \end{pmatrix} = \begin{pmatrix} r - \bar{\zeta} - \bar{\lambda}\bar{\mu} - \frac{1}{2}V_t \\ \kappa_v(\bar{v} - V_t) \end{pmatrix} dt + \sqrt{V_t}\begin{pmatrix} 1 & 0 \\ \bar{\rho}\sigma_v & \sqrt{1-\bar{\rho}^2}\sigma_v \end{pmatrix} dW_t^{\mathcal{Q}} + dZ_t, \quad (1.16)$$

where $W^{\mathcal{Q}}$ is a standard Brownian motion in $\mathbb{R}^2$ under $\mathcal{Q}$, and Z is a pure jump process in $\mathbb{R}^2$ with constant mean jump-arrival rate $\bar{\lambda}$, whose bivariate jump-size distribution ν has the form θ.

Using the ODE approach exposed in [1], we can calculate, at time t, the transform ψ of the log-price state variable Y_T as

$$\psi(u, (y, v), t, T) = \exp(\bar{\alpha}(T - t, u) + uy + \bar{\beta}(T - t, u)v), \quad (1.17)$$

where, defining $b = \sigma_v\bar{\rho}u - \kappa_v$, $a = u(1-u)$, and $\gamma = \sqrt{b^2 + a\sigma_v^2}$, we have

$$\bar{\beta}(\tau, u) = -\frac{a(1 - e^{-\gamma\tau})}{2\gamma - (\gamma + b)(1 - e^{-\gamma\tau})}, \quad (1.18)$$

$$\bar{\alpha}(\tau, u) = \alpha_0(\tau, u) - \bar{\lambda}\tau(1 + \bar{\mu}u) + \bar{\lambda}\int_0^\tau \theta(u, \bar{\beta}(s, u))ds, \quad (1.19)$$

where

$$\alpha_0(\tau, u) = -r\tau + (r - \bar{\zeta})u\tau - \kappa_v\bar{v}\Big(\frac{\gamma + b}{\sigma_v^2}\tau + \frac{2}{\sigma_v^2}\ln\big[1 - \frac{\gamma + b}{2\gamma}(1 - e^{-\gamma\tau})\big]\Big), \quad (1.20)$$

and where the term $\int_0^\tau \theta(u, \bar{\beta}(s, u))ds$ depends on the formulation of bivariate jump transform $\theta(\cdot, \cdot)$.

Let us now consider the jump transform θ defined by

$$\theta(c_1, c_2) = \bar{\lambda}^{-1}(\lambda^y\theta^y(c_1) + \lambda^v\theta^v(c_2) + \lambda^c\theta^c(c_1, c_2)), \quad (1.21)$$

where $\bar{\lambda} = \lambda^y + \lambda^v + \lambda^c$, and

$$\begin{aligned} \theta^y(c) &= \exp\big(\mu_y c + \frac{1}{2}\sigma_y^2 c^2\big), \\ \theta^v(c) &= \frac{1}{1 - \mu_v c}, \\ \theta^c(c_1, c_2) &= \frac{\exp\big(\mu_{c,y}c_1 + \frac{1}{2}\sigma_{c,y}^2 c_1^2\big)}{1 - \mu_{c,v}c_2 - \rho_J\mu_{c,v}c_1}. \end{aligned}$$

This formulation incorporates three types of jumps:

- jumps in the log-price process Y, with arrival intensity λ^y and normally distributed jump size (mean μ_y and variance σ_y^2);
- jumps in volatility V, with arrival intensity λ^v and exponentially distributed jump size with mean μ_v;
- simultaneous correlated jumps in Y and V, with arrival intensity λ_c. The marginal distribution of the jump size in V is exponential with mean $\mu_{c,v}$.

Conditional on a realization of the jump size in V, say z_v, the jump size in Y is normally distributed with mean $\mu_{c,y} + \rho_J z_v$, and variance $\sigma_{c,y}^2$. This model seems to reproduce the level of skewness implied by the volatility smirk observed in market data.

The authors in [1] also provide explicit option pricing formula in this particular case, i.e.

$$\int_0^\tau \theta(u, \bar{\beta}(s,u))ds = \bar{\lambda}^{-1}\big(\lambda^y f^y(u,\tau) + \lambda^v f^v(u,\tau) + \lambda^c f^c(u,\tau)\big), \tag{1.22}$$

where

$$\begin{aligned}
f^u(u,\tau) &= \tau \exp\big(\mu_y u + \frac{1}{2}\sigma_y^2 u^2\big),\\
f^v(u,\tau) &= \frac{\gamma - b}{\gamma - b + \mu_v a}\tau + \\
&\quad - \frac{2\mu_v a}{\gamma^2 - (b - \mu_v a)^2} \ln\big[1 - \frac{(\gamma + b) - \mu_v a}{2\gamma}(1 - e^{-\gamma\tau})\big],\\
f^c(u,\tau) &= \exp\big(\mu_{c,y} u + \sigma_{c,y}^2 \frac{u^2}{2}\big)d,
\end{aligned}$$

with $a = u(1-u)$, $b = \sigma_v \bar{\rho} u - \kappa_v$, $c = 1 - \rho_J \mu_{c,v} u$, and

$$\begin{aligned}
d =& \frac{\gamma - b}{(\gamma - b)c + \mu_{c,v} a}\tau \\
&- \frac{2\mu_{c,v} a}{(\gamma c)^2 - (bc - \mu_{c,v} a)^2} \ln\big[1 - \frac{(\gamma + b)c - \mu_{c,v} a}{2\gamma c}(1 - e^{-\gamma\tau})\big].
\end{aligned}$$

This specification nests many of the popular models used for option pricing and portfolio allocation applications and reduces to

- Stochastic volatility model with no jumps (SV model), for $\bar{\lambda} = 0$. The SV model (introduced by Heston in 1993) is a pure diffusion model where volatility's behavior is driven by a square-root process.
- Stochastic volatility model with jumps in price only (SVJ-Y), for $\lambda^y > 0$ and $\lambda^v = \lambda^c = 0$. The SVJ model (introduced by Bates in 1996 as a combination of Merton's (1976) jump-diffusion model and the SV model of Heston) has Poisson jump arrivals in returns and normal distributed sizes.

- Stochastic volatility with simultaneous and correlated Poisson jumps in price and volatility (SVJJ), for $\lambda^c > 0$ and $\lambda^y = \lambda^v = 0$.

Duffie et al. suggest that the high value of volatility of volatility in the diffusion component of V for the SVJ-Y model, highlighted also by Bates and Bakshi et al., can be cured allowing for jumps in volatility. In addition, for small values of σ_y^2, the jump sizes of Y and V are nearly perfectly anticorrelated, and this explains the association of jumps down in returns with simultaneous jumps up in volatility found in estimating SV model. These remarks lead to the conclusion that SVJJ model, when compared to SV and SVJ-Y model, better fits market data. Furthermore, the addition of a jump in volatility leads to a more pronounced smirk both at short and long maturities.

Finally we remark how the introduction of a volatility jump component to the SV and SVJ-Y models might affect the volatility smile, and how correlation between jumps in Y and V affects the volatility smirk. Investigating the following three additional cases, Duffie et al. gain a specific idea of the impact of an additional jump component in volatility.

1. SVJ-V model, i.e. the SV model, fitted to market data, is extended by introducing jumps in volatility. The addition of jumps in volatility seems to attenuate the overpricing in the SVJ model (at least for options that are not too far out of the money).

2. SVJ-Y-V model, i.e. SVJ-Y model, fitted to market data, is generalized with possible jumps in volatility. The addition of a jump in V to the SVJ model also attenuates the over-pricing of OTM calls.

3. Finally, the fitted SVJJ model is modified by varying the correlation between simultaneous jumps in Y and V. In the presence of simultaneous jumps, the levels of implied volatilities for OTM calls depend on the sign and magnitudes of the correlation between the jump amplitudes.

2 Markov Chain Monte Carlo Methods

Several studies highlight that Markov chain Monte Carlo methods (MCMC) are particularly well suited to deal with stochastic volatility models. Among all estimation models, the MCMC approach is one of the most computationally efficient and flexible, and it accurately estimates latent volatility, jump sizes and jump times. One of the advantages of using MCMC is that it provides a general methodology that can be applied also in nonlinear and non-Gaussian state models and, given the data, it returns the distribution of both state variables and parameters.

In this section we propose an outline of Markov chain Monte Carlo methods that will be used in the next chapter to extract information about latent state variables, structural parameters and market prices from observed data.

The target of our study is then the Bayesian solution to the inference problem: the posterior distribution $p(\Theta, X|Y)$, i.e., the distribution of parameters and state variables, respectively Θ and X, conditional on observed prices Y.

One of the main tools needed for the following dissertation is the Clifford-Hammersley theorem. This theorem states that a joint distribution can be characterized by its complete conditional distributions and, in our case, it reads: $p(X|\Theta, Y)$ and $p(\Theta|X, Y)$ completely characterize the joint distribution $p(\Theta, X|Y)$. Clearly, it is typically easier to characterize the complete conditional distributions, $p(\Theta|X, Y)$ and $p(X|\Theta, Y)$, then to directly analyze the higher-dimensional joint distribution, $p(\Theta, X|Y)$.

The MCMC algorithm generates a Markov chain over (Θ, X): given the initial draws $X^{(0)}$ and $\Theta^{(0)}$, the g-th draws are obtained iteratively as

$$X^{(g)} \sim p(X|\Theta^{(g-1)}, Y)$$
$$\Theta^{(g)} \sim p(\Theta|X^{(g)}, Y).$$

The sequence of random variables $\{\Theta^{(g)}, X^{(g)}\}_{g=1}^{G}$ obtained is a Markov chain, whose distribution converges to $p(\Theta, X|Y)$ under a number of metrics and mild conditions.

The MCMC algorithms, in general, consist on two steps and

- If the complete conditional distributions are known in closed form and can be directly sampled, MCMC algorithm samples through the so-called *Gibbs steps*.[3]
- Otherwise, the MCMC method translates into the application of *Metropolis-Hastings algorithm*. Here, a candidate draw is sampled from a proposal density and accepted or rejected over the application of an acceptance criterion. The criterion is chosen in order to generate random samples that form a Markov Chain with the appropriate equilibrium distribution.

Given the sample $\{\Theta^{(g)}, X^{(g)}\}_{g=1}^{G}$ from the joint posterior, parameter and state variable estimation can be performed with Monte Carlo method. If $f(\Theta, X)$ is a

[3]The algorithm is addressed to as *Gibbs sampler* if all the conditionals can be sampled through Gibbs steps.

function satisfying technical regularity conditions, the Monte Carlo estimates

$$E[f(\Theta, X)|Y] = \int f(\Theta, X)p(\Theta, X|Y)dXd\Theta \approx \frac{1}{G}\sum_{g=1}^{G} f(\Theta^{(g)}, X^{(g)}).$$

We can, in addition, analyse two types of convergence for $G \to \infty$:

1. The convergence of the distribution of the Markov chain to $p(\Theta, X|Y)$;
2. The convergence of the partial sums

$$\frac{1}{G}\sum_{g=1}^{G} f(\Theta^{(g)}, X^{(g)})$$

to the conditional expectation $E[f(\Theta, X)|Y]$.

Both types of convergence are guaranteed by the Ergodic Theorem for Markov Chains, since MCMC algorithm verifies the statement's holding conditions, see [15].

2.1 Bayesian Inference and Asset Pricing Models

In this section we present, from a Bayesian inference point of view, the principal properties of the elements that play a central role in the analysis of Markov Chain Monte Carlo methods. Bayesian inference provides a coherent approach for inference, as it consists merely on the application of probability laws to model parameters and state variables, and guarantees strong theoretical foundations.

Let X, Y and Θ be respectively: the unobserved state variables $X = \{X_t\}_{t=1}^T$, the observed prices $Y = \{Y_t\}_{t=1}^T$ and the model parameters Θ. The *posterior distribution* can be factorized by Bayes rule into its constituent components as follows:

$$p(\Theta, X|Y) \propto p(Y|X, \Theta)p(X|\Theta)p(\Theta), \tag{2.1}$$

where $p(Y|X, \Theta)$ is called *likelihood function*, $p(X|\Theta)$ represents the distribution of the state variables, and $p(\Theta)$ is the *prior distribution* of the parameters. The posterior can then be considered as a sum of the information, deducible from prices, concerning state variables and parameters.

The full-information likelihood, i.e., the distribution $p(Y|X, \Theta)$, is linked to the marginal likelihood function, $p(Y|\Theta)$, by

$$p(Y|\Theta) = \int p(Y, X|\Theta)dX = \int p(Y|X, \Theta)p(X|\Theta)dX.$$

While in continuous-time asset pricing models $p(Y|\Theta)$ is not always available in closed form (so that it requires simulation methods to perform likelihood-based inference), the full-information likelihood is usually known in closed form.

The presence of the prior distribution, $p(\Theta)$, in the expression of the posterior allows to incorporate, in a consistent manner, nonsample information, e.g. positivity of parameters or beliefs over the degree of mispricing in a model. Statistically, the

prior can impose stationarity, rule out near unit-root behavior, or separate mixture components.

The posterior distribution plays a central role also in the process of decision making. When facing with a decision problem in the presence of uncertainty, a rational decision maker chooses an action, a, to maximize expected utility $E[U]$, where

$$E[U] = \int U(a, \Theta, X) p(\Theta, X|Y) d\Theta dX$$

and $U(a, \Theta, X)$ is the utility in state X, with parameter Θ, and for action a. The uncertainty in the parameters and states must be taken into account by integrating out the uncertainty in these quantities and then maximizing expected utility by choosing the appropriate action.

The marginal posterior distribution contains, instead, the information embedded in the observed data and it is defined by

$$p(\Theta_i|Y) = \int p(\Theta_i, \Theta_{(-i)}, X|Y) dX d\Theta_{(-i)} \tag{2.2}$$

where Θ_i is the i^{th} element of the parameter vector and $\Theta_{(-i)}$ denotes the remaining parameters. The marginal posterior provides estimates (posterior means or medians) and characterizes estimation risk (posterior standard deviations, quantiles or credible sets).

Let now Y^t be the observed prices up to time t, and consider the following posterior distributions:

$$\begin{aligned} &p(X_t|Y^T),\ t = 1, \ldots, T; \\ &p(X_t|Y^t),\ t = 1, \ldots, T; \\ &p(X_{t+1}|Y^t),\ t = 1, \ldots, T. \end{aligned}$$

The evaluation of $p(X_t|Y^T)$ at time t is a static problem, called the *smoothing problem*, that requires all of the data to be solved. The problem of calculating the second and third distribution, respectively the *filtering and forecasting problem*, can be solved in an inherently sequential way. To filter latent states, we can use once again Bayes rule, obtaining:

$$p(X_t|Y^t) \propto \int p(Y_t|X_t) p(X_t|X_{t-1}) p(X_{t-1}|Y^{t-1}) dX_{t-1},$$

where $p(Y_t|X_t)$ is the likelihood, $p(X_t|X_{t-1})$ is the state evolution and $p(X_{t-1}|Y^{t-1})$ is the prior representing knowledge of the past states, given price information. Simulation based filtering methods, such as the particle and practical filter, provide computationally tractable approaches to approximate the filtering density.

The *model specification* can be evaluated using the posterior distribution, which provides also a method to compare different models: the posterior can be used to analyze the in-sample fit, e.g., it can be used to test the normality of residuals or the independence of random variables, taking into account estimation risk. When there are a finite set of models under consideration, $\{\mathcal{M}_i\}_{i=1}^M$, we can compute the

posterior odds of model i versus j, formally:

$$\frac{p(\mathcal{M}_i|Y)}{p(\mathcal{M}_j|Y)} = \frac{p(Y|\mathcal{M}_i)}{p(Y|\mathcal{M}_j)}\frac{p(\mathcal{M}_i)}{p(\mathcal{M}_j)}.$$

The ratio $p(\mathcal{M}_i)/p(\mathcal{M}_j)$, represents the Bayes factor. If it is greater than one, the data favors model i over model j and viceversa. MCMC provides the output needed to perform formal Bayesian diagnostic via tools such as Odds ratios or Bayes Factors.

The construction of an MCMC algorithm relies on the evaluation of the conditional distribution underlying the likelihood and state dynamics, i.e., $p(Y_{t+1}|Y_t, X_t, \Theta)$ and $p(X_{t+1}|X_t, \Theta)$. The next sections are dedicated to the analysis of these conditional distributions and aim at providing the necessary background to understand the general methodology.

2.1.1 Prices and the Likelihood Function

In general, two different types of likelihood are identified and price dynamics can then be modeled both as the solution to an SDE, as it emerges in models of equity prices or exchange rates, or via a deterministic function between prices and state variables and parameters, as in option pricing and term structure modeling.

