APPLIED MATHEMATICS THESIS

UNIVERSITÀ DEGLI STUDI

DI PAROVA

Ph.D THESIS

Scaling and Multiscaling in Financial Indexes: A Simple Model.

Ph.D Candidate: Alessandro Andreoli

Relators: Ch.mo Prof. Paolo Dai Pra
Ch.mo Prof. Francesco Caravenna

Anno Accademico 2010-2011

Alla mia famiglia e
ai miei nonni

Contents

Sommario

In questa tesi di Dottorato proponiamo un semplice modello stocastico per serie storiche, che presenta le seguenti caratteristiche: è analiticamente trattabile, facile da simulare e cattura alcuni importanti fatti stilizzati degli indici finanziari, fra cui proprietà di scaling. Inoltre, mostriamo che il modello ha un ottimo fit con la serie storica del *Dow Jones Industrial Average* fra il 1935-2009, e infine deriviamo alcuni risultati di option pricing per tale modello.

Recenti sviluppi nella modellazione stocastica di serie storiche sono state fortemente influenzate dall'analisi di indici finanziari. Il modello base, da cui si ricava la celebre formula di & Scholes [17, 24], assume che il logaritmo X_t del prezzo dell'indice sottostante, una volta detrendizzato, sia dato da

$$\mathrm{d}X_t = \sigma \, \mathrm{d}W_t,$$

dove σ (la *volatilità*) è costante e $(W_t)_{t\geq 0}$ è un moto Browniano standard. È ben noto che, nonostante il suo successo, questo modello non è consistente con molti *fatti stilizzati* che sono stati empiricamente verificati su molte serie storiche. Alcuni fra i più importanti di questi fatti sono i seguenti:

- la volatilità non è costante: in particolare, può esibire picchi molto alti, spesso interpretati come *shocks* nel mercato;
- la distribuzione empirica degli incrementi $X_{t+h} - X_t$ dei logaritmi dei prezzi (detti *log-returns*), esibisce code più pesanti di una Gaussiana;
- log-returns corrispondenti a intervalli disgiunti di tempo sono effettivamente scorrelati, ma non indipendenti: la correlazione fra i valori assoluti $|X_{t+h} - X_t|$ e $|X_{s+h} - X_s|$ presenta infatti un decadimento piuttosto lento in $|t - s|$, almeno per valori moderati dello stesso. Questo fenomeno è noto come *clustering of volatility*.

Al fine di ottenere un fit migliore con i dati reali, sono stati proposti molti differenti modelli per descrivere i processi della volatilità e del prezzo. A tempo discreto, modelli autoregressivi quali ARCH, GARCH e generalizzazioni [15, 10, 3, 11] sono ampiamente utilizzati, mentre a tempo continuo hanno ottenuto molto successo modelli con volatilità stocastica [6].

Più recentemente (see [14, 9, 26]), sono stati evidenziati altri fatti stilizzati riguardanti le *proprietà di scaling* della distribuzione empirica dei log-returns. Consideriamo la serie storica di un indice $(s_i)_{1\leq i\leq T}$ in un periodo $T \gg 1$ giorni e denotiamo con p_h the *distribuzione*

empirica dei log-returns (detrendizzati) corrispondenti all'intervallo di h giorni:

$$p_h(\cdot) := \frac{1}{T-h}\sum_{i=1}^{T-h}\delta_{x_{i+h}-x_i}(\cdot)\,, \qquad x_i := \log(s_i) - \overline{d}_i\,,$$

dove $\overline{d}_i$ è il tasso lineare locale di crescita di $\log(s_i)$. L'analisi statistica di vari indici mostra che, per h entro un'opportuna scala di tempo, p_h segue approsimativamente una relazione di scaling diffusivo del tipo

$$p_h(\mathrm{d}r) \simeq \frac{1}{\sqrt{h}}\, g\left(\frac{r}{\sqrt{h}}\right)\mathrm{d}r,$$

dove g è una densità con code più pesanti di una Gaussiana.

Inoltre, se si considera il q-esimo momento empirico $m_q(h)$, definito da

$$m_q(h) \;:=\; \frac{1}{T-h}\sum_{i=1}^{T-h}|x_{i+h}-x_i|^q \;=\; \int |r|^q\, p_h(\mathrm{d}r)\,,$$

si osserva che per momenti di ordine piccolo $q \le \bar{q}$, $m_q(h)$ scala come $h^{\frac{q}{2}}$, mentre, per momenti di ordine maggiore $q > \bar{q}$, esso scala in modo differente $m_q(h) \approx h^{A(q)}$, con $A(q)$ strettamente minore di $q/2$; questo fenomeno è noto come *multiscaling dei momenti.*

Un'analisi dettagliata di tali caretteristiche statistiche viene trattato nel Capitolo 1.

Nel Capitolo 2, partendo da quanto sottolineato nel primo capitolo e rendendo spunto dal lavoro di Baldovin e Stella [26], definiamo un *semplice* modello stocastico a tempo continuo compatibile con *tutti* i fatti stilizzati citati; questo punto è non triviale, infatti, nonostante la varietà di modelli che si possono trovare in letteratura, non siamo a conoscenza di uno che soddisfi tutte queste richieste.

Il meccaniscmo alla base del nostro modello è un *random update della volatilità.*

Dati due numeri reali $D \in (0, 1/2]$, $\lambda \in (0,\infty)$ e una probabilità ν on $(0,\infty)$ (che si possono pensare come i parametri generali del modello), il nostro modello si basa sulle tre seguenti fonti di aleatorietà:

- un moto Browniano standard $W = (W_t)_{t\ge 0}$;
- un processo di punto di Poisson $\mathcal{T} = (\tau_n)_{n\in\mathbb{Z}}$ su $\mathbb{R}$ con intensità λ (per convenzione, chiameremo i punti di $\mathcal{T}$ in modo tale che $\tau_0 < 0 < \tau_1$);
- una successione $\Sigma = (\sigma_n)_{n\ge 0}$ di variabili aleatorie positive e indipendenti identicamente distribuite.

Assumiamo inoltre che $W, \mathcal{T}, \Sigma$ siano indipendenti.

Per $t \ge 0$, definiamo poi

$$i(t) \;:=\; \sup\{n \ge 0 :\, \tau_n \le t\} \;=\; \#\{\mathcal{T} \cap (0,t]\}$$

(quindi $\tau_{i(t)}$ è l'ultimo punto di $\mathcal{T}$ prima di t), e introduciamo il processo $I = (I_t)_{t\ge 0}$

$$I_t \;:=\; \sigma^2_{i(t)}\theta\left(\lambda(t-\tau_{i(t)})\right) \;+\; \sum_{k=1}^{i(t)}\sigma^2_{k-1}\theta\left(\lambda(\tau_k-\tau_{k-1})\right) \;-\; \sigma^2_0\theta\left(-\lambda\tau_0\right),$$

dove $\theta : [0, +\infty) \to [0, +\infty)$ è una funzione di riscalamento temporale che soddisfa le seguenti proprietà:

(1) θ è concava, strettamente crescente, $\theta(0) = 0$ e $\theta(s) \to +\infty$ per $s \to +\infty$.

(2) θ è $\mathcal{C}^1$ su $(0, +\infty)$, e esiste una costante $0 < D < \frac{1}{2}$ tale che

$$\lim_{s \downarrow 0} \frac{\theta'(s)}{2Ds^{2D-1}} = 1.$$

In particolare quindi, $\theta(s)$ si comporta come s^{2D} vicino a 0, e questo andamento è responsabile delle singularità di I_t nei punti τ_i.

Il nostro processo $X = (X_t)_{t \geq 0}$ si può quindi definire ponendo

$$X_t := W_{I_t} \,,$$

e quindi può essere visto come un independent random time change di un moto Browniano.

Inoltre, sempre nel Capitolo 2 dimostriamo alcune proprietà basilari del processo, e in particolare mostriamo che:

(A) *X ha incrementi stazionari.*

(B) *X può essere rappresentato con un processo a volatilità stocastica*:

$$\mathrm{d}X_t = v_t \, \mathrm{d}B_t \,,$$

dove $(B_t)_{t \geq 0}$ è un moto Browniano Standard. Più precisamente, denotando $I'(s) := \frac{\mathrm{d}}{\mathrm{d}s} I(s)$, si ha che B_t e v_t sono definite da

$$B_t := \int_0^{I_t} \frac{1}{\sqrt{I'(I^{-1}(u))}} \, \mathrm{d}W_u \,, \qquad v_t := \sqrt{I'(t)} = \sigma_{i(t)} \sqrt{\lambda \, \sigma' \left(\lambda (t - \tau_{i(t)}) \right)} \,.$$

(C) *X è una martingala con media zero e di quadrato integrabile* (assunto $\mathbb{E}[\sigma^2] < \infty$).

Nel Capitolo 3, vengono provati i principali risultati (asintotici) che abbiamo ottenuto per il processo X. Essi riguardano alcuni dei fatti stilizzati presentati nel Capitolo 1, in particolare: lo scaling diffusivo della distribuzione dei log-returns, il multiscaling dei momenti e il clustering of volatility.

Teorema. *Per $h \downarrow 0$ abbiamo la seguente convergenza in distribuzione*

$$\frac{(X_{t+h} - X_t)}{\sqrt{h}} \xrightarrow[h \downarrow 0]{d} f(x) \, \mathrm{d}x \,,$$

dove f è una mistura di densità Gaussiane, precisamente

$$f(x) = \int_0^\infty \nu(\mathrm{d}\sigma) \int_0^\infty \mathrm{d}t \, \lambda e^{-\lambda t} \, \frac{1}{\sigma \sqrt{2\pi \theta'(\lambda t)}} \, e^{-\frac{x^2}{2\sigma^2 \theta'(\lambda t)}} \,.$$

Theorem (Multiscaling dei momenti). *Sia $q > 0$, e si assuma $\mathbb{E}[\sigma^q] < +\infty$. Allora la quantità $m_q(h) := \mathbb{E}[|X_{t+h} - X_t|^q] = \mathbb{E}[|X_h|^q]$ è finita e mostra il seguente comportamento asintotico per $h \downarrow 0$:*

$$m_q(h) \sim \begin{cases} C_q\,(\lambda h)^{\frac{q}{2}} & \text{se } q < q^* \\ C_q\,(\lambda h)^{\frac{q}{2}} \log(\frac{1}{h}) & \text{se } q = q^* \\ C_q\,(\lambda h)^{Dq+1} & \text{se } q > q^* \end{cases}, \qquad \text{dove } \; q^* := \frac{1}{(\frac{1}{2} - D)}\,.$$

Theorem (Volatility autocorrelation). *Si assuma $\mathbb{E}[\sigma^2] < \infty$. La correlazione degli incrementi del processo X mostra il seguente comportamento asintotico per $h \downarrow 0$:*

$$\lim_{h \downarrow 0} \rho(|X_{s+h} - X_s|, |X_{t+h} - X_t|) \;=\; \rho(t-s) \;:=\; \frac{2}{\pi\, Var\left[\sigma\,|W_1|\,\sqrt{\theta'(S)}\right]}\, e^{-\lambda|t-s|}\, \phi(\lambda|t-s|)\,,$$

dove $\phi(x) := Cov\left(\sigma\,\sqrt{\theta'(S)}\,,\, \sigma\,\sqrt{\theta'(S+x)}\right)$ e $S \sim Exp(1)$ è indipendente da σ e W_1.

Nel Capitolo 4 consideriamo alcuni aspetti del nostro modello da un punto di vista numerico. Confrontiamo le predizioni teoriche del modello e i dati simulati con le serie storiche del Dow Jones Industrial Average (DJIA) e dello Standard & Poor's 500 (S&P 500), la prima su un periodo di 75 anni e la seconda su un periodo di 61 anni. Per tale confronto numerico, abbiamo deciso di focalizzarci soprattutto sulle seguenti quantità:

(a) Il *multiscaling dei momenti.*

(b) Il *decadimento della volatility autocorrelation.*

(c) La *distribuzione* di X_t.

(d) La *volatilità locale.*

L'idea è di calcolare *empiricamente* queste quantità sulle serie storiche e poi di confrontare i risultati con le previsioni *teoriche* del nostro modello (questo è reso possibile dal fatto che il processo $(X_t)_{t \geq 0}$ ha incrementi ergodici).

Per quanto riguarda il lavoro numerico, abbiamo usato le due seguenti versioni del modello:

- la prima con $\theta(t) \;=\; t^{2D}$
- la seconda con $\theta(t) \;:=\; \begin{cases} t^{2D} & \text{se } 0 < t < C \\ C^{2D} + 2\,D\,C^{2D-1}\,(t - C) & \text{se } t \geq C \end{cases}$,

  con C parametro scalare.