In the first case, asset prices solve the parameterized stochastic integral equation

$$Y_{t+1} = Y_t + \int_t^{t+1} \mu_y(Y_s, X_s, \Theta)ds + \int_t^{t+1} \sigma_y(Y_s, X_s, \Theta)dW_s + \sum_{j=N_t}^{N_{t+1}} \xi_j,$$

where the dynamics are driven by the state variables, a vector of Brownian motions $\{W_t\}_{t\geq 0}$, and a vector point process $\{N_t\}_{t\geq 0}$ with stochastic intensity λ_t, and where ξ_j is a jump with $\mathcal{F}_{\tau_j -}$ distribution $\Pi_{\tau_j -}$. We suppose that such random variables are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ and that characteristics have sufficient regularity for a well-defined solution to exist. The distribution implied by the solution of the stochastic differential equation, $p(Y_{t+1}|Y_t, X_t, \Theta)$, generates the likelihood function.

In the second case, at least one of the asset prices is a known function of the state variables and parameters, so we can write $Y_t = f(X_t, \Theta)$. In general, neither the parameters nor the state variables are observed. In multi-factor term structure models, generally the short rate process is a function of a set of state variables, $r_s = r(X_s)$, and bond prices are given by

$$f(X_t, \Theta) = E^Q[e^{-\int_t^T r(X_s)ds}|X_t]$$

where Q is an equivalent martingale measure on the original probability space and the function f can be computed either analytically or as the solution to ordinary or partial differential equation. Thanks to the Fundamental Theorem of Asset pricing, in option pricing models, there exists a probability measure Q, equivalent to P, such that prices are discounted expected values of payoffs under Q. In the case of a call option, it reads

$$f(X_t, \Theta) = E^Q[e^{-\int_t^T r(X_s)ds}(X_T - K)_+|X_t]$$

being $(\cdot)_+$ the positive component.

In option pricing applications, it is common to assume the existence of a pricing error, ε_t. This assumption derives from the fact that parameters are typically unknown and that possible noisy fluctuations can affect prices measurement. In the case of an additive pricing error, we have

$$Y_t = f(X_t, \Theta) + \varepsilon_t$$

for $\varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon)$.

2.1.2 State Variable Dynamics

The state variables, X_t, are also modeled as solutions to stochastic differential equations. The state variables are commonly specified as diffusion models, jump-diffusion models and Markov switching diffusion. In the following we will only examine the case of diffusion and jump-diffusion models; for further information on Markov switching diffusion see [16].

State variables are commonly specified as diffusions, following the work of Black and Scholes [3]. The state variables dynamics are generated by

$$X_{t+1} = X_t + \int_t^{t+1} \mu(X_s, \Theta) ds + \int_t^{t+1} \sigma(X_s, \Theta) dW_s$$

where W_t is a vector of Brownian motions under the P-measure and we assume sufficient regularity on μ, σ and X_0 for a well-behaved solution to exist. Diffusion is then characterized by its continuous sample path and by a Markov structure.

After the work of Merton [9], the continuity assumption has been relaxed adding a marked point process (N_t) to the diffusion component. Such point process counts the number of jump times $\{\tau_j\}_{j=1}^{\infty}$ prior to time t. At each time τ_j, a jump ξ_j arrives and induces a discontinuity in the diffusion of state variables, $X_{\tau_j} - X_{\tau_{j-}} = \xi_j$. The process X_t then solves

$$X_{t+1} = X_t + \int_t^{t+1} \mu(X_s, \Theta) ds + \int_t^{t+1} \sigma(X_s, \Theta) dW_s + \sum_{j=N_t}^{N_{t+1}} \xi_j. \tag{2.3}$$

Here, the dynamics of the state variables are characterized by μ, σ and by the arrival intensity of point process, $\lambda_t = \lambda(X_t)$, and the F_{τ_j-} conditional distribution of the jump sizes, $\Pi(X_{\tau_j-}, \Theta)$.

2.1.3 Parameter Distribution

As already mentioned, the last component of the joint posterior distribution is represented by $p(\Theta)$, the prior distribution of Θ, which contains non-sample information regarding the parameters. In literature, parameterized distribution are preferred and this leads to the necessity of choosing both a distribution for the prior and the parameters that index the distribution. The choice of distribution and prior

parameters allows the imposition of non-sample information or, alternatively, little information.[4]

It is convenient to use standard conjugate prior distributions which provide a way of finding closed-form, easy to simulate, conditional posteriors. A conjugate prior is a distribution for which the conditional posterior is the same distribution with different parameters.[5]

The use of informative priors is legitimated by statistical and economic motivations: consider Merton's jump diffusion model for log-returns $Y_t = \log(S_{t+\Delta}/S_t)$, [9]. In this model, returns are given by

$$Y_t = \mu + \sigma(W_{t+\Delta} - W_t) + \sum_{j=N_t}^{N_{t+\Delta}} \xi_j \tag{2.4}$$

and the jump sizes are normally distributed, $\mathcal{N}(\mu_J, \sigma_J^2)$. Here, the maximum likelihood estimator is not defined as the likelihood takes infinite values from some parameters. In this case, it is necessary to use at least partially informative priors to overcome the aforementioned likelihood degeneracies.

Informative priors can be also used to impose stationarity on the state variables. In the stochastic volatility model discussed in the previous chapter, κ_v is often chosen to be very small introducing near-unit root behavior and it is possible to obtain stationarity imposing mean-reversion: this enters via the prior on the speed of mean reversion that forces κ_v to be positive and bounded away from zero. Stationarity is a useful property in practical applications such as option pricing or portfolio formation.

The impact of specific prior parameters on the parameter posterior is usually evaluated through performing sensitivity analysis.

2.1.4 Time-Discretization

From the preceding discussion, it emerges that the state variable dynamics and the likelihood, $p(X_{t+1}|X_t, \Theta)$ and $p(Y_{t+1}|Y_t, X_t, \Theta)$, are both abstractly given as conditional distribution arising from the solution of stochastic differential equations. Unfortunately, the transition densities of the prices or the state variables are known in closed form only in few simple cases, e.g., a square root process, Gaussian process or geometric Brownian motion.

Let us consider a diffusive specification for the state variables and let Δ be the time interval between two observations. The conditional distribution of $X_{t+\Delta}$, given

[4]Priors that provide little or no information regarding the location of the parameters are called *uninformative or diffuse priors.*

[5]For example, if we consider a geometric Brownian motion model for returns, then continuously compounded returns, Y_t, are normally distributed, $Y_t \sim \mathcal{N}(\mu, \sigma^2)$. Assuming a normal prior on μ, $\mu \sim \mathcal{N}(a, A)$, the conditional posterior distribution $p(\mu|\sigma^2, Y)$ is also normally distributed, $\mathcal{N}(a^*, A^*)$, where the starred parameters depend on the data, sample size and on a and A. Here, the posterior mean is a weighted combination of the prior mean and the sample information, with the weights determined by the relative variances. Choosing A to be very large generates an uninformative prior.

X_t, is generated by

$$X_{t+\Delta} = X_t + \int_t^{t+\Delta} \mu(X_s, \Theta)ds + \int_t^{t+\Delta} \sigma(X_s, \Theta)dW_s.$$

If the drift and diffusion functions are continuous functions of the state, for short time increments we can assume that

$$\int_t^{t+\Delta} \mu(X_s, \Theta)ds \approx \mu(X_t, \Theta)\Delta$$
$$\int_t^{t+\Delta} \sigma(X_s, \Theta)dW_s \approx \sigma(X_t, \Theta)(W_{t+\Delta} - W_t).$$

This observation leads to the following approximation for the state variables:

$$X_{t+\Delta} = X_t + \mu(X_t, \Theta)\Delta + \sigma(X_t, \Theta)(W_{t+\Delta} - W_t).$$

The "Euler" discretization built in this way implies that the induced distribution of the state increments is conditionally normal,

$$p(X_{t+\Delta} - X_t|X_t, \Theta) \sim \mathcal{N}(\mu(X_t, \Theta)\Delta, \Sigma(X_t, \Theta)\Delta),$$

where $\Sigma = \sigma\sigma'$ and the state dynamics $p(X|\Theta)$ are given by the products of normal distributions.

In the case of a jump-diffusion, (2.3), we time-discretize the point process N_t that generates jump times. Note that N_t verifies

$$Prob(N_{t+\Delta} - N_t = 1) \approx \lambda_t\Delta.$$

Defining an indicator variable $J_{t+\Delta}$ such that $J_{t+\Delta} = 1$ (with probability $\lambda_t\Delta$), the jump size distribution is approximated by $\xi_{t+\Delta} \sim \Pi(X_t, \Theta)$ and the time-discretization of the jump-diffusion model becomes:

$$X_{t+\Delta} = X_t + \mu(X_t, \Theta)\Delta + \sigma(X_t, \Theta)(W_{t+\Delta} - W_t) + J_{t+\Delta}\xi_{t+\Delta}.$$

Given the jump-diffusion model discretization, the state space can be expanded to include the jump times and the jump sizes, building a jump-augmented state vector, $[X_t, J_t, \xi_t]$. The increments are normally distributed, conditional on current state and the jump times and sizes:

$$X_{t+\Delta}|X_t, J^x_{t+\Delta}, \xi^x_{t+\Delta} \sim \mathcal{N}(X_t + \mu_t\Delta + J^x_{t+\Delta}\xi^x_{t+\Delta}, \sigma_t\sigma_t'\Delta) \tag{2.5}$$

where the dependence of the drift and diffusion on the parameters and state variables have been removed. A Markov chain can be time-discretized analogously.

Time-discretization allows the use of standard MCMC techniques since it generates simplified conditional distribution structure. Eraker Joannes and Polson proved in [17] that this Euler approximation does not introduce any sistematic biases in an equity price model with stochastic volatility, jumps in returns and jumps in volatility. In other cases, the Euler approximation may not provide an accurate approximation

to the true dynamics. MCMC solves this problem by "filling in" asset prices or state variable values at times in between observation dates.

2.1.5 Asset Pricing Models

We now present models that conveniently fit into the framework developed above and that are especially well-suited for MCMC estimation.

1. Continuous-time Equity Price Models: Equity prices are typically modeled in continuous-time since this specification often leads to analytically tractable solutions in portfolio allocation and option pricing applications. Initially equity prices were assumed to follow geometric Brownian motion dynamics

$$dS_t = \mu S_t dt + \sigma S_t dW_t^s$$

where W_t^s is a scalar Brownian motion. As we discussed in the first chapter, empirical tests end up with rejecting the geometric Brownian motion model and suggested the use of models with jumps, stochastic expected returns and volatility:

$$dS_t = \mu_t S_{t-} dt + S_{t-}\sqrt{V_{t-}} dW_t^s + d\big(\sum_{j=1}^{N_t^s} S_{\tau_j -}(e^{\xi_j^s} - 1)\big)$$

where the expected returns are typically assumed to follow a Gaussian diffusion process and the volatility is a jump-diffusion:

$$\begin{aligned} d\mu_t &= \kappa_\mu(\theta_\mu - \mu_t)dt + \sigma_\mu dW_t^\mu \\ dV_t &= \kappa_v(\theta_v - V_{t-})dt + \sigma\sqrt{V_{t-}} dW_t^v + d\big(\sum_{j=1}^{N_t^v} \xi_j^v\big). \end{aligned}$$

In this model, the observed data is typically the log-returns, $Y_t = \log(S_t/S_{t-1})$, and the state variables are the time-varying mean and volatility $X_t = [\mu_t, V_t]$. As an alternative, the log-volatility model $d\log(V_t) = \kappa_v(\theta_v - \log(V_t))dt + \sigma_v dW_t^v$ is also popular for empirical applications.

2. Equity index option pricing models: In the previous sections, we presented equity models where the only observed data are the continuously compounded equity returns. Option prices sharpen inference by providing information about the market prices of volatility and jump risks that are embedded only in derivative prices. Including an option adds another level to the state space model:

$$\begin{aligned} C_t &= E^Q[e^{-r(T-t)}(S_T - K)_+ | V_t, S_t] = f(S_t, V_t, K, T-t, \Theta) \\ dS_t &= \mu_t S_{t-} dt + S_{t-}\sqrt{V_{t-}} dW_t^s + d\big(\sum_{j=1}^{N_t^s} S_{\tau_j -}(e^{\xi_j^s} - 1)\big) \\ dV_t &= \kappa_v(\theta_v - V_{t-})dt + \sigma_v\sqrt{V_{t-}} dW_t^v + d\big(\sum_{j=1}^{N_t^v} \xi_j^v\big) \end{aligned}$$

where C_t is the price of a call option with strike at K and maturity T. In this case, the observed data is $Y_t = [C_t, \log(S_t/S_{t-1})]$. The fact that the option price is only known up to a numerical integration poses no problems for an MCMC based estimation approach as shown by Eraker, [18].

2.2 MCMC Methods and Theory

The fundamental building block of Markov Chain Monte Carlo methods is the Clifford-Hammersley theorem. This statement, in its more general version, (Besag (1974), [19]) provides conditions for when a set of conditional distributions characterizes a unique joint distribution. In continuous-time asset pricing models, $p(\Theta, X|Y)$ is in general a complicated, high-dimensional distribution and it is prohibitive to directly generate samples from this distribution. The key idea of MCMC is to break the joint distribution into its complete set of conditionals via Clifford-Hammersley theorem. In this way, the dimensionality of the problem is attacked: conditional distributions are of lower dimension and hence easier to sample. In particular, in our setting, the theorem indicates that $p(\Theta|X, Y)$ and $p(X|\Theta, Y)$ uniquely determine $p(\Theta, X|Y)$.

When the direct sampling from $p(\Theta|X, Y)$ and $p(X|\Theta, Y)$ is not possible or the dimension of the conditional posteriors is still prohibitive, a new application of the Clifford-Hammersley theorem can help to further simplify the problem. Let us consider $p(\Theta|X, Y)$ and suppose that the Kdimensional vector Θ can be partitioned into $k \leq K$ uni- or multidimensional components $\Theta = (\Theta_1, \ldots, \Theta_k)$.

Given the partition, the iterative application of Clifford-Hammersley theorem implies that

$$\begin{aligned} &p(\Theta_1|\Theta_2, \ldots, \Theta_k, X, Y) \\ &p(\Theta_2|\Theta_1, \ldots, \Theta_k, X, Y) \\ &\quad \ldots \\ &p(\Theta_k|\Theta_1, \ldots, \Theta_{k-1}, X, Y) \end{aligned}$$

are conditional distributions that uniquely determine $p(\Theta|X, Y)$.

Analogously, the joint distribution of the state vector $p(X|\Theta, Y)$ can be characterized by its own complete set of conditionals: $p(X_t|\Theta, X_{(-t)}, Y)$ for $t = 1, \ldots, T$ where $X_{(-t)}$ denotes the elements of X excluding X_t. Then for a $T + K$ dimensional posterior, iterating the application of Clifford-Hammersley theorem, it is possible to obtain the same information drawing $T + K$ one dimensional distributions. A proof of the Clifford-Hammersley theorem, based on the Besag formula in [19], uses the insight that for any pair (Θ^0, X^0) of points, the joint density $p(\Theta, X|Y)$ is determined, as long as positivity is satisfied, by

$$\frac{p(\Theta, X|Y)}{p(\Theta^0, X^0|Y)} = \frac{p(\Theta|X^0, Y)p(X|\Theta, Y)}{p(\Theta^0|X^0, Y)p(X^0|\Theta, Y)}. \tag{2.6}$$

Thus, knowledge of $p(\Theta|X, Y)$ and $p(X|\Theta, Y)$, up to a constant of proportionality, is equivalent to knowledge of the joint distribution. The positivity condition in our framework can be translated into the request that for each point in the sample

space, $p(\Theta, X|Y)$ and the marginal distributions have positive mass. This condition is always satisfied under very mild regularity conditions.

We now provide a simple description of the algorithms used in MCMC methods.

2.2.1 Gibbs Sampling

When iterative direct sampling from all of the complete conditionals is possible via standard methods, the resulting MCMC algorithm is a Gibbs sampler. In general, given $(\Theta^{(0)}, X^{(0)})$, a Gibbs sampler is defined by

$$\begin{aligned} &1.\ \text{Draw } \Theta^{(1)} \sim p(\Theta|X^{(0)}, Y) \\ &2.\ \text{Draw } X^{(1)} \sim p(X|\Theta^{(1)}, Y). \end{aligned}$$

Iterating the two steps, the Gibbs sampler generates a sequence of random variables, $\{\Theta^{(g)}, X^{(g)}\}_{g=1}^{G}$, that converges to $p(\Theta, X|Y)$. The algorithm runs until it converges, and then a sample is drawn from the limiting distribution.

If it is not possible to generate direct draws from $p(\Theta|X, Y)$ and $p(X|\Theta, Y)$, the Gibbs sampler becomes:

Given $(\Theta^{(0)}, X^{(0)})$

1. Draw $\Theta_i^{(1)} \sim p(\Theta_i|\Theta_1^{(0)}, \ldots, \Theta_{i-1}^{(1)}, \Theta_{i+1}^{(0)}, \ldots, \Theta_r^{(0)}, X^{(0)}, Y)$ for $i = 1, \ldots, r$
2. Draw $X^{(1)} \sim p(X|\Theta^{(0)}, Y)$.