Una volta specificato il modello, abbiamo usato un approccio di minimizzazione ai minimi quadrati per la stima dei parametri, utilizzando un'opportuna funzione "costo" $\mathcal{L}_C$; il fit ottenuto in questo modo fra il modello calibrato e le serie storiche prese in considerazione è risultato notevole.

Infine, nell'ultimo capitolo presentiamo quanto siamo finora riusciti a fare nel campo dell'Option Pricing col nostro modello; benché non siamo ancora riusciti a trovare formule chiuse per il prezzaggio di opzioni, siamo comunque riusciti a ottenere qualche risultato interessante con metodi di tipo Montecarlo.

In particoloare, abbiamo studiato un mercato $\mathcal{M}$ composto da due soli titoli:

(1) il titolo non rischioso $S^{[0]} \equiv 1$;

(2) il titolo rischioso $S^{[1]}$. I prezzi del titolo rischioso, che denotiamo con S_t, seguono la dinamica

$$\mathrm{d}(\log S_t) = v_t \mathrm{d}B_t + r_t \mathrm{d}t, \qquad \text{i.e.} \qquad S_t = s_0 \exp\left(\int_0^t v_s \mathrm{d}B_s + \int_0^t r_s \, \mathrm{d}s \right),$$

dove r_s è il trend locale,

$$B_t = \int_0^{I_t} \frac{1}{I'(I^{-1}(s))} \mathrm{d}W_s, \qquad \text{e} \qquad v_t = \sqrt{I'(t)}.$$

Poiché tale mercato è evidentemente *incompleto*, abbiamo dovuto scegliere una *misura martingala equivalente* Q per poter applicare la General Pricing Formula per un'opzione X:

$$\Pi_X(t) = \mathbb{E}^Q \left[X | \mathcal{F}_t \right];$$

principalmente per ragioni di semplicità (e poiché comunque ci sembrava ragionevole come scelta), abbiamo deciso di definire Q nel seguente modo:

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = \exp\left(- \int_0^T \frac{v_t}{2} + \frac{r_t}{v_t} \mathrm{d}B_t - \frac{1}{2} \int_0^T \left(\frac{v_t}{2} + \frac{r_t}{v_t} \right)^2 \mathrm{d}t \right).$$

Sotto tale misura, il processo dei prezzi (scontati) segue una dinamica quasi identica a quella del nostro modello, e richiede di stimare solamente un nuovo parametro $\tau_{i(t)}$ (moralmente, la data dell'ultima crisi). Per stimare questo parametro abbiamo introdotto due algoritmi euristici, il primo basato sulla serie dei rezzi storici dell'indice, il secondo sui prezzi delle Call options basate su quell'indice. Inoltre, un'ulteriore applicazione del primo algoritmo ci ha permesso di identificare le buona precisione le date delle maggiori crisi economiche direttamente dalla serie storica dei prezzi dell'indice.

Infine, concludiamo la Tesi con un'analisi (ancora un po' rozza) della *volatility term structure* dei prezzi delle Call options ottenuti col nostro modello.

Summary

In this Ph.D Thesis we propose a simple stochastic model for time series which is analytically tractable, easy to simulate and which captures some relevant stylized facts of financial indexes, including *scaling properties*. We show that the model fits the *Dow Jones Industrial Average* time series in the period 1935-2009 with a remarkable accuracy and finally we derive some results on option pricing.

Recent developments in stochastic modelling of time series have been strongly influenced by the analysis of financial indexes. The basic model, that has given rise to the celebrated Black & Scholes formula [17, 24], assumes that the logarithm X_t of the price of the underlying index, after subtracting the trend, is given by

$$\mathrm{d}X_t = \sigma \, \mathrm{d}W_t,$$

where σ (the *volatility*) is a constant and $(W_t)_{t\geq 0}$ is a standard Brownian motion. It is well-know that, despite its success, this model is not consistent with a number of *stylized facts* that are empirically detected in many real time series. Some of these facts are the following:

- the volatility is not constant: in particular, it may have high peaks, that may be interpreted as *shocks* in the market;
- the empirical distribution of the increments $X_{t+h} - X_t$ of the logarithm of the price (the *log-returns*) has tails heavier than Gaussian;
- log-returns corresponding to disjoint time-interval are uncorrelated, but not independent: in fact, the correlation between the absolute values $|X_{t+h} - X_t|$ and $|X_{s+h} - X_s|$ has a slow decay in $|t-s|$, up to moderate values for $|t-s|$. This phenomenon is known as *clustering of volatility*.

In order to have a better fit with real data, many different models have been proposed to describe the volatility and the price process. In discrete-time, autoregressive models such as ARCH, GARCH and generalizations [15, 10, 3, 11] have been widely used, while in continuous time stochastic volatility models have had a great succes [6].

More recently (see [14, 9, 26]), other stylized facts of financial indexes have been pointed out, concerning the *scaling properties* of the empirical distribution of the log-returns. Consider the time series of an index $(s_i)_{1\leq i\leq T}$ over a period of $T \gg 1$ days and denote by p_h the

empirical distribution of the (detrended) log-returns corresponding to an interval of h days:

$$p_h(\cdot) := \frac{1}{T-h}\sum_{i=1}^{T-h}\delta_{x_{i+h}-x_i}(\cdot)\,, \qquad x_i := \log(s_i) - \overline{d}_i\,,$$

where $\overline{d}_i$ is the local rate of linear growth of $\log(s_i)$. The statistical analysis of various indexes shows that, for h within a suitable time scale, p_h obeys approximately a diffusive scaling relation

$$p_h(\mathrm{d}r) \simeq \frac{1}{\sqrt{h}}\, g\!\left(\frac{r}{\sqrt{h}}\right)\mathrm{d}r,$$

where g is a probability density with tails heavier than Gaussian.

Moreover, if one considers the q-th empirical moment $m_q(h)$, defined by

$$m_q(h) \;:=\; \frac{1}{T-h}\sum_{i=1}^{T-h}|x_{i+h}-x_i|^q \;=\; \int |r|^q\, p_h(\mathrm{d}r)\,,$$

what one observes for moments of small order $q \le \bar{q}$ is that $m_q(h)$ scales as $h^{\frac{q}{2}}$, while, for moments of higher order $q > \bar{q}$, the different scaling relation $h^{A(q)}$, with $A(q) < q/2$, takes place; this is the so-called *multiscaling of moments*.

A detailed description of all these statistic features of financial indexes is the content of Chapter 1.

In Chapter 2, starting from these properties and inspired by the work of Baldovin and Stella [26], we define a *simple* continuous-time stochastic model which agrees with *all* mentioned stylized facts; this is a non-trivial point, despite of the variety of models that can be found in the literature.

The basic mechanism of our model is a *random update of the volatility*. Given two real numbers $D \in (0, 1/2]$, $\lambda \in (0,\infty)$ and a probability ν on $(0,\infty)$ (these may be viewed as our general parameters), our model is defined upon the following three sources of alea:

- a standard Brownian motion $W = (W_t)_{t\ge 0}$;
- a Poisson point process $\mathcal{T} = (\tau_n)_{n\in\mathbb{Z}}$ on $\mathbb{R}$ with intensity λ (by convention we label the points of $\mathcal{T}$ so that $\tau_0 < 0 < \tau_1$);
- a sequence $\Sigma = (\sigma_n)_{n\ge 0}$ of independent and identically distributed positive random variables.

We assume that $W, \mathcal{T}, \Sigma$ are independent.

For $t \ge 0$, define

$$i(t) \;:=\; \sup\{n \ge 0 :\, \tau_n \le t\} \;=\; \#\{\mathcal{T}\cap(0,t]\}\,,$$

so that $\tau_{i(t)}$ is the location of the last point in $\mathcal{T}$ before t; then, we can introduce the proces $I = (I_t)_{t\ge 0}$ by

$$I_t \;:=\; \sigma^2_{i(t)}\theta\left(\lambda(t-\tau_{i(t)})\right) \;+\; \sum_{k=1}^{i(t)}\sigma^2_{k-1}\theta\left(\lambda(\tau_k-\tau_{k-1})\right) \;-\; \sigma^2_0\theta\left(-\lambda\tau_0\right),$$

where $\theta : [0, +\infty) \to [0, +\infty)$ is a time rescaling function that satisfies the following properties:

(1) θ is concave, strictly increasing, $\theta(0) = 0$, and $\theta(s) \to +\infty$ as $s \to +\infty$.

(2) θ is $\mathcal{C}^1$ on $(0, +\infty)$, and there exists a constant $0 < D < \frac{1}{2}$ such that

$$\lim_{s \downarrow 0} \frac{\theta'(s)}{2Ds^{2D-1}} = 1.$$

In particular then, $\theta(s)$ behaves as s^{2D} near $s = 0$, and this behaviour produces singularities on I_t at the points τ_i.

Our basic process $X = (X_t)_{t \geq 0}$ can be then defined by setting

$$X_t := W_{I_t} ,$$

that is an independent random time change of a Brownian motion.

Moreover, in this chapter we derive some basic properties for the process, in particular we prove that:

(A) *X has stationary increments.*

(B) *X can be represented as a stochastic volatility process*:

$$\mathrm{d}X_t = v_t \, \mathrm{d}B_t ,$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion. More precisely, denoting $I'(s) := \frac{\mathrm{d}}{\mathrm{d}s} I(s)$, the variables B_t and v_t are defined by

$$B_t := \int_0^{I_t} \frac{1}{\sqrt{I'(I^{-1}(u))}} \, \mathrm{d}W_u , \qquad v_t := \sqrt{I'(t)} = \sigma_{i(t)} \sqrt{\lambda \, \sigma' \left(\lambda (t - \tau_{i(t)}) \right)} .$$

(C) *X is a zero-mean, square-integrable martingale* (provided $\mathbb{E}[\sigma^2] < \infty$).

In Chapter 3, we prove three main asymptotic results for the process X. They concern some of the stylized facts that we have mentioned in Chapter 1, in particular: the diffusive scaling of the distributions of log-returns, the multiscaling of moments and the clustering of volatility.

Theorem. *As $h \downarrow 0$ we have the convergence in distribution*

$$\frac{(X_{t+h} - X_t)}{\sqrt{h}} \xrightarrow[h \downarrow 0]{d} f(x) \, \mathrm{d}x ,$$

where f is a mixture of centered Gaussian densities, namely

$$f(x) = \int_0^\infty \nu(\mathrm{d}\sigma) \int_0^\infty \mathrm{d}t \, \lambda e^{-\lambda t} \, \frac{1}{\sigma \sqrt{2\pi \theta'(\lambda t)}} \, e^{-\frac{x^2}{2\sigma^2 \theta'(\lambda t)}} .$$

Theorem (Multiscaling of moments). *Let $q > 0$, and assume $\mathbb{E}[\sigma^q] < +\infty$. Then the quantity $m_q(h) := \mathbb{E}[|X_{t+h} - X_t|^q] = \mathbb{E}[|X_h|^q]$ is finite and has the following asymptotic behavior as $h \downarrow 0$:*

$$m_q(h) \sim \begin{cases} C_q\,(\lambda h)^{\frac{q}{2}} & \text{if } q < q^* \\ C_q\,(\lambda h)^{\frac{q}{2}} \log(\frac{1}{h}) & \text{if } q = q^* \\ C_q\,(\lambda h)^{Dq+1} & \text{if } q > q^* \end{cases}, \qquad \text{where } \; q^* := \frac{1}{(\frac{1}{2} - D)}\,.$$

Theorem (Volatility autocorrelation). *Assume that $\mathbb{E}[\sigma^2] < \infty$. The correlation of the increments of the process X has the following asymptotic behavior as $h \downarrow 0$:*

$$\lim_{h \downarrow 0} \rho(|X_{s+h} - X_s|, |X_{t+h} - X_t|) = \rho(t-s) := \frac{2}{\pi\, Var\left[\sigma\,|W_1|\,\sqrt{\theta'(S)}\right]}\, e^{-\lambda|t-s|}\, \phi(\lambda|t-s|)\,,$$

where $\phi(x) := Cov\left(\sigma\,\sqrt{\theta'(S)}\,,\, \sigma\,\sqrt{\theta'(S+x)}\right)$ and $S \sim Exp(1)$ is independent of σ and W_1.

In Chapter 4 we consider some aspects of our model from a numerical viewpoint. We compare the theoretical predictions and the simulated data of our model with the time series of the Dow Jones Industrial Average (DJIA) and of the Standard & Poor's 500 (S&P 500), the first over a period of 75 years and the second over a period of 61 years.

For the numerical comparison of our process $(X_t)_{t \geq 0}$ with the empirical time series, we have decided to focus on the following quantities:

(a) The *multiscaling of moments*.

(b) The *volatility autocorrelation decay*.

(c) The *distribution* of X_t.

(d) The *local volatility*.