If the states cannot be drawn in a block, then iterative applications of Clifford-Hammersley can be used to obtain a factorization of $p(X|\Theta, Y)$ into a set of lower dimensional distributions.

The Gibbs sampler requires that one can conveniently draw from the complete set of conditional distributions.

2.2.2 Metropolis-Hastings

When one or more of the conditional distributions cannot be conveniently sampled or when it is impossible to find efficient algorithms for sampling from a known distribution, the Gibbs sampler does not apply. In these cases, we need to use a different technique: the Metropolis-Hastings algorithms.

Let us inspect the case where one of the parameter posterior conditionals, namely $\pi(\Theta_i) := p(\Theta_i|\Theta_{(-i)}, X, Y)$, can be evaluated (as a function of Θ_i), but it is not possible to generate a sample from the distribution. Consider a single parameter and suppose we are trying to sample from a one-dimensional distribution, $\pi(\Theta)$, i.e., we are suppressing the dependence of other parameters and states in the conditional posterior, $p(\Theta_i|\Theta_{(i)}, X, Y)$. To generate samples from $\pi(\Theta)$, a Metropolis-Hastings algorithm requires the specification of a recognizable proposal or candidate density $q(\Theta^{(g+1)}|\Theta^{(g)})$. This distribution will generally depend on the other parameters, the state variables and the previous draws for the parameter being drawn. As in Metropolis, et al., [20], we only require that we can easily evaluate density ratio

$\pi(\Theta^{(g+1)})/\pi(\Theta^{(g)})$ and this assumption is satisfied in the majority of continuous-time models.

The Metropolis-Hastings algorithm samples iteratively as in the Gibbs sampler, but it first draws a candidate point that will be accepted or rejected based on the acceptance probability. The Metropolis-Hastings algorithm consists of the following two stage procedure:

$$\text{Step 1 : Draw } \Theta_i^{(g+1)} \text{ from the proposal density } q(\Theta_i^{(g+1)}|\Theta_i^{(g)}) \tag{2.7}$$

$$\text{Step 2 : Accept } \Theta_i^{(g+1)} \text{ with probability } \alpha(\Theta_i^{(g+1)}, \Theta_i^{(g)}) \tag{2.8}$$

where

$$\alpha(\Theta_i^{(g+1)}, \Theta_i^{(g)}) = \min\Big(\frac{\pi(\Theta_i^{(g+1)})/q(\Theta_i^{(g+1)}|\Theta_i^{(g)})}{\pi(\Theta_i^{(g)})/q(\Theta_i^{(g)}|\Theta_i^{(g+1)})}, 1\Big). \tag{2.9}$$

Specifically, implementing Metropolis-Hastings requires: drawing a candidate $\hat{\Theta}_i$ from $q(\Theta_i|\Theta_i^{(g)})$, drawing $u \sim Uniform[0,1]$, accepting the draw, that is, set $\Theta_i^{(g+1)} = \hat{\Theta}_i$ if $u < \alpha(\Theta_i^{(g)}, \Theta_i^{(g+1)})$, and otherwise rejecting the draw, that is, set $\Theta_i^{(g+1)} = \Theta_i^{(g)}$. This algorithm splits the conditional distribution into two parts: a recognizable distribution to generate candidate points and an unrecognizable part from which the acceptance criteria arise. The acceptance criterion insures that the algorithm has the correct equilibrium distribution. Continuing in this manner, the algorithm generates samples $\{\Theta^{(g)}\}_{g=1}^{G}$ whose limiting distribution is $\pi(\Theta)$.

It is then straightforward to see that:

1. Gibbs sampling is a special case of Metropolis-Hastings, where $q(\Theta^{(g+1)}|\Theta^{(g)}) \propto \pi(\Theta^{(g+1)})$. This implies that the acceptance probability is always one and the algorithm always moves;

2. The Metropolis-Hastings algorithm allows the functional form of the density to be non-analytic, as occurs when pricing functions require the solution of partial or ordinary differential equations. It is sufficient to evaluate the true density at two given points;

3. When there are constraints in the parameter space, one can just reject these draws. In addition, sampling can be done conditional on specific regions, providing a convenient approach for analyzing parameter restrictions imposed by economic models.

Note that the choice of proposal density can affect the performance of the algorithm: if the proposal density has thin tails relative to the target, the algorithm may converge slowly. In some cases, the algorithm may never converge, getting stuck in a region of the parameter space.

There are two important special cases of the general Metropolis-Hastings algorithm which deserve special attention.

Metropolis-Hastings algorithm can draw $\Theta^{(g+1)}$ directly from proposal density, $q(\Theta^{(g+1)}|\Theta^{(g)})$, which has a dependence from the previous Markov state $\Theta^{(g)}$ (and, in general, other parameters and states) or from a distribution independent of the

previous state, $q(\Theta^{(g+1)}|\Theta^{(g)}) = q(\Theta^{(g+1)})$. The second is known as an *independence Metropolis-Hastings* algorithm:

$$\text{Step 1 : Draw } \Theta_i^{(g+1)} \text{ from the proposal density } q(\Theta_i^{(g+1)}) \tag{2.10}$$

$$\text{Step 2 : Accept } \Theta_i^{(g+1)} \text{ with probability } \alpha(\Theta_i^{(g+1)}, \Theta_i^{(g)}) \tag{2.11}$$

where

$$\alpha(\Theta_i^{(g+1)}, \Theta_i^{(g)}) = \min\Big(\frac{\pi(\Theta_i^{(g+1)})/q(\Theta_i^{(g+1)})}{\pi(\Theta_i^{(g)})/q(\Theta_i^{(g)})}, 1\Big).$$

The candidate draws, $\Theta^{(g+1)}$, are then drawn independently from the previous state, but in general the sequence $\{\Theta^{(g)}\}_{g=1}^{G}$ will not be independent while the acceptance probability depends on previous draws. When using independence Metropolis, it is common to pick the proposal density to closely match certain properties of the target distribution.

The original algorithm considered by Metropolis, et al. (1953), [20], is the so-called *Random-walk Metropolis*. It draws a candidate from $\Theta^{(g+1)} = \Theta^{(g)} + \varepsilon_t$, where ε_t is an independent mean zero error term (e.g. a symmetric density function with fat tails, like a t-distribution). The choice of the proposal density is generic, ignoring the structural features of the target density and the symmetry in the proposal density, $q(\Theta^{(g+1)}|\Theta^{(g)}) = q(\Theta^{(g)}|\Theta^{(g+1)})$, leads to a simplification of the algorithm:

$$\text{Step 1 : Draw } \Theta_i^{(g+1)} \text{ from the proposal density } q(\Theta_i^{(g+1)}|\Theta_i^{g}) \tag{2.12}$$

$$\text{Step 2 : Accept } \Theta_i^{(g+1)} \text{ with probability } \alpha(\Theta_i^{(g+1)}, \Theta_i^{(g)}) \tag{2.13}$$

where

$$\alpha(\Theta_i^{(g+1)}, \Theta_i^{(g)}) = \min\Big(\frac{\pi(\Theta_i^{(g+1)})}{\pi(\Theta_i^{(g)})}, 1\Big).$$

In random walk Metropolis-Hastings algorithms, the variance of the error term is under control and the algorithm must be tuned, by adjusting the variance of the error term, to obtain an acceptable level of accepted draws, generally in the range of 20-40%.

2.2.3 Convergence Theory

The sequence of draws for parameters and state variables generated by the MCMC algorithm, is constructed to be a Markov chain characterized by its starting value $\Theta^{(0)}$ and its conditional distribution or transition kernel $P(\Theta^{(g+1)}, \Theta^{(g)})$, where we abstract from the latent variables.

As exposed in [15], the ergodic theory of Markov chains is the framework in which it is possible to inspect convergence properties of this sequence.

In general, a Markov chain is characterized by its g-step transition probability,

$$P^{(g)}(x, A) = Prob\{\Theta^{(g)} \in A | \Theta^{(0)} = x\}$$

If the chain is irreducible and aperiodic, then it has a unique equilibrium or stationary

distribution, π.

Given a Markov chain with invariant distribution π, we say that the chain is:

- *irreducible* if, for any initial state, it has positive probability of eventually entering any set which has π-positive probability;
- *aperiodic* if there are no portions of the state space that the chain visits at regularly spaced time intervals.

If an irreducible and aperiodic chain has a proper invariant distribution, then π is unique and it is the equilibrium distribution of the chain. That is

$$\lim_{g\to\infty} Prob\{\Theta^{(g)} \in A|\Theta^{(0)}\} = \pi(A).$$

The general theory of Markov chains also provides explicit convergence rates, which in many cases results to be geometric, see Meyn and Tweedie, [21].

Even though it could be difficult to verify the convergence of Markov chains, chains generated by Metropolis-Hastings algorithms present properties which allow convergence conditions to be verified in general, avoiding references to the specifics of the algorithm.

The target distribution π for a Metropolis-Hastings algorithm is given and proper, since it is a posterior distribution. The invariance of the distribution π is verified through the analysis of the detailed balance (time-reversibility) condition. A transition function P is said to satisfy the detailed balance condition if there exists a function π such that

$$P(x, y)\pi(x) = P(y, x)\pi(y)$$

for any points x and y in the state space, meaning that, if the chain is stationary, the probability that the chain reaches a point x from a point y, starting at y, equals the probability of reaching y from x starting at x. It follows that π is the invariant distribution, since $\pi(y) = \int P(x, y)\pi(dx)$.

For the Gibbs sampler, time-reversibility derives directly from Clifford-Hammersley theorem. The Gibbs sampler cycles through the one-dimensional conditional distributions generating the following transition density:

$$P(x, y) = \prod_{i=1}^{k} p(x_i|x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_k).$$

By Cliffor-Hammersley theorem,

$$\frac{\pi(x)}{\pi(y)} = \prod_{i=1}^{k} \frac{p(x_i|x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_k)}{p(y_i|y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_k)}$$

and

$$\frac{\pi(x)}{\pi(y)} = \frac{P(x, y)}{P(y, x)},$$

which is precisely the time-reversibility condition.

For Metropolis-Hastings algorithms, the transition function is given by

$$P(x,y) = \alpha(x,y)Q(x,y) + (1 - r(x))\delta_x(y) \tag{2.14}$$

where $r(x) = \int \alpha(x,y)Q(x,y)dy$ and $Q(x,y) = q(y|x)$. The detailed balance condition holds straightforwardly for both terms in the right hand side of (2.14) as shown above. Thus Gibbs samplers and Metropolis-Hastings algorithms generate Markov chains that are time-reversible and have the target distribution as an invariant distribution.

Also π-irreducibility can be verified: one sufficient condition is that $\pi(y) > 0$ implies that $Q(x,y) > 0$. For the Gibbs sampler, these conditions can be relaxed to the assumption that x and y communicate, which effectively means that starting from x one can eventually reach state y. Since all π-irreducible Metropolis algorithms are Harris recurrent (Tierney (1994), [22]), it is possible to verify aperiodicity, see also Robert and Casella [15]. Hence, there exists a unique stationary distribution to which the Markov chain generated by Metropolis-Hastings algorithms converges and hence the chain is ergodic.

The interest in convergence of the Markov chains is in general strictly theoretical. In practice, we are interested in sample averages of functionals along the chain: in order to estimate the posterior mean for a given parameter, we inspect the convergence of $\frac{1}{G}\sum_{g=1}^{G} f(\Theta^{(g)})$. The two forms of convergence operating are the distributional convergence of the chain, and the convergence of the sample average. Both convergences are provided by the following

Proposition 2.1. (Ergodic Averaging) *Let $\Theta^{(g)}$ be an ergodic chain with stationary distribution π and f a real-valued function with $\int |f|d\pi < \infty$. Then for all $\Theta^{(g)}$ for any initial starting value $\Theta^{(0)}$*

$$\lim_{G\to\infty} \frac{1}{G}\sum_{g=1}^{G} f(\Theta^{(g)}) = \int f(\Theta)\pi(\Theta)d\Theta$$

almost surely.

Proof. See Robert and Casella, [15]. □

Going further, we have the ergodic central limit theorem:

Proposition 2.2. (Central Limit Theorem) *Let $\Theta^{(g)}$ be an ergodic chain with stationary distribution π and f be real-valued such that $\int |f|d\pi < \infty$. Then there exists a real number $\sigma(f)$ such that*

$$\sqrt{G}\Big(\frac{1}{G}\sum_{g=1}^{G} f(\Theta^{(g)} - \int f(\Theta)d\pi\Big)$$

converges in distribution to a mean zero normal distribution with variance $\sigma^2(f)$ for any starting value.

Proof. See Robert and Casella, [15]. □

The geometric convergence of all Gibbs samplers under a minorization condition was proven in Roberts and Polson (1994), [23]. There are various results on the geometric convergence of the Metropolis-Hastings algorithm, which rely on the tail behavior of the target and proposal density. Mengersen and Tweedie (1996), [24], show that a sufficient condition for the geometric ergodicity of independence Metropolis-Hastings algorithms is that the tails of the proposal density dominate the tails of the target, which requires that the proposal density q is such that q/π is bounded over the entire support.

In addition, several studies provide polynomial convergence in a number of different cases. Polynomial convergence is obviously faster than geometric one and guarantees convergence in finite or computing time. It has been proven that MCMC algorithms that draw from specific distributions, such as log-concave, generate polynomial convergent algorithms. Data augmentation can then be used to convert a non-log-concave sampling problem into a log-concave problem. Thus the convergence of MCMC algorithm can be significantly improved via careful data augmentation.

Having presented a brief highlight of the formal convergence theory, it is now worth inspecting more deeply the information content of the sequence $\{\Theta^{(g)}\}_{g=1}^{G}$. Although the convergence is guaranteed from the theory, MCMC algorithms generate chains whose convergence of realized output is impossible to formally diagnose. In this framework, it is important to exploit the informational content of the output of the chain.

Popular observed-chain based diagnostics include:

- calculating parameter trace plots, i.e., plots of $\Theta_i^{(g)}$ versus g, which show the history of the chain simplifying the diagnose for chains that get stuck in a region of the state space;
- the analysis of the correlation structure of draws by computing the autocorrelation function (ACF);
- calculating Monte Carlo estimates for the standard errors of $\frac{1}{G}\sum_{g=1}^{G} f(\Theta^{(g)})$.

2.3 Asset Pricing Applications

Markov chain Monte Carlo methods are particularly well suited for calibrating stochastic volatility models parameters. In these models, in fact, the state space is generally non-Gaussian and nonlinear and, in some cases, cannot be expressed in an analytic form. In the following, we present the MCMC approach to different stochastic volatility models, such as the log-stochastic volatility model, Heston's (1993) square-root stochastic volatility model, [4], and the double-jump model of Duffie, Pan, and Singleton (2000), [1].

2.3.1 Log-Stochastic Volatility Model

The log-stochastic volatility model is defined by:

$$\begin{aligned} d\log(S_t) &= \mu_t dt + \sqrt{V_t} dW_t^s \\ d\log(V_{t+1}) &= \kappa_v(\theta_v - \log(V_t))dt + \sigma_v dW_t^v. \end{aligned}$$

were Brownian motions are assumed to be independent. To abstract from conditional mean dynamics we set $\mu_t = 0$. An Euler time discretization is given by

$$\begin{aligned} Y_t &= \sqrt{V_{t-1}}\varepsilon_t^s \\ \log(V_t) &= \alpha_v + \beta_v \log(V_{t-1}) + \sigma_v \varepsilon_t^v, \end{aligned}$$

where Y_t are the continuously compounded returns, and the reparametrization $\alpha_v = \kappa_v\theta_v$ and $\beta_v = 1 - \kappa_v$ allows us to use standard conjugate updating theory for the parameters.

Define the parameter and state vector respectively as $\Theta := \{\alpha_v, \beta_v, \sigma_v^2\}$ and $X := V = \{V_t\}_{t=1}^T$. By Clifford-Hammersley theorem, $p(\Theta, V|Y)$ is completely characterized by $p(\alpha_v, \beta_v|\sigma_v, V, Y)$, $p(\sigma_v^2|\alpha_v, \beta_v, V, Y)$ and $p(V|\alpha_v, \beta_v, \sigma_v^2, Y)$. Jacquier, Polson and Rossi, [2], assume conjugate priors for the parameters, $\alpha_v, \beta_v \sim \mathcal{N}$ and $\sigma_v^2 \sim \mathcal{IG}$, which implies that

$$p(\alpha_v, \beta_v|\sigma_v, V, Y) \propto \prod_{t=1}^{T} p(V_t|V_{t-1}, \alpha_v, \beta_v, \sigma_v)p(\alpha_v, \beta_v) \propto \mathcal{N}$$

and, for σ_v,

$$p(\sigma_v^2|\alpha_v, \beta_v, V, Y) \propto \prod_{t=1}^{T} p(V_t|V_{t-1}, \alpha_v, \beta_v, \sigma_v)p(\sigma_v^2) \propto \mathcal{IG}.$$

The full joint posterior for volatility is then

$$p(V|\Theta, Y) \propto p(Y|\Theta, V)p(V|\Theta) \propto \prod_{t=1}^{T} p(V_t|V_{t-1}, V_{t+1}, \Theta, Y)$$

for

$$p(V_t|V_{t-1}, V_{t+1}, \Theta, Y) = p(Y_t|V_t, \Theta)p(V_t|V_{t-1}, \Theta)p(V_{t+1}|V_t, \Theta).$$

The conditional variance posterior is a function of V_t, whose distribution is not recognizable and then requires Metropolis-Hastings to sample from it.

Since joint volatility posterior, $p(V|\Theta, Y)$, cannot directly draw from without approximations, we consider a "single state" Metropolis updating scheme. The MCMC algorithm becomes:

$$\begin{aligned} p(\alpha_v, \beta_v|\sigma_v, V, Y) &\sim \mathcal{N} \\ p(\sigma_v^2|\alpha_v, \beta_v, V, Y) &\sim \mathcal{IG} \\ p(V_t|V_{t-1}, V_{t+1}, \Theta, Y) &: Metropolis. \end{aligned}$$

In [2], the states are updated through an independence Metropolis-Hastings.

The first term in the posterior is an inverse Gamma and the second log-normal term can be approximated by a suitable chosen inverse Gamma. This suggests to chose the proposal density, $q(V_t)$, to be a Gamma density. Addressing the true conditional density as $\pi(V_t) := p(V_t|V_{t-1}, V_{t+1}, \Theta, Y)$, the Metropolis-Hastings step is given by:

$$\begin{aligned}&\text{Step 1. Draw } V_t^{(g+1)} \text{ from } q(V_t)\\ &\text{Step 2. Accept } V_t^{(g+1)} \text{ with probability } \alpha(V_t^{(g+1)}, V_t^{(g)})\end{aligned}$$

where

$$\alpha(V_t^{(g+1)}, V_t^{(g)}) = \min\Big(\frac{\pi(V_t^{(g+1)})q(V_t^{(g)})}{\pi(V_t^{(g)})q(V_t^{(g+1)})}, 1\Big).$$

As the gamma distribution bounds the tails of the true conditional density, the algorithm is geometrically convergent.

It is also possible to reproduce the leverage effect adding a correlation, $corr(W_t^v, W_t^s) = \rho$, between the shocks in volatility and price. The leverage effect results in the addition of a correlation between equity returns and changes in volatility and of an additional parameter. To incorporate the leverage effects, in discrete time the model can be written as:

$$\begin{aligned}Y_t &= \sqrt{V_{t-1}}\varepsilon_t^s\\ \log(V_t) &= \alpha_v + \beta_v \log(V_{t-1}) + \sigma_v[\rho\varepsilon_t^s + \sqrt{1-\rho^2}\varepsilon_t^v]\end{aligned}$$

where ε_t^s and ε_t^v are uncorrelated. Considering the reparametrization: $\phi_v = \sigma_v\rho$ and $\omega_v = \sigma_v^2(1-\rho^2)$. Assuming $\alpha_v, \beta_v \sim \mathcal{N}$, $\phi_v \sim \mathcal{N}$ and $\omega_v \sim \mathcal{IG}$, the MCMC algorithm becomes:

$$\begin{aligned}p(\alpha_v, \beta_v|\sigma_v, V, Y) &\sim \mathcal{N}\\ p(\phi_v, \omega_v|\alpha_v, \beta_v, V, Y) &\sim \mathcal{N} - \mathcal{IG}\\ p(V_t|V_{t-1}, V_{t+1}, \Theta, Y) &: Metropolis.\end{aligned}$$

2.3.2 Heston's Square-Root Volatility Model

The Heston square-root stochastic volatility model is defined by:

$$\begin{aligned}dS_t &= S_t(r_t + \eta_v V_t + \frac{1}{2}V_t)dt + S_t\sqrt{V_t}dW_t^s(P)\\ dV_t &= \kappa_v(\theta_v - V_t)dt + \sigma_v\sqrt{V_t}dW_t^v(P).\end{aligned}$$

Let ρ be the constant correlation of Brownian motions. The Euler discretization of the model is given by

$$\begin{aligned}Y_t &= \eta_v V_{t-1} + \sqrt{V_{t-1}}\varepsilon_t^s\\ V_t &= \alpha_v + \beta_v V_{t-1} + \sigma_v\sqrt{V_{t-1}}\varepsilon_t^v\end{aligned}$$

where Y_t are excess returns, and we have re-defined the parameters in the drift process of the volatility process.

Unfortunately, this time-discretization do not share volatility positivity property with the original model: the shocks to V_t are normally distributed and then simulated volatility V_t could be also negative. This shortcoming can be overcome in three different ways:

- The original SDE could be transformed by Ito's lemma into logarithms, and the solution simulated in logs, obtaining:

$$\log(V_t) = h_t$$
$$dh_t = e^{h_t}[k_v(\theta_v - e^{h_t}) - \frac{1}{2}\sigma_v^2]dt + e^{h_t/2}dW_t^v.$$

  Simulating this process in discrete-time in logarithms guarantees that $V_t = \exp(h_t) > 0$ but makes parameters' updating more complicated as volatility appears both in the drift and in the diffusion part.

- Following Eraker, Johannes and Polson [17], for certain time series, the problem can be ignored hoping that it does not affect the results.

- The impact of discretization could be reduced filling in missing data points.

Assume normal independent priors for η_v and (α_v, β_v), an inverted Gamma prior for σ_v^2, and a uniform prior for ρ. The complete conditionals, given by Clifford-Hammersley theorem, are: $p(\alpha_v, \beta_v|\sigma_v, \rho, V, Y)$, $p(\sigma_v^2|\alpha_v, \beta_v, \rho, V, Y)$, $p(\rho|\alpha_v, \beta_v, \sigma_v^2, V, Y)$ and $p(V|\alpha_v, \beta_v, \sigma_v^2, Y)$.

The arising MCMC algorithm is:

$$\begin{aligned}
p(\eta_v|\alpha_v, \beta_v, \sigma_v, \rho, V, Y) &\sim \mathcal{N} \\
p(\alpha_v, \beta_v|\sigma_v, \rho, V, Y) &\sim \mathcal{N} \\
p(\sigma_v^2|\alpha_v, \beta_v, \rho, V, Y) &\sim \mathcal{IG} \\
p(\rho|\alpha_v, \beta_v, \sigma_v^2, V, Y) &: \mathit{Metropolis} \\
p(V_t|V_{t-1}, V_{t+1}, \Theta, Y) &: \mathit{Metropolis}
\end{aligned}$$

The parameter posteriors are analogous to those inspected in the previous paragraph. Eraker, Johannes, and Polson use a random walk Metropolis-Hastings algorithm for both the correlation parameter and the volatility states.

2.3.3 Stochastic volatility with jumps in returns and volatility

Consider the double-jump model of Duffie, Pan, and Singleton, [1], the equity price, S_t, and its stochastic variance, V_t, jointly solve

$$dS_t = S_t(r_t + \eta_v V_t)dt + S_t\sqrt{V_t}dW_t^s(P) + d\Big(\sum_{j=1}^{N_t(P)} S_{\tau_j -}(e^{Z_j^s(P)} - 1)\Big)$$
$$dV_t = \kappa_v(\theta_v - V_t)dt + \sigma_v\sqrt{V_t}dW_t^v(P) + d\Big(\sum_{j=1}^{N_t(P)} Z_j^v(P)\Big)$$

where $W_t^s(P)$ and $W_t^v(P)$ are correlated Brownian motions, $corr(W_t^s(P), W_t^v(P)) = \rho$, $N_t(P) \sim Poisson(\lambda)$, τ_j are the jump times, $Z_j^s(P)|Z_j^v \sim \mathcal{N}(\mu_s + \rho_s Z_j^v, \sigma_s^2)$ are the return jumps, $Z_j^v(P) \sim exp(\mu_v)$ are the variance jumps, and r_t is the spot interest rate.

Eraker, Johannes and Polson estimate stochastic volatility models with jumps in returns and volatility using MCMC methods. Eraker (2004), [18], extends Eraker, Johannes, and Polson to incorporate option prices.

The Euler time-discretization of this model is given by

$$Y_t = \mu + \eta_v V_{t-1} + \sqrt{V_{t-1}}\varepsilon_t^s + J_t Z_t^s$$
$$V_t = \alpha_v + \beta_v V_{t-1} + \sigma_v\sqrt{V_{t-1}}\varepsilon_t^v + J_t Z_t^v.$$

Applying Clifford-Hammersley theorem, parameters and states factorize into the following groups $[(\mu, \eta_v), (\alpha_v, \beta_v), \sigma_v^2, \rho, \lambda, \mu_v, (\mu_s, \rho_s), \sigma_s^2, J, Z^s, Z^v, V]$. We assume normal independent priors for (μ, η_v), (α_v, β_v), and (μ_s, ρ_s), inverted Gamma priors for σ_v^2 and σ_s^2, a Beta prior for λ, a Gamma prior for μ_v, and a uniform prior for ρ.

Due to the modular nature of MCMC algorithms and given the results already presented, the derivation of the conditional posteriors for many of the model's parameters is straightforward: the conditional posteriors for the "diffusive" parameters can be obtained adjusting return and volatility series of Heston model. Conditional on jump times and sizes, we can define the jump-adjusted returns and volatilities to get

$$\tilde{r}_t = Y_t - J_t Z_t^s = \mu + \eta_v V_{t-1} + \sqrt{V_{t-1}}\varepsilon_t^s$$
$$\tilde{V}_t = V_t - J_t Z_t^v = \alpha_v + \beta_v V_{t-1} + \sigma_v\sqrt{V_{t-1}}\varepsilon_t^v$$

which implies that the functional forms conditional posteriors for $(\mu, \eta_v), (\alpha_v, \beta_v), \sigma_v^2$, and ρ are the same as in the previous paragraph and drawing λ is analogous. Conditional on the jump sizes, the parameters of the jump distributions are conjugate.

The MCMC algorithm draws from the conditional parameter posteriors

$$\begin{aligned}
p(\mu, \eta_v | \ldots, J, Z, V, Y) &\sim \mathcal{N} \\
p(\alpha_v, \beta_v | \ldots, J, Z, V, Y) &\sim \mathcal{N} \\
p(\sigma_v^2 | \ldots, J, Z, V, Y) &\sim \mathcal{IG} \\
p(\lambda | J) &\sim \mathcal{B} \\
p(\mu_s, \rho_s | \ldots, J, Z^s, Z^v) &\sim \mathcal{N} \\
p(\sigma_s^2 | \ldots, J, Z^s, Z^v) &\sim \mathcal{IG} \\
p(\mu_v | \ldots, J, Z, V, Y) &\sim \mathcal{G} \\
p(\rho | \alpha_v, \beta_v, \sigma_v^2, V, Y) &\ : \ Metropolis
\end{aligned}$$

and the conditional state variable posteriors

$$\begin{aligned}
p(Z_t^v | \ldots, Z_t^s, J_t, V_t, V_{t-1}) &\sim \mathcal{TN} \\
p(Z_t^s | \ldots, Z_t^s, J_t, Y_t, V_t, V_{t-1}) &\sim \mathcal{N} \\
p(J_t = 1 | \ldots, Z_t^s, Z_t^v, Y_t, V_t, V_{t-1}) &\sim Bernoulli \\
p(V_t | V_{t-1}, V_{t+1}, \Theta, Y) &\ : \ Metropolis
\end{aligned}$$

For both ρ and the volatilities, Eraker, Johannes, and Polson use a random walk Metropolis algorithm, properly tuned, to deliver acceptance rates in the 30-60% range. For additional information see [17] and [18].

3 Cryptocurrency Dynamics

Cryptocurrencies appeared in literature in 2008 when the pseudonymous Satoshi Nakamoto published a white paper, [25], describing the implementation of a digital currency called Bitcoin (BTC) that used blockchain technology. Bitcoin was officially released in 2009 and is the first open source distributed cryptocurrency. Since then it has captured worldwide attention and interest and today, more than twelve years later, hundreds of cryptocurrencies and countless other applications of blockchain technology are readily available. To quantify the importance of this new kind of currencies in today's economy, we recall that, on February 28th, 2021, there were more than 8.600 cryptocurrencies with a total market capitalization of 1,44T USD (according to coinmarketcap.com). Bitcoin dominates today's cryptocurrencies market, representing almost the 60% of the total capitalization.

The revolutionary characteristics of cryptocurrencies are the use of immutable databases, the sophisticated cryptography techniques, that permit safer and traceable trades, and distributed ledger technologies. These aspects are embodied and summarized in the underlying structure, i.e. the blockchain, and there is plenty of papers explaining the mechanics of blockchains and cryptocurrencies, see for example Härdle et al. (2018), [26].

The rise of cryptocurrencies threaten several traditional functions in finance: cryptocurrencies embrace a peer-to-peer mechanism that effectively eliminates intermediaries such as financial institutions and credit cards. This technology leads to the possibility of cheaper, more secure and near real-time transactions without the necessity of entering the traditional banking infrastructure.

In 2018, Lansky, [27], proposed the following definition of cryptocurrency:

Definition 3.1. *Cryptocurrency is a system that meets all of the following conditions:*

1. *The system does not require a central authority, its state is maintained through distributed consensus;*
2. *The system keeps an overview of cryptocurrency units and their ownership;*
3. *The system defines whether new cryptocurrency units can be created. In the affirmative case, the system defines the circumstances of their origin and how to determine the ownership of the new units;*
4. *Ownership of cryptocurrency units can be proved exclusively cryptographically;*
5. *The system allows transactions to be performed in which ownership of the cryptographic units is changed. A transaction statement can only be issued by an entity proving the current ownership of these units;*
6. *If two different instructions for changing the ownership of the same cryptographic units are simultaneously entered, the system performs at most one of them.*

A cryptocurrency can be viewed as a decentralized autonomous organization (DAO), an open-source peer-to-peer digital autonomous network. In this framework,

the money supply is set by an algorithmic rule, and the integrity of the network replaces the dependence on integrity of human participants. The growth of cryptocurrency technology therefore poses a challenge to traditional monetary authorities and central banks.

In addition, cryptocurrencies are the only type of currencies with the following three features:

- Pseudo-anonymity: A user executing cryptocurrency transactions cannot be easily identified. Nonetheless, users may reveal their identity spontaneously or can be identified through the use of external data, and then the cryptocurrency conversely ensures that their transactions are transparent.

- Independence from central authority: Cryptocurrencies are decentralised and independent of central authorities empowered to change the consensus rules of the cryptocurrency system. Any change to the consensus rules can only be achieved by approval of the majority of the cryptocurrency operators. A cryptocurrency cannot be abolished or regulated by force since it is not controlled by a central authority; it can only cease to exist by itself, when users of the cryptocurrency lose confidence in it, e.g., after technical attacks or hacks. Nevertheless, individual users of a cryptocurrency can voluntarily decide for a form of regulation of the transactions executed by them.

- Double spending attack protection: Payments are marked so that cryptocurrency units cannot be used to validate two different transactions.

The rapidly growing interest in cryptocurrencies stimulates the study of its impact, as a new digital asset class, in contemporary financial markets. Since Bitcoin has a dominant role in the market, several studies have started suggesting econometric methods to model the dynamics of BTC prices. Scaillet et al. (2019), [28], show that jumps are much more frequent in the BTC market than, for example, in the US equity market (see e.g., Eraker (2004) [18]), therefore jumps should be considered when modeling BTC prices.

Even if research on the BTC derivatives is still limited, the actual availability of BTC futures and options traded on independent exchange platform (e.g., Deribit and Binance, among others), encourages the creation and formalization of a derivatives market for cryptocurrencies. One of the first step in this direction was the Commodity Futures Trading Commission (CFTC) approval of LedgerX for clearing derivatives, on July 2017.[6] It is not a direct market for options on cryptocurrencies, but provides an initial approximation to relate cryptocurrencies with derivatives. A fundamental turning point was reached in December 2017, when the CME (Chicago Mercantile Exchange) launched Bitcoin futures.[7] After that, also the CBOE (Chicago Board Options Exchange), the largest U.S. options exchange, announced the advent of Cboe Bitcoin Futures.

The limited research on pricing and hedging Bitcoin derivatives is attributed, almost partially, to the fact that pricing BTC derivatives encounters econometric

[6]LedgerX is an institutional trading and clearing platform focused on trading and clearing swaps and options on digital currencies.