Roughly speaking, the idea is to compute *empirically* these quantities on the time series and then to compare the results with the *theoretical* predictions of our model (this is justified by the the fact that our process $(X_t)_{t \geq 0}$ has ergodic increments).

For the numerical analysis, we used the following two easy versions of the model:

- the first with $\theta(t) = t^{2D}$
- the second with $\theta(t) := \begin{cases} t^{2D} & \text{if } 0 < t < C \\ C^{2D} + 2\,D\,C^{2D-1}\,(t - C) & \text{if } t \geq C \end{cases}$,

  with C a scalar parameter.

Once specified the model, we use a least-squared minimization approach for the estimation of the parameters, considering a loss function $\mathcal{L}_C$, and we find out a really impressive fit between the model and the emprical time series.

Finally, in the last Chapter we present what we have done by now about Option Pricing with this model; although we didn't find any closed pricing formula yet, by montecarlo methods we managed to obtain some useful results.

We study a market $\mathcal{M}$ composed by only two assets:

(1) the risk-free asset $S^{[0]} \equiv 1$;

(2) the riskly asset $S^{[1]}$. The prices of the riskly asset, that will be denoted by S_t, follows the dynamics

$$\mathrm{d}(\log S_t) = v_t \mathrm{d}B_t + r_t \mathrm{d}t, \qquad \text{i.e.} \qquad S_t = s_0 \exp\left(\int_0^t v_s \mathrm{d}B_s + \int_0^t r_s \, \mathrm{d}s \right),$$

where r_s is the local trend,

$$B_t = \int_0^{I_t} \frac{1}{I'(I^{-1}(s))} \mathrm{d}W_s, \qquad \text{and} \qquad v_t = \sqrt{I'(t)}.$$

Since the market is clearly *incomplete*, we had to fix an *equivalent martingale measure* Q in order to apply the General Pricing Formula to a claim X:

$$\Pi_X(t) = \mathbb{E}^Q \left[X | \mathcal{F}_t \right].$$

Due overall to simplicity reasons, we defined Q by the Girsanov formula

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = \exp\left(- \int_0^T \frac{v_t}{2} + \frac{r_t}{v_t} \mathrm{d}B_t - \frac{1}{2} \int_0^T \left(\frac{v_t}{2} + \frac{r_t}{v_t} \right)^2 \mathrm{d}t \right);$$

under this measure, the (discounted) price process is just a very sligth modification of our model. In particular, it requires to estimate only one new parameter, namely $\tau_{i(t)}$, that we estimate by two different euristic algorithms: the first algorithm is based on hystorical index prices data, the second on Call options prices data. Moreover, as an application of the first algorithm, we identify with a good precision the dates of big economical crises in history direclty from the hystorical prices of the index.

Finally, we conclude the Thesis with a (still rough) analysis of the volatility term structure for the Call options prices derived according with our model.

Introduction

General Notations & Definitions

Here are some notations to be used in all the thesis.

- If $(\Omega, \mathcal{F}, P)$ is a probability space, we denote by $\mathbb{E}[X]$ the expectation of any integrable random variable X; if there is some ambiguity as to the measure P, we write $\mathbb{E}^P[X]$. If $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$, the conditional expectation of the variable X is well-defined whenever X is integrable or nonnegative or nonpositive, and we denote by $\mathbb{E}[X|\mathcal{G}]$ any version of it.

- A *filtration* $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}^+}$ is any increasing and right-continuous family of sub-σ-fields (that is $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leq t$ and $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$). We will denote by $\mathcal{F}_t^S$ the filtration generated by the stochastic process S, that is the minimal filtration such that S_t is $\mathcal{F}_t$-measurable.

- A *filtered probability space* $(\Omega, \mathcal{F}, P, \mathbb{F})$ is called *complete* if the σ-field $\mathcal{F}$ is P-complete and if every $\mathcal{F}_t$ contains all P-null sets of $\mathcal{F}$. In general we will consider complete filtered probability spaces if not differentely specified.

- A process S is called *càd (resp. càg, resp. càdlàg)* if all its paths are right-continuous (resp. are left-continuous, resp. are right-continuous and admit left-hand limit). When S is càdlag we define two other processes, S^- and ΔS, by

$$S_0^- = S_0, \quad S_t^- = \lim_{s\uparrow t} S_s \quad \text{for } t > 0$$

$$\Delta S_t = S_t - S_t^-$$

- A process $(S_t)_{t\in\mathbb{R}^+}$ is *adapted* to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$ if X_t is $\mathcal{F}_t$ measurable for every $t \in \mathbb{R}^+$.

  The *optional σ-field* is the σ-field $\mathcal{O}$ on $\Omega \times \mathbb{R}^+$ that is generated by all càdlàg adapted processes.

  The *predictable σ-field* is the σ-field $\mathcal{P}$ on $\Omega \times \mathbb{R}^+$ that is generated by all càg adapted processes.

  A process or random set that is $\mathcal{O}$-measurable (resp. $\mathcal{P}$-measurable) is called *optional* (resp *predictable*)

- If S is a process and T is a mapping: $\Omega \to \overline{\mathbb{R}}^+$, we define the *process stopped at time* T, denoted by S^T, by
$$S_t^T = S_{t\wedge T}$$

- A random set A (that is a subset $A \subset \Omega \times \mathbb{R}$) is called *evanescent* if the set
$$\{\omega : \exists t \in \mathbb{R}^+ \text{ with } (\omega, t) \in A\}$$
is P-null; two processes S and Z are called *indistinguishable* if the set
$$\{S = Z\} = \{(\omega, t) : S_t(\omega) \neq Z_t(\omega)\}$$
is evanescent. As for random variables, in most cases $S = Z$ for stochastic processes means "up to an evanescent set".

- A *stopping time* is a mapping $T : \Omega \to \overline{\mathbb{R}}^+$ such that $\{T \leq t\} \in \mathcal{F}_t$ for all $t \in \mathbb{R}^+$.

- If $\mathcal{C}$ is a class of processes, we denote by $\mathcal{C}_{loc}$ the *localized class*, defined as follows: a process S belongs to $\mathcal{C}_{loc}$ if there exists an increasing sequence (T_n) of stopping times (depending on S) such that $\lim_{n\to\infty} T_n = \infty$ *a.s.* and that each stopped process S^{T_n} belongs to $\mathcal{C}$. The sequence (T_n) is called a *localizing sequence* for S (relative to $\mathcal{C}$)

- A process S is a *martingale* (resp. *submartingale*, resp. *supermartingale*) with respect to the filtration $\mathbb{F}$ if S is $\mathbb{F}$-adapted, every S_t is integrable and
$$S_s = \mathbb{E}[S_t|\mathcal{F}_s] \quad \big(\text{resp. } S_s \leq \mathbb{E}[S_t|\mathcal{F}_s],\ \text{resp. } S_s \geq \mathbb{E}[S_t|\mathcal{F}_s]\big),$$
for every $s \leq t$.