[7]The CME is the worlds leading and most diverse derivatives marketplace.

challenges from the extraordinary occurrence of jumps as this market is unregulated, lacks of central settlement and is highly speculation-driven. In addition, Cheah and Fry (2015), [29], and Kristoufek (2015), [30], found that Bitcoin shares both standard financial asset and speculative one properties, and the considerable speculative component makes BTC particularly susceptible to bubbles. It is then vital to understand the asset price formation, given the absence of fundamentals that support the price (such as sales, assets and revenue) compared to some other assets. This highlights the necessity of a flexible model to capture the sudden jumps appearing in both the returns and variance processes.

Following Chen et al. (2020), [31], we inspect and reproduce BTC dynamics using the stochastic volatility with the correlated jump (SVCJ) model of Duffie Pan and Singleton (2000), [1], presented in the first chapter. Several empirical studies have applied the SVCJ model in different markets. For example, Eraker et al. (2003), [17], and Eraker (2004),[18], use the SVCJ model to estimate equity market returns and perform equity option pricing finding empirical evidence of jumps in returns and volatility in the US equity market.

Chen et al. are indeed among the first who extensively inspect the stochastic and econometric properties of Bitcoin and incorporate these properties in the BTC option pricing. The results are relevant in terms of model selection for characterizing the Bitcoin dynamics. The need for incorporating jumps in the returns and volatility processes of BTC is investigated, and they find that jumps play a crucial role in the option prices. This approach seems to be readily applicable to pricing BTC options in reality.

3.1 SVCJ Model Calibration and Results

In order to estimate the Bitcoin dynamics with the SVCJ model, we recall the continuous time model of Duffie Pan and Singleton, [1], presented in Chapter 1.

Let $\{S_t\}$ be the price process, $\{d\log S_t\}$ the log returns and $\{V_t\}$ be the volatility process. The SVCJ dynamics is described by:

$$d\log S_t = \mu dt + \sqrt{V_t} dW_t^{(S)} + Z_t^y dN_t \tag{3.1}$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma_V \sqrt{V_t} dW_t^{(V)} + Z_t^v dN_t \tag{3.2}$$

$$Cov(dW_t^{(S)}, dW_t^{(V)}) = \rho dt \tag{3.3}$$

$$P(dN_t = 1) = \lambda dt \tag{3.4}$$

where κ is the mean reversion rate and θ is the mean reversion level (like in the Cox-Ingersoll-Ross model), $W^{(S)}$ and $W^{(V)}$ are correlated standard Brownian motions with correlation coefficient ρ, N_t is a pure jump process with constant mean jump-arrival rate λ and Z_t^y and Z_t^v are the random jump sizes.

The correlation ρ between the diffusion terms is introduced with the aim to capture the possible leverage effects between returns and volatility. Since the jump-driving Poisson process is the same in both equations (3.1) and (3.2), i.e. log-returns and volatility dynamics, the jump sizes can be correlated. In our framework, the random jump size Z_t^y conditional on Z_t^v is assumed to have a Gaussian distribution

with a mean of $\mu_y + \rho_j Z_t^v$ and standard deviation set to σ_y while the jump in volatility Z_t^v is assumed to follow an exponential distribution with mean μ_v. In formulae:

$$Z_t^y | Z_t^v \sim \mathcal{N}(\mu_y + \rho_j Z_t^v, \sigma_y^2); \quad Z_t^v \sim Exp(\mu_v). \tag{3.5}$$

Also the jumps may be correlated, in this case the correlation term is ρ_j. The stochastic volatility process $\sqrt{V_t}$ is modeled as a square root process. The parameter θ is the long-run mean of V_t if no jumps in volatility occur, and the process reverts to this level at a speed controlled by the parameter κ. The parameter σ_V is referred to as the volatility of volatility, and it measures the variance responsiveness to diffusive volatility shocks. The expected log-return, in the absence of jumps, is measured by the parameter μ.

For the reasons discussed in Chapter 2, we estimate the SVCJ model using the MCMC method. This allows for a wide class of numerical fitting procedures that can be guided by a variation of the priors. Since there are no official BTC options yet, the Markov chain Monte Carlo method is more flexible in estimating the stochastic variance jumps and better reflects the market price of risk. The estimation is based on the following Euler discretization:

$$\begin{aligned} Y_t &= \mu + \sqrt{V_{t-1}}\varepsilon_t^y + Z_t^y J_t \\ V_t &= \alpha + \beta V_{t-1} + \sigma_V \sqrt{V_{t-1}}\varepsilon_t^v + Z_t^v J_t, \end{aligned}$$

where

- $Y_{t+1} = log(S_{t+1}/S_t)$ is the log return;
- $\alpha = \kappa\theta$, $\beta = 1 - \kappa$;
- $\varepsilon_t^y, \varepsilon_t^v$ are $\mathcal{N}(0,1)$-distributed variables with correlation ρ;
- J_t is a Bernoulli random variable with $P(J_t = 1) = \lambda$;
- the jump sizes Z_t^y and Z_t^v are distributed as specified in (3.5).

We define the parameter vector and the latent variance, jump sizes and jump respectively as

$$\begin{aligned} \Theta &= \{\mu, \mu_y, \sigma_y, \lambda, \alpha, \beta, \sigma_V, \rho, \rho_j, \mu_v\} \\ X_t &= \{V_t, Z_t^y, Z_t^v, J_t\}. \end{aligned}$$

Following Chen et al. (2020), [31], we choose the prior distributions specified in Asgharian and Nossman (2011), [32], who estimate a large group of international equity market returns with jump-diffusion models using the MCMC method, specifically,

$$\begin{array}{ccc} \mu \sim \mathcal{N}(0, 25) & \mu_y \sim \mathcal{N}(0, 100) & \sigma_y^2 \sim \mathcal{IG}(10, 40) \\ \lambda \sim Be(2, 40) & (\alpha, \beta) \sim \mathcal{N}(0_{2\times 1}, I_{2\times 2}) & \sigma_V^2 \sim \mathcal{IG}(2.5, 0.1) \\ \rho \sim U(1, 1) & \rho_j \sim \mathcal{N}(0, 0.5) & \mu_v \sim \mathcal{IG}(10, 20) \end{array}$$

where $\mathcal{IG}$ and Be stand for Inverse Gaussian and Beta Distribution respectively. The full posterior distributions of the parameters and the latent-state variables can be found in Asgharian and Nossman (2011), [32].

The reported mean of the posterior that is taken as an estimate of Θ is quite robust relative to changes in variance of the prior distributions, see [31]. The posterior for all parameters except σ_V and ρ are conjugate (i.e. the posterior distribution is analogous to that of the prior but with different parameters). We observe that the posterior for J_t is a Bernoulli distribution, the jump sizes Z_t^y and Z_t^v follow a posterior normal distribution and a truncated normal distribution respectively, and the posteriors for ρ, σ_V^2 and V_t are nonstandard distributions.

It follows that draws for the joint distribution of J_t, Z_t^y and Z_t^v can be easily obtained, while the posteriors for ρ, σ_V^2 and V_t must be sampled using the Metropolis-Hastings algorithm (in our case, random-walk method for ρ and V_t, and independence sampling for σ_V^2). For the estimation of posterior moments, we perform 5000 iterations with a burn-in for the first 1000 simulations, in order to reduce the impact of the starting values.

The SVCJ model discerns returns related to sudden unexpected jumps from large diffusive returns caused by periods of high volatility. For the BTC case the aim is to link the latent historical jump times to news and known interventions. Given the total number of iteration N, the estimates $\hat{J}_t \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} J_t^i$ indicate the posterior probability that there is a jump at time t. In Johannes et al. (1999), [33], a jump is considered to be occurred at t if the estimated jump probability is sufficiently large, i.e., it is greater than an appropriately chosen threshold value ζ:

$$\tilde{J}_t = 1\{\hat{J}_t > \zeta\}, \quad t = 1, 2, \ldots, T.$$

We can choose ζ empirically so that the number of inferred jump times divided by the number of observations is approximately equal to the estimate of λ.

3.1.1 Data

We estimate the BTC returns by taking the log first differences of prices, then use returns to estimate the SVCJ model. The analyses are carried out based on daily closing prices. The data cover the period from October 2017 to 28 February 2021 and are collected from coingecko.com. The dynamics of BTC daily prices and BTC returns are depicted in Figure 3.1. It shows that the BTC return is clearly more volatile than the traded stock return, along with more frequent jumps or the scattered volatility spikes. This makes the standard set of stationary models, such as ARIMA and GARCH, unable to fit the BTC returns well due to the presence of jumps, as specified by Chen et al. (2020), [31].

Several large jumps triggered by a series of big events in the BTC market can be detected from the returns series, see also Kim et al. (2019), [34]. The first peak of Bitcoin price was registered at the end of 2017, due to the widespread interest in cryptocurrencies. After that, the price moved downward until July 2019 when a new peak was reached after Facebook announcement on the Libra project. However, the unequivocal proof that the BTC (and cryptocurrency in general) market is strictly linked to the news and has a speculative nature came on March 11th 2020, when

the BTC price crashed down by 65% after the World Health Organisation (WHO) officially characterized the Covid-19 as a pandemic.

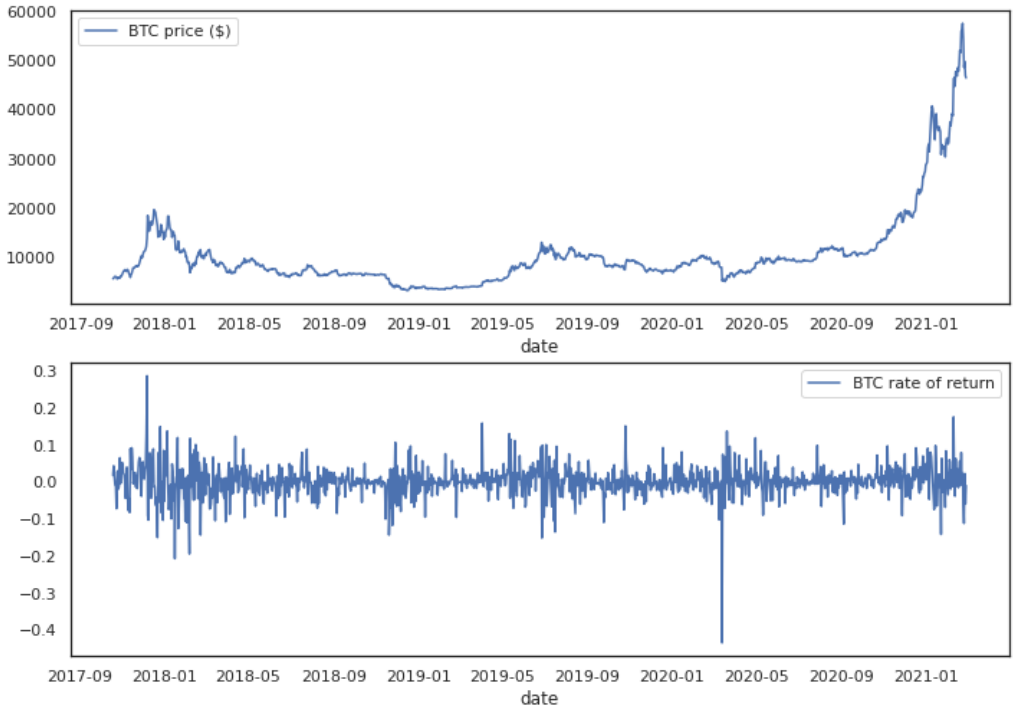

Figure 3.1: Historical BTC/USD prices and associated log-returns

The large upward and downward movements in BTC prices make the returns of BTC display high volatility and scattered spikes/jumps.

We can compare BTC with other widespread cryptocurrencies such as Ethereum (ETH) and Cardano (ADA), having the second and third highest market capitalization (with dominance rate equal to 11% for ETH and 2,5% for ADA). Figure 3.2 and Figure 3.3 show the historical price and return of ETH and ADA, respectively. We observe that the price patterns of ETH and ADA reflect the dynamics depicted in Figure 3.1: an almost zero price variation from the cryptocurrency release (not included in the plot) until a first notable increase at the end of 2017, a drop on March 11th 2020 and then a new incredible growth. The maximum price for Bitcoin was reached on 22.02.2021 (57.669,30 USD), for Ethereum on 20.02.2021 (2.042,93 USD), for Cardano on 27.02.2020 (1,48 USD).

As in the case of BTC, returns oscillate around zero with frequent changes, positive and negative. Those changes give the idea of a jump. Analyzing the three return plot together, it emerges that the dynamics of returns is not related with market capitalization. Bitcoin, Ethereum and Cardano substantially share the same returns pattern.

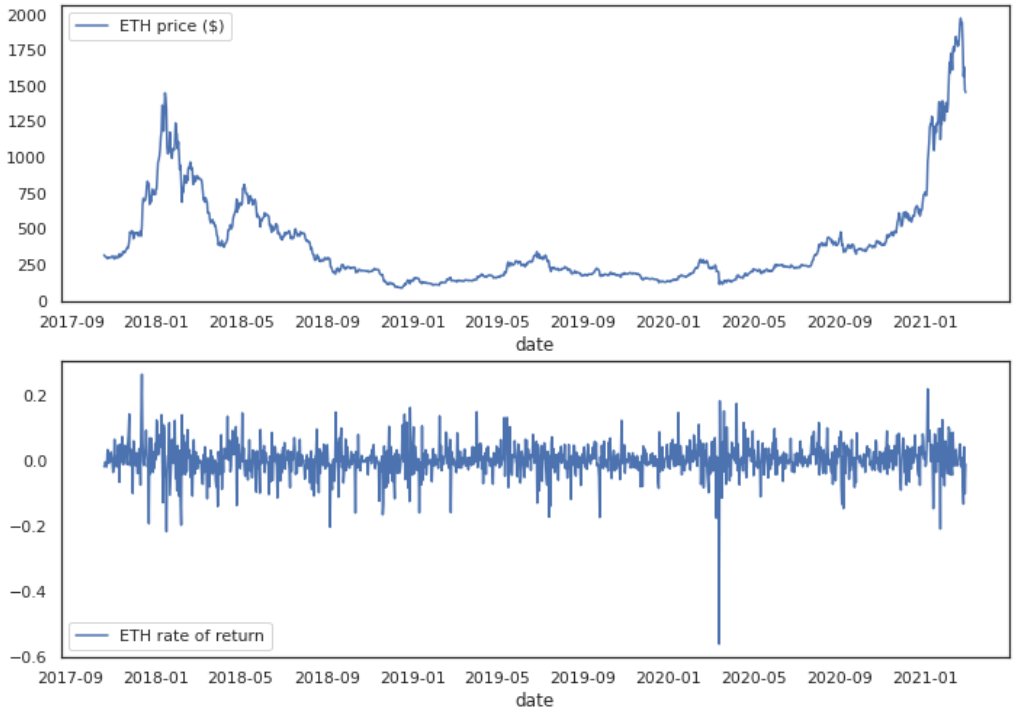

Figure 3.2: Historical ETH/USD prices and associated log-returns

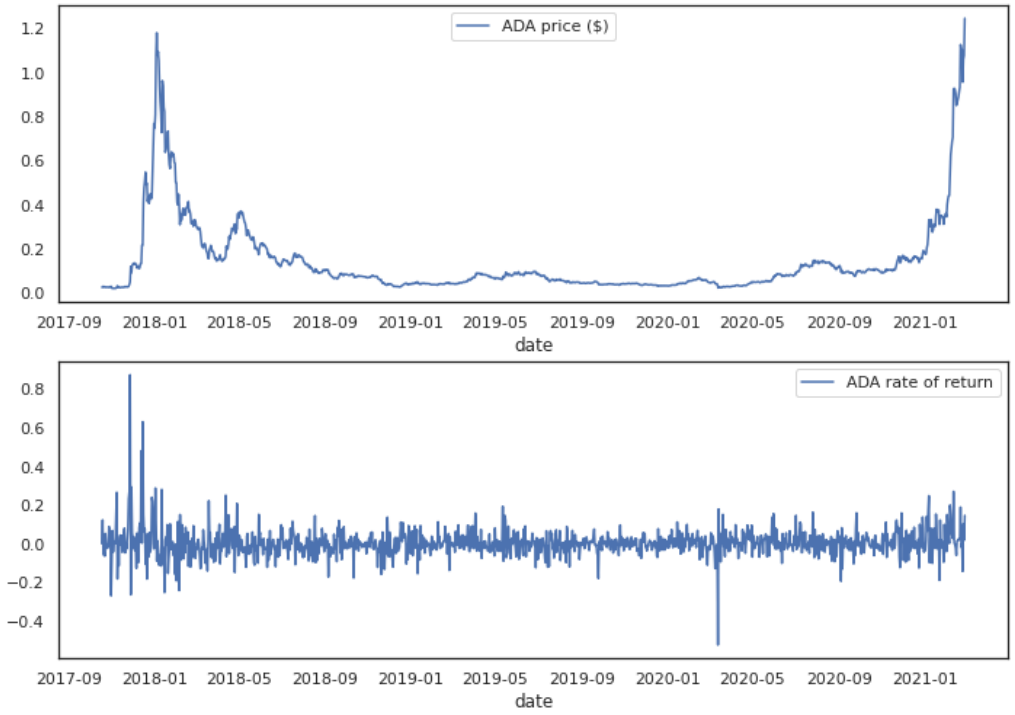

Figure 3.3: Historical ETH/USD prices and associated log-returns

It appears then clear that there exists at least an empirical correlation between log-returns of the aforementioned cryptocurrencies. To give a deeper insight of this property we include the pairplot of cryptocurrencies log-returns, Figure 3.4.

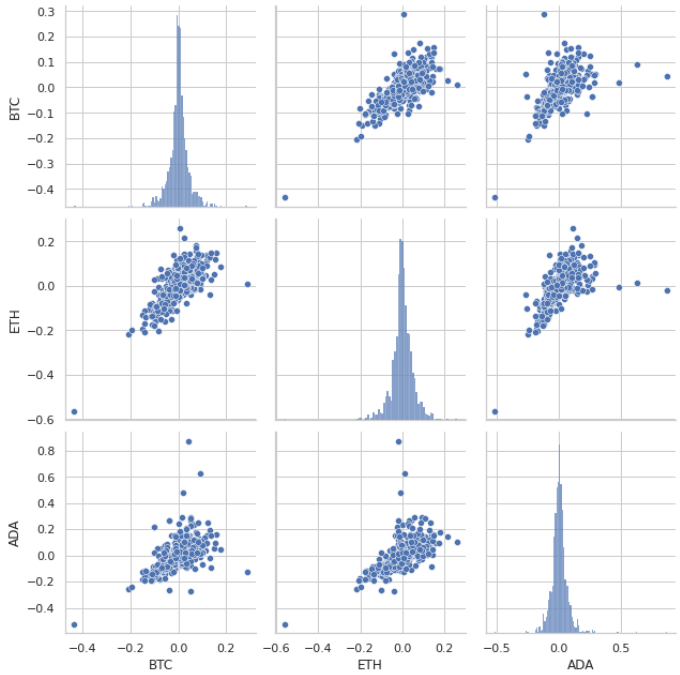

Figure 3.4: Pairplot of criptocurrencies log-returns

Focusing on BTC, the returns distribution is highlighted in Figure 3.5.

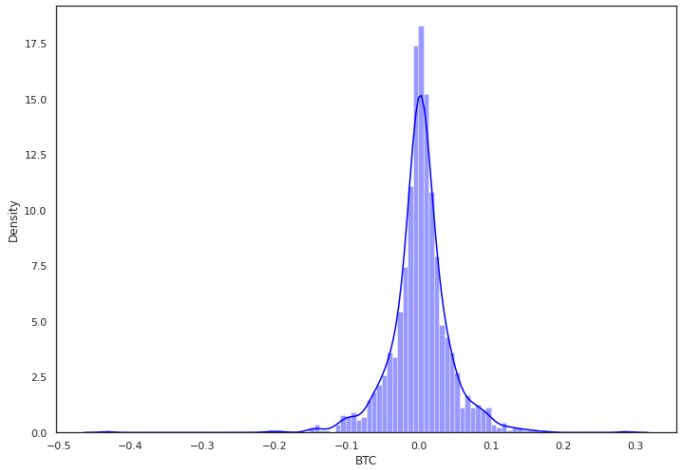

Figure 3.5: Histogram of BTC log-returns

3.1.2 Parameter Estimation

We use the returns to calculate the SVCJ model. The main code for the calculation of the SVCJ reproduces the one used by Chen et al., [31], in their estimations. We perform the SVCJ estimation considering BTC historical data from February 2017 to February 2021, a total of 5000 iterations were done in each case with a burn-in of 1000 to minimize initial value influence.

In the following we present the trace plot for the estimated parameters and their moving average.

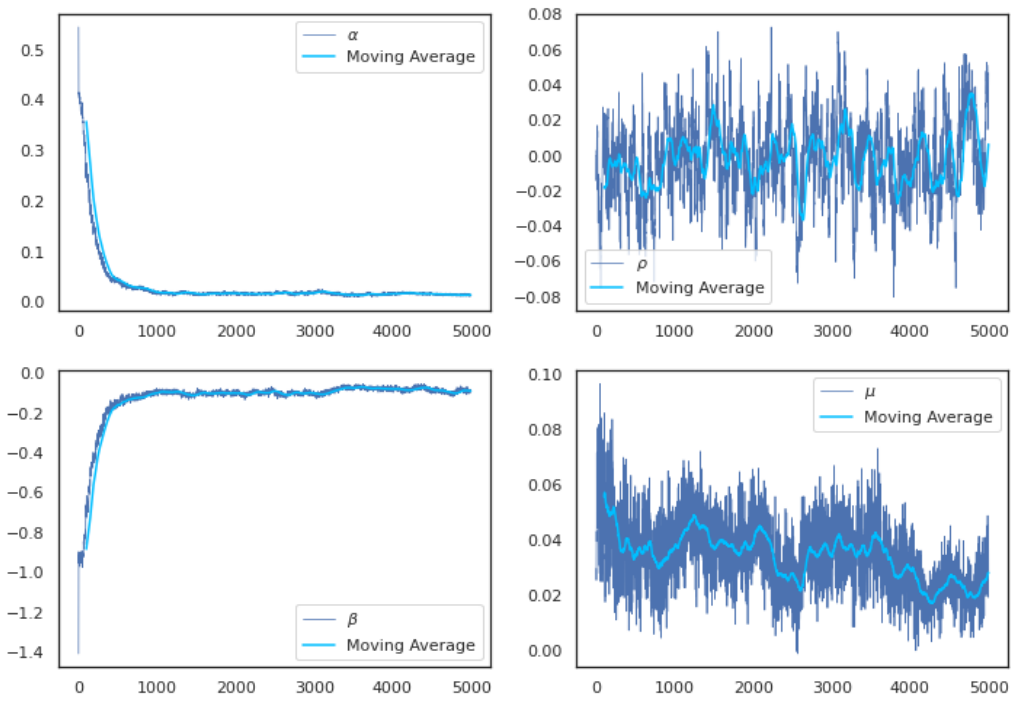

Figure 3.6: Trace plot SVCJ parameters: α, β, ρ, μ

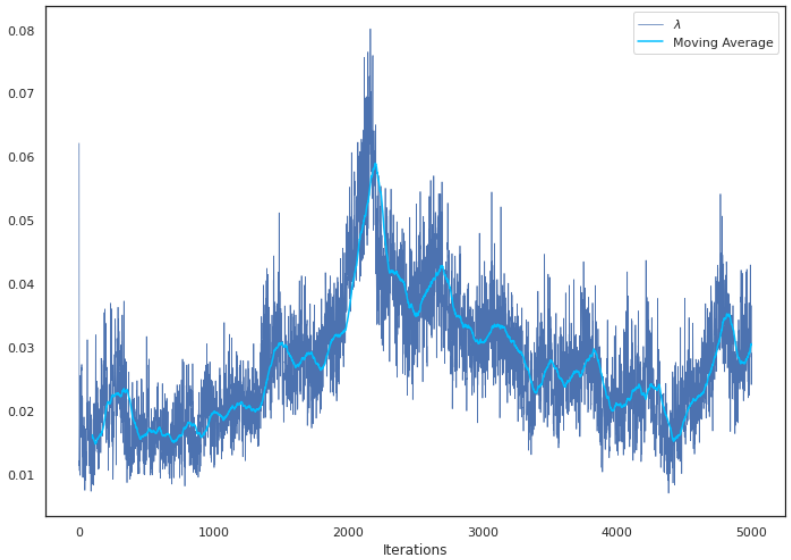

Figure 3.7: Trace plot SVCJ parameters: λ

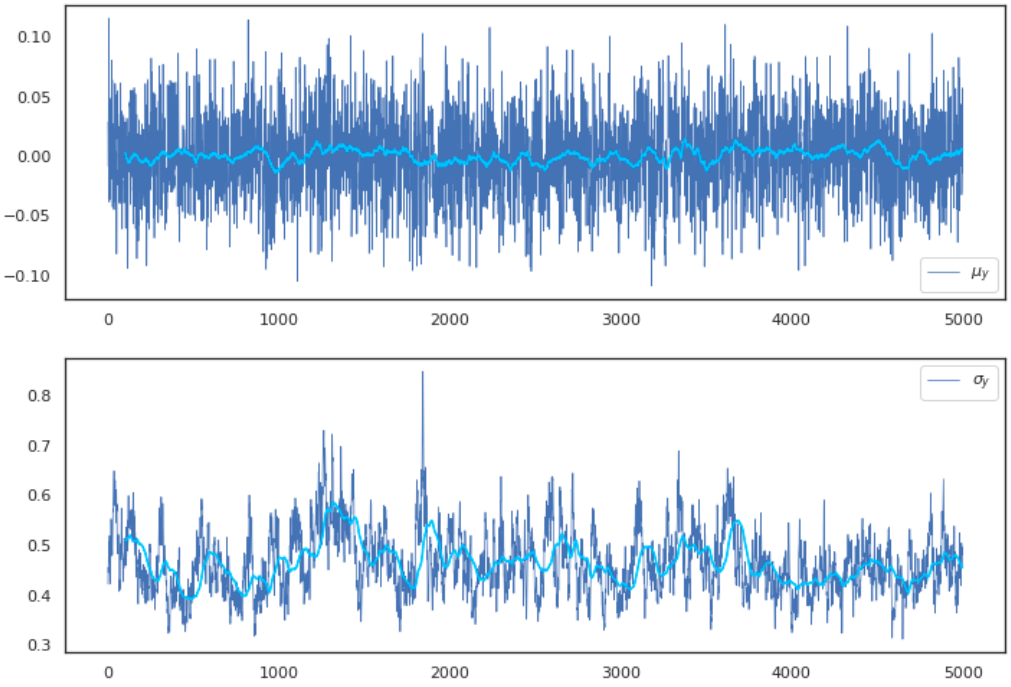

Figure 3.8: Trace plot SVCJ parameters: μ_y, σ_y

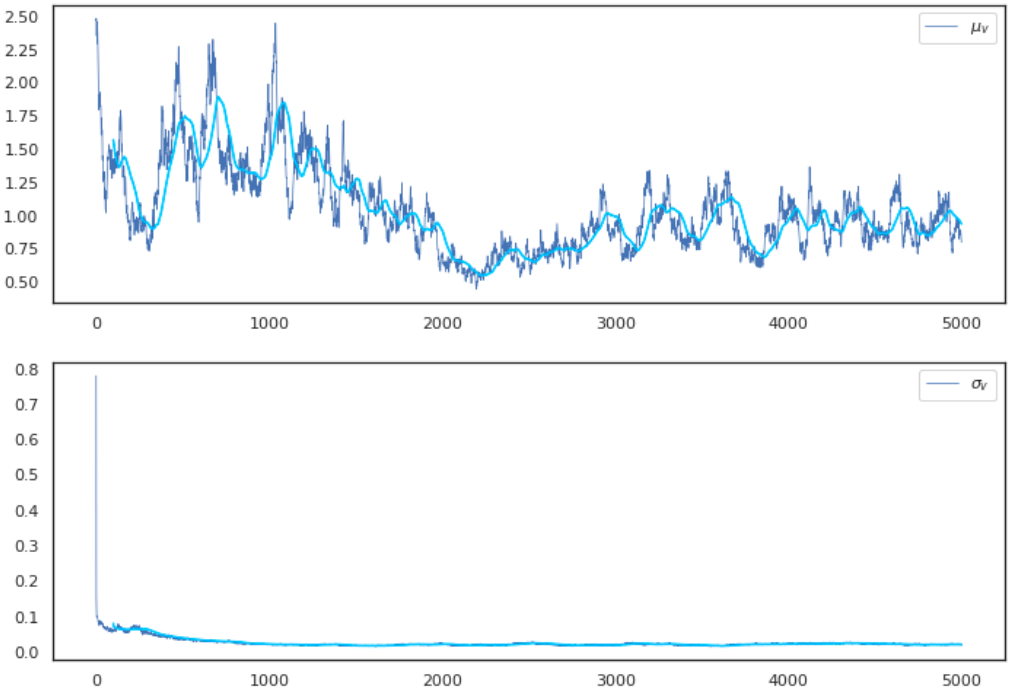

Figure 3.9: Trace plot SVCJ parameters: μ_v, σ_v

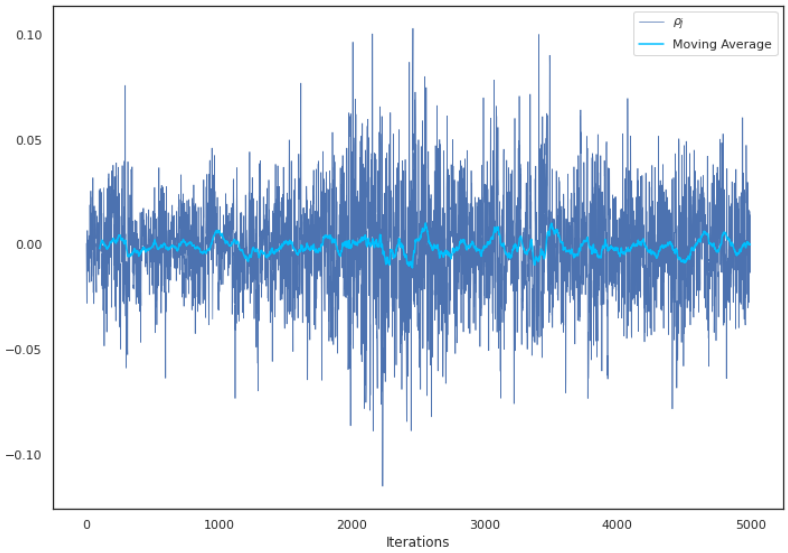

Figure 3.10: Trace plot SVCJ parameters: ρ_j

We now inspect the dynamics followed by the mean of each parameter estimated considering different amounts of data. In particular, we have adopted the two following approaches:

1. For each day t_i between $t_0 = 28.02.2018$ and $t_N = 28.02.2021$, let I_i be the discrete time interval from $t_0 = 28.02.2018$ to t_i, included. For each $i \in \{0, 1, \ldots, N\}$, we calibrate the SVCJ model giving as an input the historical daily price and returns registered for each day in I_i and performing, again, 5000 iterations. The first group of plots (i.e., Figures 3.11, 3.12, 3.13, 3.14, 3.15) shows, for each parameter, for each day t_i, the mean value calculated from the iterations generated including as an input the historical data from 28.02.2018 to t_i.

2. For each day t_i between $t_0 = 28.02.2018$ and $t_N = 28.02.2021$ we define J_i as the discrete time interval obtained considering the 365 days preceding t_i, i.e., a horizon of one year ending with t_i. For each $i \in \{0, 1, \ldots, N\}$, we calibrate the SVCJ giving as an input the historical daily price and returns registered for each day in J_i and performing 5000 iterations. The second group of plots (i.e., Figures 3.16, 3.17, 3.18, 3.19, 3.20) shows, for each parameter, for each day t_i, the mean value calculated from all the iterations generated considering the historical data for each day in J_i.

Analysing the results and the plots obtained adopting the first approach, we notice that some parameters display a sort of empirical convergence or are pretty stable and close to zero (e.g. μ_y, ρ_j), while λ, σ_y and μ_v oscillate in a small interval.