  We denote by $\mathcal{M}$ the class of all *uniformly integrable martingales*, that is of all martingales S such that the family of random variables $(S_t)_{t\in\mathbb{R}^+}$ is uniformly integrable.

  We say that a martingale S is *square-integrable* if $\mathbb{E}[S_t^2] < \infty$ for all $t \in \mathbb{R}^+$, while we denote by $\mathcal{H}^2$ the class of all $\mathbf{L}^2$*-martingales*, that is of all martingales S such that $\sup_{t\in\mathbb{R}^+} \mathbb{E}[S_t^2] < \infty$.

- A *local martingale* (resp. a *locally square-integrable martingale*) is a process that belongs to the localized class $\mathcal{M}_{loc}$ (resp. $\mathcal{H}^2_{loc}$) constructed from $\mathcal{M}$ (resp. $\mathcal{H}^2$) by the localization procedure above.

- A *Wiener process* (or *Brownian motion*) on $(\Omega, \mathcal{F}, \mathbb{F}, P)$ is a continuous adapted process W such that $W_0 \equiv 0$ and:

  (1) $W_t - W_s \sim N(0, t-s)$ for all $0 < s \leq t$, where $N(m, \sigma^2)$ denote the Gaussian random variable with mean m and variance σ^2;

  (2) $W_t - W_s$ is independent of the σ-field $\mathcal{F}_s$ for all $0 \leq s \leq t$

  The function $\sigma^2(t) = \mathbb{E}[W_t^2]$ is called the *variance function* (or, in financial terms, the volatility) of W. The Wiener process is a continuous martingale.

- A *point process* on $\mathbb{R}^n$ is a mapping from $(\Omega, \mathcal{F}, P)$ to the locally finite subsets of $\mathbb{R}^n$.

- If $T = (T_n)_{n\in\mathbb{N}}$ is a strictly increasing sequence of positive random variables, with $T_0 = 0$ *a.s.*, the process $N = (N_t)_{t\in\mathbb{R}^+}$ is called the *counting process* associated to the sequence $(T_n)_{n\in\mathbb{N}}$ if
$$N_t = \sum_{n\geq 1} \mathbf{1}_{\{t\geq T_n\}}.$$
An adapted counting process N is a *Poisson process* on $(\Omega, \mathcal{F}, \mathbf{F}, P)$ if
  (1) $N_t - N_s \sim N_v - N_u$ for all s, t, u, v such that $0 \leq s < t$, $0 \leq u < v$, $t - s = v - u$;
  (2) $N_t - N_s$ is independent of the σ-field $\mathcal{F}_s$ for all $0 \leq s \leq t$.

  The function $\lambda(t) = \mathbb{E}[N_t]$ is called the *intensity* of N. It turns out that $N_t \sim Po(\lambda(t))$, where $Po(\lambda)$ denotes a Poisson random variable of parameter λ.

  Given a Poisson process N with constant intensity $\lambda(t) \equiv \lambda > 0$, and a sequence of *i.i.d.* random variables V_i with density $\nu(V_1)$, independent from N, the process Y such that
$$Y_t = \sum_{i=1}^{N(t)} V_i$$
is called *compound Poisson process* with intensity λ and jump distribution $\nu(V_1)$.

- Let ν be a borel measure on $\mathbb{R}^n$: a point process N on $\mathbb{R}^n$ is a *Poisson point process* with intensity ν if
  (1) $N(A) \sim Po(\nu(A))$ for every bounded borel set $A \subset \mathbb{R}^n$;
  (2) for any k disjoint Borel sets $A_1, \ldots, A_k \subset \mathbb{R}^n$ the random variables
$$N(A_1), \ldots, N(A_k)$$
  are independent.

Lévy Processes

In this section we will give an essential overview on Lévy Processes and semimartingales. Lévy processes are a rather general class of stochastic processes that are often used as price models in Financial Mathematics, semimartingales are a larger class of processes and represent the most general case for which the stochastic integration is defined. For the proves of all the theorems presented in this section we remind to [19] and [20].

Definition 0.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space. An $\mathbb{R}^d-$valued process S such that $S_0 = 0$ is a *Lévy Process* if

- for every $s, t \geq 0$, $S_{t+s} - S_s$ is independent of $\mathcal{F}_s^S$,
- for every $s, t \geq 0$ the random variables $S_{t+s} - S_s$ and S_t have the same law
- S is continuous in probability, *i.e.* for fixed t $P(|S_t - S_u| > \epsilon) \to 0$ when $u \to t$ for every $\epsilon > 0$

Remark: the second property of Lévy processes (called *stationariity*) is necessary to allow for some statistical analysis of time series modelled by a Lévy process.

Example 0.2. Brownian Motion, Poisson process and compound Poisson process are examples of Lévy processes. If S is a Lévy process, C a matrix and D a vector, $CS_t + Dt$ is also a Lévy process. More generally, the sum of two independent Lévy processes is a Lévy process.

Example 0.3. The *Gamma process* is a Lévy process S such that $S_t \sim \Gamma(at, \beta)$, with $\beta \geq 0$. More precisely, the density for S_t is

$$f_t(s) = \frac{\beta^{at}}{\Gamma(at)} s^{at-1} e^{-\beta s} \mathbf{1}_{\{s>0\}}.$$

The next propositions and representation theorems will show that a Lévy Process is essentially composed of four parts:

- a drift,
- a diffusive term,
- a bounded jump term,
- an unbounded jump term.