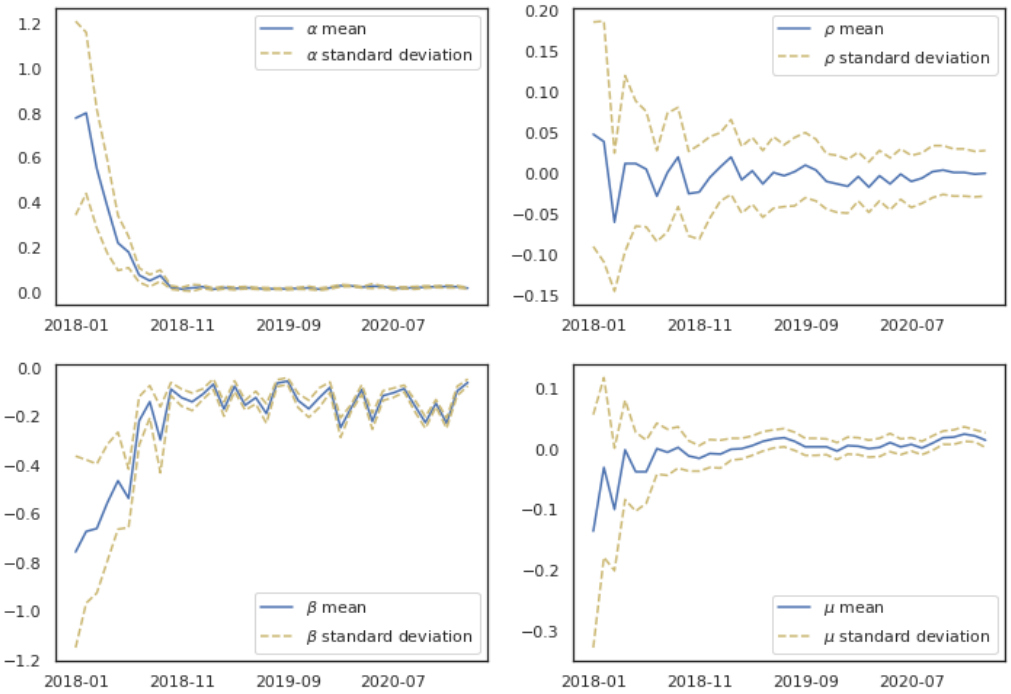

Figure 3.11: Approach 1: calibration of α, β, ρ, μ

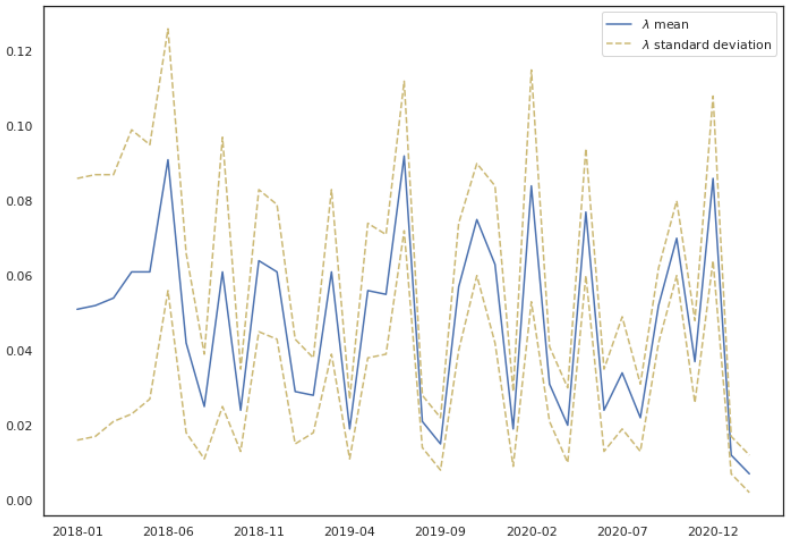

Figure 3.12: Approach 1: calibration of λ

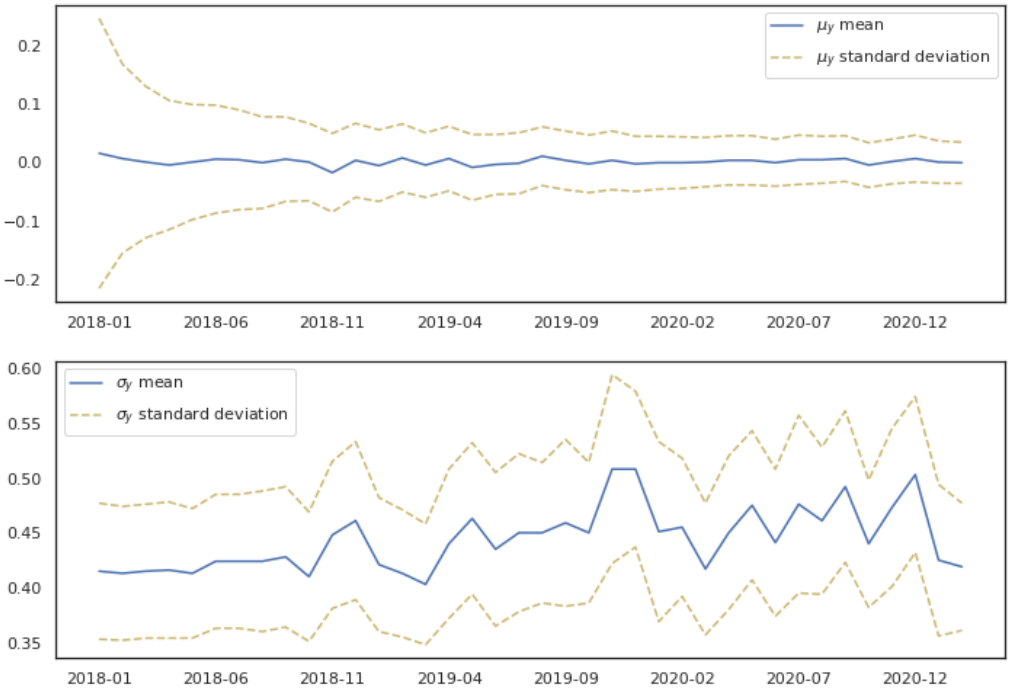

Figure 3.13: Approach 1: calibration of μ_y, σ_y

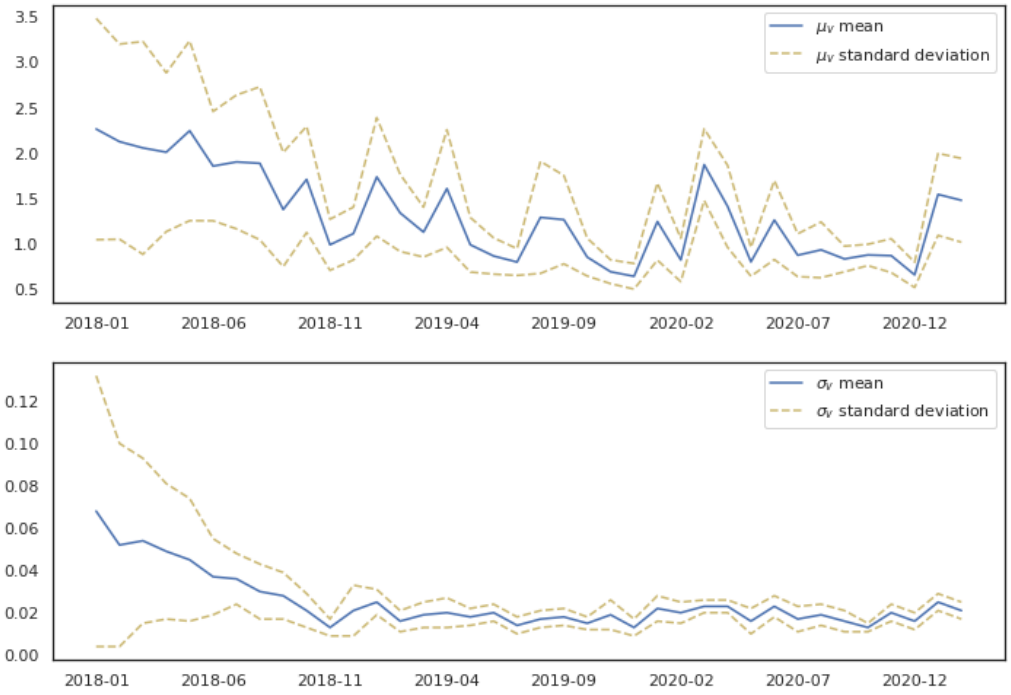

Figure 3.14: Approach 1: calibration of μ_v, σ_v

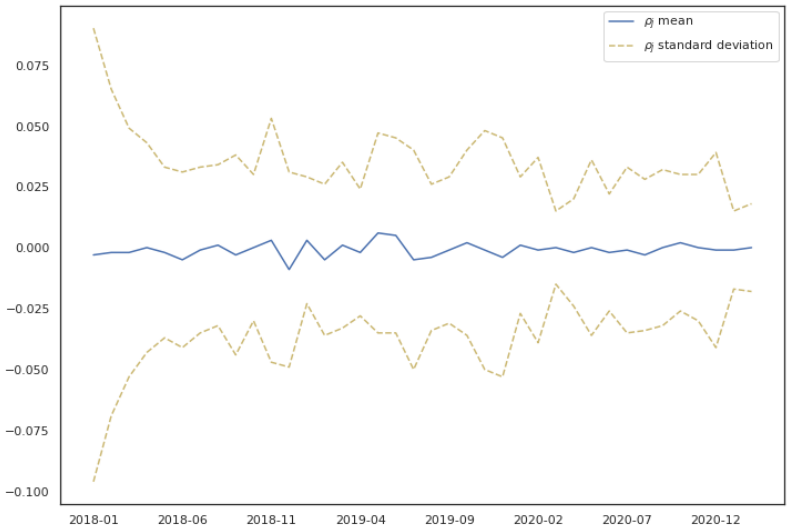

Figure 3.15: Approach 1: calibration of ρ_j

We can conclude that the first approach leads to an overall empirically acceptable parameter calibration when the time interval is sufficiently wide. The analyses carried adopting the second approach suggest that a good fitting for the SVCJ model can be reached considering a horizon of one year for BTC historical data used in the calibration method.

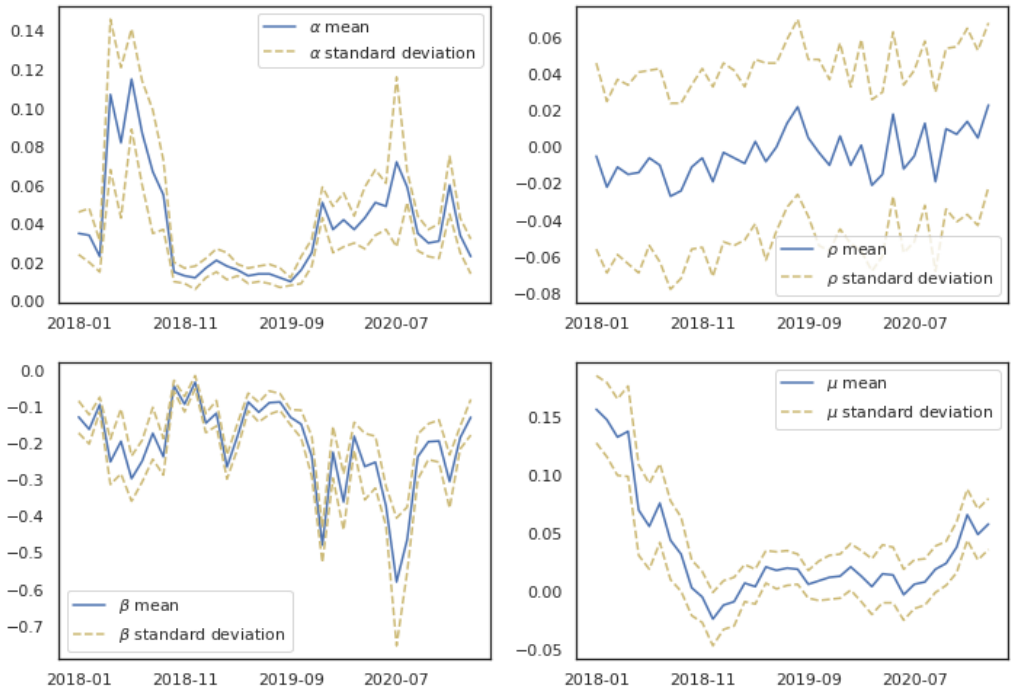

Figure 3.16: Approach 2: calibration of α, β, ρ, μ

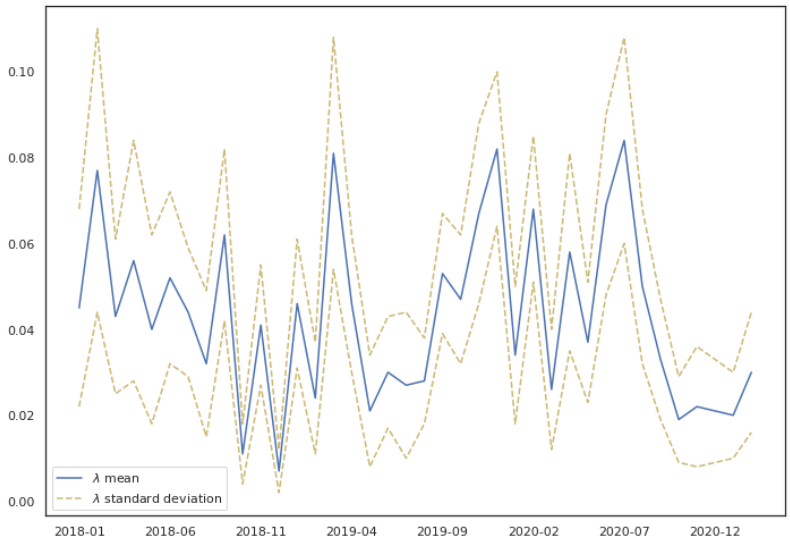

Figure 3.17: Approach 2: calibration of λ

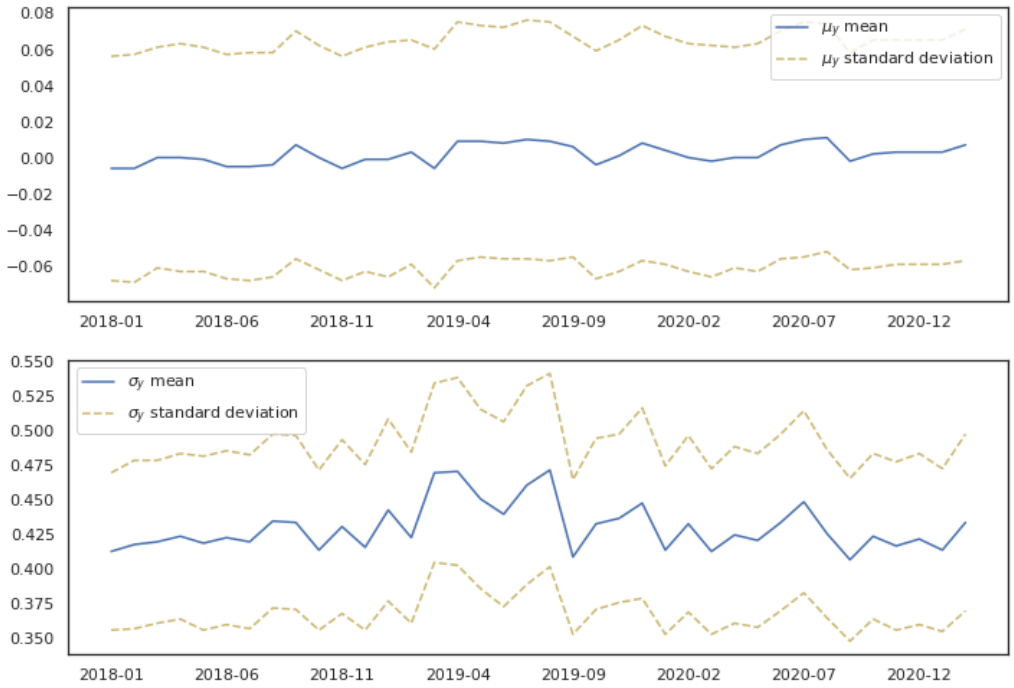

Figure 3.18: Approach 2: calibration of μ_y, σ_y

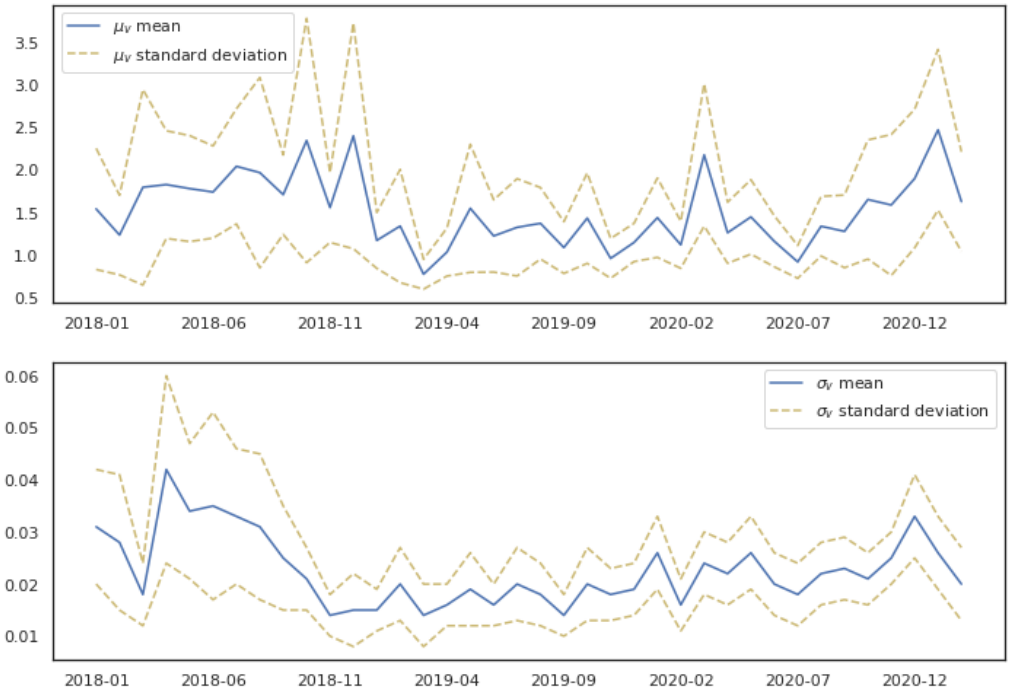

Figure 3.19: Approach 2: calibration of μ_v, σ_v

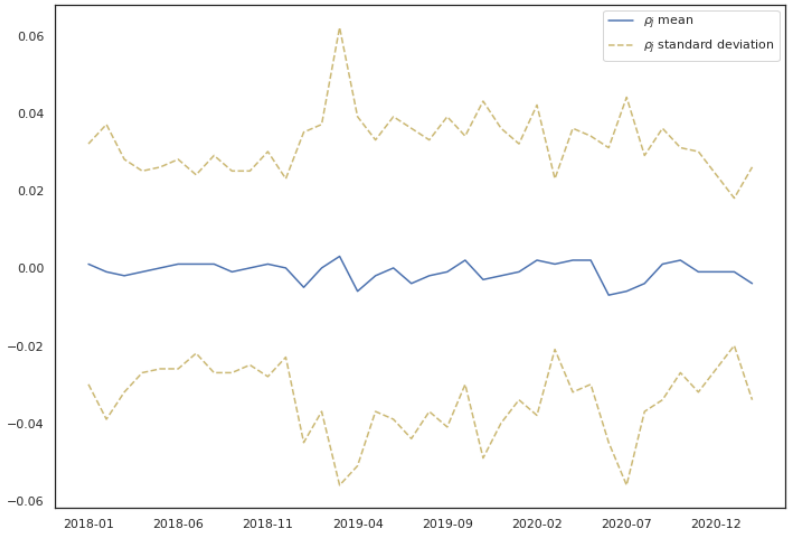

Figure 3.20: Approach 2: calibration of ρ_j

The SVCJ model estimates also jumps in returns and volatility, presented in Figure 3.21.