Definition 0.4. A random variable X with distribution μ and taking values in $\mathbb{R}^d$ is said to be *infinitely divisible* if its characteristic function may be written, for any integer n, as the $n^{th}-$power of another characteristic function $\widehat{\mu}_n$, that is if

$$\widehat{\mu}(u) = (\widehat{\mu}_n(u))^n.$$

With a slight abuse of language, we shall say that such a characteristic function (or distribution function) is infinitely divisible. Equivalently, X is infinitely divisible if

$$\forall n,\ \exists\, X_i^{(n)}\ i.i.d.\ \text{such that}\ X =^d X_1^{(n)} + X_2^{(n)} + \cdots + X_n^{(n)}.$$

Definition 0.5. A *Lévy measure* on $\mathbb{R}^d$ is a positive measure ν on $\mathbb{R}^d \backslash \{0\}$ such that

$$\int_{\mathbb{R}^d \backslash \{0\}} (1 \wedge |x|^2) \nu(\mathrm{d}x) < \infty.$$

Proposition 0.6 (Lévy-Khintchine Representation)**.** *If X is an infinitely divisible random variable with characteristic function $\widehat{\mu}$, there exists a unique triple (m, A, ν), where $m \in \mathbb{R}^d$, A is a positive matrix and ν is a Lévy measure, such that*

$$\widehat{\mu}(u) = \exp\left(iu \cdot m - \frac{1}{2} u \cdot Au + \int_{\mathbb{R}^d} (e^{iu \cdot x} - 1 - iu \cdot x\, \mathbf{1}_{|x| \leq 1}) \nu(\mathrm{d}x) \right) \tag{0.0.1}$$

The triple (m, A, ν) is called the *characteristic triple* of μ.

Remark 0.7. When $\int_{\mathbb{R}^d\setminus\{0\}} \mathbf{1}_{|x|\leq 1}|x|\nu(\mathrm{d}x) < \infty$, we can write (0.0.1) in the *reduced form*

$$\widehat{\mu}(u) = \exp\Big(iu\cdot m_0 - \frac{1}{2}u\cdot Au + \int_{\mathbb{R}^d}(e^{iu\cdot x} - 1 - iu\cdot x)\nu(\mathrm{d}x)\Big), \tag{0.0.2}$$

where $m_0 = m - \int_{\mathbb{R}^d} x\,\mathbf{1}_{|x|\leq 1}\nu(\mathrm{d}x)$.

An important and useful property of Lévy Processes is that they allow trajectories with *jumps*. In fact, let S be a Lévy process; then, for every bounded Borel set $\Lambda \in \mathbb{R}^d$ such that $0 \notin \overline{\Lambda}$ (where $\overline{\Lambda}$ is the closure of Λ) we define

$$N_t^\Lambda = \sum_{0<s\leq t} \mathbf{1}_\Lambda(\Delta S_s),$$

that is the number of jumps up to time t which take values in Λ.

Proposition 0.8. *For each fixed $(t,\omega) \in \mathbb{R}^+\times\Omega$, the map $\Lambda \to N_t^\Lambda(\omega)$ define a $\sigma-$finite measure on $\mathbb{R}^d\setminus 0$ denoted by $N_t(\omega,\mathrm{d}x)$.*

Definition 0.9. The $\sigma-$additive measure ν defined on $\mathbb{R}^d\setminus 0$ by

$$\nu(\Lambda) = \mathbb{E}[N_1^\Lambda]$$

is called the *Lévy measure* of the process S.

Proposition 0.10. *Let S be a Lévy process with Lévy measure ν.*

(1) *Assume $\nu(\Lambda) < \infty$. Then, the process*

$$N_t^\Lambda = \sum_{0<s\leq t} \mathbf{1}_\Lambda(\Delta S_s),\ \ t \geq 0$$

is a Poisson process with intensity $\nu(\Lambda)$

(2) *Let Λ_i be a finite collection of sets and assume that for all i $\nu(\Lambda_i) < \infty$. The processes N_i^Λ are independent if and only if $\nu(\Lambda_i\cap\Lambda_j) = 0$, $i \neq j$ and then in particular if the sets Λ_i are disjoint. If $\Lambda = \cup_i\Lambda_i$, with disjoint Λ_i, then $N^\Lambda = \sum_i N^{\Lambda_i}$ is a Poisson process with intensity $\sum_i \nu(\Lambda_i)$.*

Let Λ be a Borel set of $\mathbb{R}^d$ with $0 \notin \overline{\Lambda}$, and let f be a positive Borel function defined on Λ. We have, by definition of the random measure $N_t(\omega,\mathrm{d}x)$ (given in Proposition 0.8),

$$\int_\Lambda f(x)N_t(\omega,\mathrm{d}x) = \sum_{0<s\leq t} f(\Delta S_s(\omega))\mathbf{1}_\Lambda(\Delta S_s(\omega)).$$

Proposition 0.11. *Let S be a Lévy process taking values in $\mathbb{R}^d$. Then, for each t, the random variable S_t is infinitely divisible and its characteristic function is given by the Lévy-Khintchine formula: for $u \in \mathbb{R}^d$,*

$$\mathbb{E}[\exp(iu\cdot S_t)] = \exp\Big(t\Big(iu\cdot m - \frac{u\cdot Au}{2} + \int_{\mathbb{R}^d}(e^{iu\cdot s} - 1 - iu\cdot s\,\mathbf{1}_{|s|\leq 1}\nu(\mathrm{d}s))\Big)\Big),$$

where $m \in \mathbb{R}^d$, A is a positive semi-definite matrix, and ν is a Lévy measure on $\mathbb{R}^d\setminus\{0\}$. The random variable S_t admits the characteristic triple $(tm, tA, t\nu)$ and we say in short that S is a (m,A,ν) Lévy process.

- If $\nu(\mathbb{R}^d\backslash\{0\}) < \infty$, the process S has a finite number of jumps in any finite time interval. In finance, when $\nu(\mathbb{R}^d) < \infty$, one refers to *finite activity*.

  If moreover $A = 0$, the process S is a compound Poisson process with drift.

- If $\nu(\mathbb{R}^d\backslash\{0\}) = \infty$, the process corresponds to *infinite activity*. Assume moreover that $A = 0$. Then

  (1) If $\int_{|s|\leq 1} |s|\nu(\mathrm{d}s) < \infty$, the paths of S are of bounded variation on any finite time interval.

  (2) If $\int_{|s|\leq 1} |s|\nu(\mathrm{d}s) = \infty$, the paths of S are no longer of bounded variation on any finite interval.

Example 0.12. Some examples of Lévy processes and their characteristic triples:

- *Drifted Brownian Motion.* The process $(mt + \sigma W_t,\, t \geq 0)$, where W is a Brownian Motion, is a Lévy process with characteristic triple $(m, \sigma^2, 0)$. This is the only Lévy process with continuous paths.

- *Gamma process.* the characteristic triple for the Gamma process is $(0, 0, \frac{a}{s}e^{-\beta s}\mathbf{1}_{\{s>0\}}\mathrm{d}s)$ and it is a pure jump, increasing process.