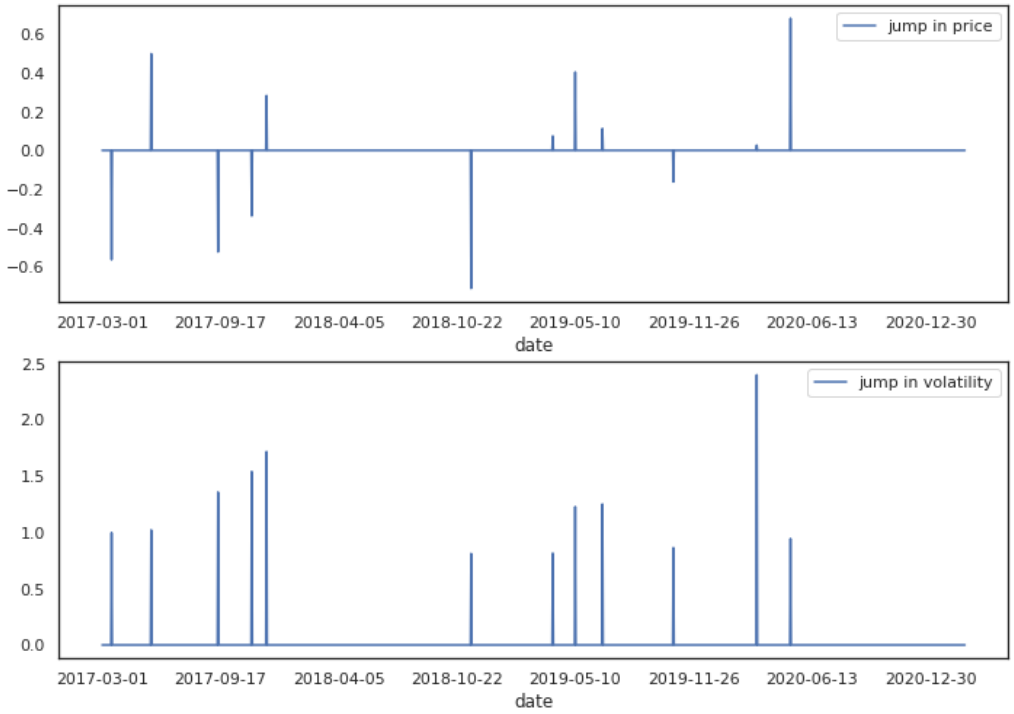

Figure 3.21: SVCJ estimated jumps in price and volatility

The estimated volatility under the SVCJ model is shown in Figure 3.22. It is worth noticing the huge increase in volatility in March 2020 that reflects the plunge of BTC prices and jumps in returns caused by the decision and affirmations of the WHO in relation to Covid-19 pandemic.

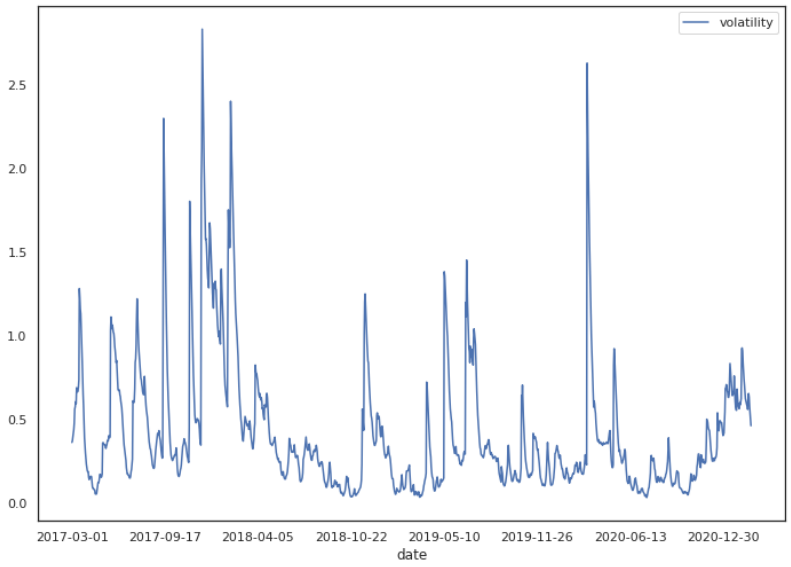

Figure 3.22: SVCJ estimated volatility

Considering the SVCJ residuals, we can refer to the QQ-plot depicted in Figure 3.23. The SVCJ model residuals seems to follow a normal distribution which speaks about a good model fitting.

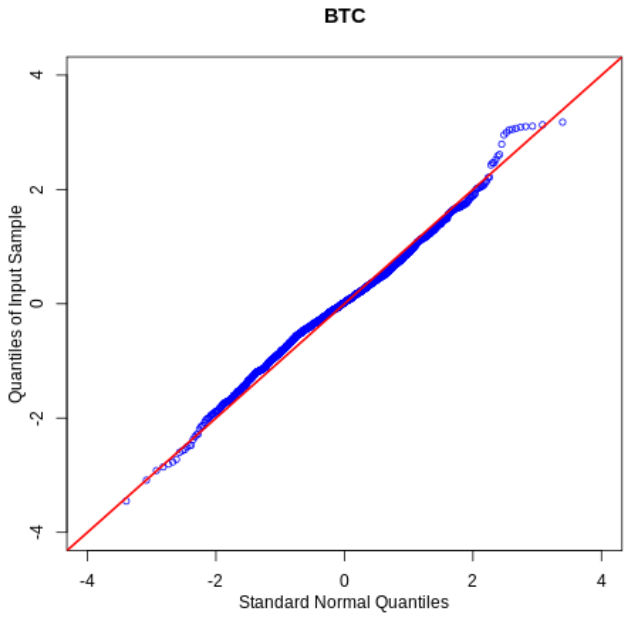

Figure 3.23: SVCJ residuals QQ-plot

Once the SVCJ model is calibrated, it is possible to simulate BTC prices path. In Figure 3.24 we show five different possible price dynamics, obtained fixing the initial price at 49.000,00 USD.

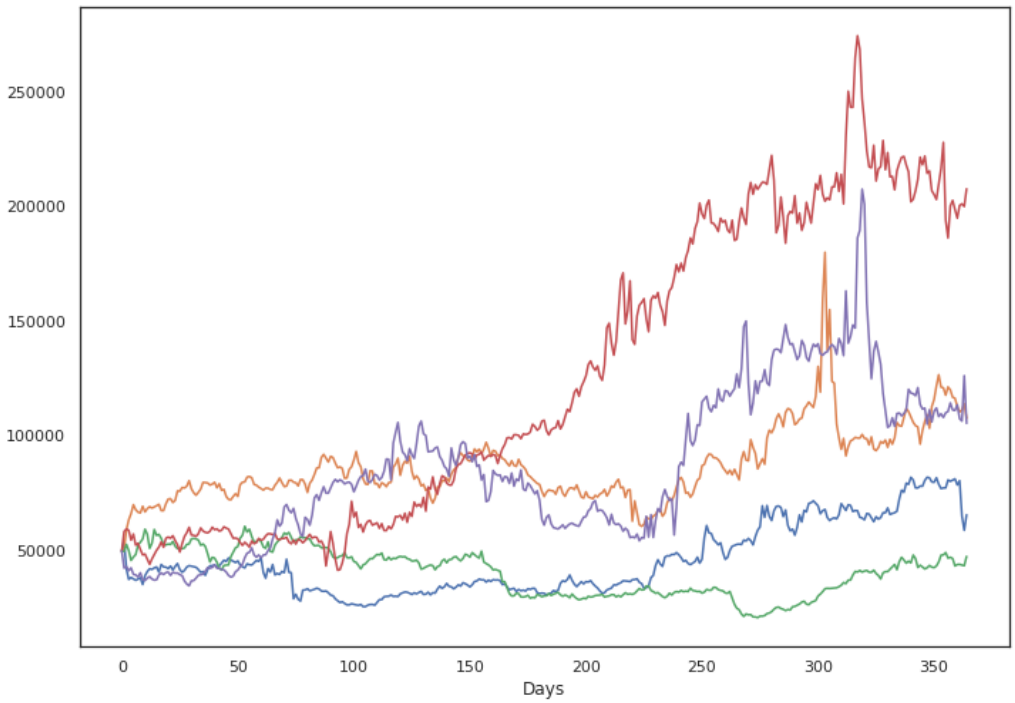

Figure 3.24: BTC/USD price simulated path

Conclusion

Cryptocurrencies cannot be considered neither a reliable store of value nor a measure of value. The exchange rate's high volatility and the price of cryptocurrencies are influenced by the supply and demand ratio, market trustworthiness and news.

The sensitivity of investors and currency holders to market news pairs with the classical market factors (e.g., political instability, investment climate and regulatory framework, internal factors associated with changes in technologies and various processes in the system) and translates into an extremely risky asset both in the payment systems framework and in speculation-driven investments. The cryptocurrency market is relatively young, therefore standard market metrics cannot be applied in general and, due to the strong volatility of most cryptocurrencies, they are often used for speculative purposes. This speculation-driven approach to the market does not contribute to digital currencies development and formation as participants in the payment system. Such capital is considered to be fictional capital, which does not contribute to the development of the economy, but only participates in the redistribution of capital.

Bitcoin is the first officially traded cryptocurrency and, despite its riskiness, it benefits of the highest level of confidence among other digital currencies, in fact, when prices fall against the news about Covid-19 pandemic, the society of traders and investors begins to buy the asset at a more attractive price. Actually, Bitcoin and Ethereum occupy market's leading position in terms of capitalization and volume of transactions, sharing a "long" history and investors' trust.

This market is subject to a constant evolution: new cryptocurrency systems appear everyday and the rising competition contributes to the active market development. Potential investors are attracted by the incredible volatility of these assets and cryptocurrencies are becoming a tool for speculative transactions and capital redistribution. We should then expect further developments in the cryptocurrency market in terms of participation of additional agents, such as central and commercial banks, and institution of a derivatives market.

In this context, the understanding of cryptocurrencies price behavior plays a fundamental role. Here we have described the stochastic volatility correlated jump (SVCJ) model of Duffie Pan and Singleton (2000), [1], and verified that it fits sufficiently well the BTC data. The several analyses performed demonstrate that SVCJ model can be reasonably used for cryptocurrencies price path simulations.

Further developments of this work can include the study of cryptocurrency option pricing framework and the realisation of pricing tools for vanilla options on cryptocurrencies.

References

[1] D. Duffie, J. Pan, and K. Singleton, "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, vol. 68, pp. 1343–1376, 02 2000.

[2] T. Doan, E. Jacquier, N. Polson, and P. Rossi, "Bayesian analysis of stochastic volatility models," *Journal of Business Economic Statistics*, vol. 12, pp. 371–89, 02 1994.

[3] M. Scholes and F. Black, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, pp. 637–54, 02 1973.

[4] S. Heston, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of Financial Studies*, vol. 6, pp. 327–43, 02 1993.

[5] R. Cont and P. Tankov, *Financial Modelling With Jump Processes*, vol. 101. 01 2006.

[6] A. Pascucci, *PDE and Martingale Methods in Option Pricing*. 01 2011.

[7] D. Applebaum, *Lévy Processes and Stochastic Calculus*. 07 2004.

[8] A. Pascucci, *PDE and Martingale Methods in Option Pricing*. Milan, Italy: Springer-Verlag, 2010.

[9] R. Merton, "Option prices when underlying stock returns are discontinuous," *Journal of Financial Economics*, vol. 3, pp. 125–144, 01 1976.

[10] S. Kou, "A jump-diffusion model for option pricing," *Management Science*, vol. 48, pp. 1086–1101, 08 2001.

[11] D. Bates, "Jumps and stochastic volatility: Exchange rate process implicit in deutsche mark options," *Review of Financial Studies*, vol. 9, pp. 69–107, 02 1996.

[12] C. Cao, G. Bakshi, and Z. Chen, "Empirical performance of alternative option pricing models," *Journal of Finance*, vol. 52, 05 1997.

[13] D. Bates, "Post-'87 crash fears in sp 500 futures options," *National Bureau of Economic Research, Inc, NBER Working Papers*, vol. 94, 01 1997.

[14] J. Pan, "The jump-risk premia implicit in options: Evidence from an integrated time-series study," *Journal of Financial Economics*, vol. 63, pp. 3–50, 01 2002.

[15] C. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 2004.

[16] M. Johannes and N. Polson, "Mcmc methods for continuous-time financial econometrics," *Handbook of Financial Econometrics, Vol 2*, vol. 2, 12 2003.

[17] B. Eraker, M. Johannes, and N. Polson, "The impact of jumps in volatility and returns," *Journal of Finance*, vol. 58, pp. 1269–1300, 02 2003.

[18] B. Eraker, "Do stock prices and volatility jump? reconciling evidence from spot and option prices," *The Journal of Finance*, vol. 59, pp. 1367 – 1404, 06 2004.

[19] J. Besag, "Spatial interaction and statistical analysis of lattice systems," *J R Stat Soc B*, vol. 48, 01 1974.

[20] Metropolis, Nicholas, A. W., Rosenbluth, M. N., A. H., Teller, and Edward, "Equation of state calculations for fast computing machines," *Journal of Chemical Physics 6*, vol. 21, pp. 1087–, 01 1953.

[21] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*. 01 2009.

[22] L. Tierney, "Markov chains for exploring posterior distributions," *The Annals of Statistics*, vol. 22, 12 1994.

[23] G. Roberts and N. Polson, "On the geometric convergence of the gibbs sampler," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 56, pp. 377–384, 07 1994.

[24] K. Mengersen and R. Tweedie, "Rates of convergence of the hastings and metropolis algorithms," *The Annals of Statistics*, vol. 24, 02 1996.

[25] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 03 2009.

[26] W. Hrdle, C. Harvey, and R. Reule, "Understanding cryptocurrencies*," *Journal of Financial Econometrics*, vol. 18, pp. 181–208, 03 2020.

[27] J. Lansky, "Possible state approaches to cryptocurrencies," *Journal of Systems Integration*, vol. 8, 01 2018.

[28] O. Scaillet, A. Treccani, and C. Trevisan, "High-frequency jump analysis of the bitcoin market*," *Journal of Financial Econometrics*, 06 2018.

[29] J. E.-T. Cheah and J. Fry, "Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin," *Economics Letters*, vol. 130, 02 2015.

[30] L. Kristoufek, "What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis," *PLoS ONE*, 06 2014.

[31] W. Wang, A. Hou, W. Härdle, and C. Chen, "Pricing cryptocurrency options," *Journal of Financial Econometrics*, 03 2020.

[32] H. Asgharian and M. Nossman, "Risk contagion among international stock markets'," *Journal of International Money and Finance*, vol. 30, pp. 22–38, 02 2008.

[33] M. Johannes, R. Kumar, and N. Polson, "State dependent jump models: How do us equity markets jump," 03 1999.

[34] A. Kim, S. Trimborn, and W. Härdle, "Vcrix - a volatility index for cryptocurrencies," *SSRN Electronic Journal*, 01 2019.