- *subordinators.* Subordinators are Lévy processes that take values in $[0, \infty[$ (or equivalently, which have increasing paths). In this case the parameters of the characteristic triple are $m \geq 0$, $\sigma = 0$ and the Lévy measure ν is a measure on $]0, \infty[$ with $\int_{]0,\infty[}(1 \wedge s)\nu(\mathrm{d}s) < \infty$.

Proposition 0.13 (Lévy-Ito's Decomposition)**.** *If S is an $\mathbb{R}^d-$valued Lévy process, it can be decomposed into*

$$S = Y^{(0)} + Y^{(1)} + Y^{(2)} + Y^{(3)},$$

where $Y^{(0)}$ is a constant drift, $Y^{(1)}$ is a linear transform of a Brownian Motion, $Y^{(2)}$ is a compound Poisson process with jump sizes greater than or equal to 1 and $Y^{(3)}$ is a Lévy process with jump sizes smaller than 1. The processes $Y^{(i)}$ are independent.

From the Lévy-Ito's Decomposition follows this representation formula for Lévy processes:

$$X_t = mt + Z_t + \int_0^t \int x\mathbf{1}_{\{|z|\leq 1\}}\widetilde{N}(\mathrm{d}s, \mathrm{d}x) + \sum_{s\leq t} \Delta X_s\, \mathbf{1}_{\{|\Delta X_s|>1\}}, \tag{0.0.3}$$

where Z is an $\mathbb{R}^d-$valued Brownian Motion with correlation matrix A, and $\widetilde{N}(\mathrm{d}t, \mathrm{d}x) = N(\mathrm{d}t, \mathrm{d}x) - \mathrm{d}t\,\nu(\mathrm{d}x)$.

- If $\int_{\{|x|\leq 1\}} |x|\nu(\mathrm{d}x) < \infty$, the process X can be represented as

$$X_t = m_0\, t + Z_t + \int_0^t \int xN(\mathrm{d}s, \mathrm{d}x),$$

  where $m_0 = m - \int x\mathbf{1}_{\{|x|\leq 1\}}\nu(\mathrm{d}x)$. If ν is a finite measure, the process $\int_0^t \int xN(\mathrm{d}s, \mathrm{d}x)$ is a compound Poisson process: indeed, let $T_k, k \geq 1$ be the jump times of X and

define $N_t = \sum \mathbf{1}_{\{T_k \leq t\}}$ the associated counting process. The process N is a Poisson process with intensity $\lambda = \nu(\mathbb{R}^d \backslash \{0\})$. Furthermore, the random variables $Y_k = \int xN(\{T_k\}, \mathrm{d}x)$ are *i.i.d.* with law $\nu(\mathrm{d}x)/\lambda$ and $\sum_1^{N_t} Y_k = \int_0^t \int xN(\mathrm{d}s, \mathrm{d}x)$.

- if $\int_{\{|x|\leq 1\}} |x|\nu(\mathrm{d}x) = \infty$, it is not possible to separate the integral

$$\int_{|x|\leq 1} x(N_t(\cdot, \mathrm{d}x) - t\nu(\mathrm{d}x))$$

into the two parts $\int_{|x|\leq 1} xN_t(\cdot, \mathrm{d}x)$ and $t \int_{|x|\leq 1} x\nu(\mathrm{d}x)$ which may both be ill-defined.

A short digression on semimartingales

Lévy processes are particular cases of a more general class of sthocastic processes: the *semimartingales*. Semimartingales are the most general processes for which there exists an integration theory.

Definition 0.14. A stochastic process X is a *semimartingale* if it is of the form

$$X = X_0 + M + A,$$

where M is a local martingale and A is a càdlàg adapted process of locally bounded variation.

By the Lévy-Ito's Decomposition we have that a Lévy process is in particular a semimartingale.

Let X be a (m, σ^2, ν) real-valued Lévy process and $f \in C^{1,2}(\mathbb{R}^+ \times \mathbb{R}, \mathbb{R})$. Then, since X is a semimartingale, we can apply Ito's formula in the optional or predictable form:

- The optional Ito's formula is

$$f(t, X_t) = f(0, X_0) + \int_0^t \partial_t f(s, X_s)\mathrm{d}s + \frac{\sigma^2}{2}\int_0^t \partial_{xx} f(s, X_s)\mathrm{d}s + \int_0^t \partial_x f(s, X_{s^-})\mathrm{d}X_s +$$
$$+ \sum_{s\leq t} f(s, X_{s^-} + \Delta X_s) - f(s, X_{s^-}) - (\Delta X_s)\partial_x f(s, X_{s^-}).$$

- If f and its first and second derivatives w.r.t. x are bounded, the predictable Ito's formula is

$$\mathrm{d}f(t, X_t) = \partial_x f(t, X_t)\sigma\, \mathrm{d}W_t + \int_{\mathbb{R}} (f(t, X_{t^-} + x) - f(t, X_{t^-}))\, \widetilde{N}(\mathrm{d}t, \mathrm{d}x) +$$
$$+ \Big[\frac{1}{2}\sigma^2 \partial_{xx} f(t, X_t) + m\partial_x f(t, X_t) + \partial_t f(t, X_t) +$$
$$+ \int_{\mathbb{R}} (f(t, X_{t^-} + x) - f(t, X_{t^-}) - x\partial_x f(t, X_{t^-})\mathbf{1}_{|x|\leq 1})\nu(\mathrm{d}x)\Big]\mathrm{d}t.$$

Finally for this section, we want to introduce the quadratic variation (and its discrete part) of a semimartingale, that appears in the construction of an importan financial model, the COGARCH model, see Example 0.45.

Definition 0.15. Two local martingales X and Y are called *orthogonal* if their product $X\,Y$ is again a local martingale.

Definition 0.16. A local martingale X is called *purely discontinuous* if $X_0 = 0$ and if it is orthogonal to all continuous local martingales

Theorem 0.17. *Any local martingale X admits unique decomposition*

$$X = X_0 + X^c + X^d,$$

where $X_0^c = X_0^d = 0$, X^c is a continuous local martingale, and X^d is a purely discontinuous local martingale.

X^c is called the *continuous part* of X and X^d its *purely discontinuous part.*

Analoguosly, a similar decomposition holds for semimartingales:

Proposition 0.18. *Let $X = X_0+M+A$ be a semimartingale, with the notation in Definition 0.14; then, there is a unique continuous local martingale X^c such that $M^c = X^c$.*

X^c is called the *continuous martingale part* of X, while $X - X_0 - X^c$ is called the *discrete part* of X.

Definition 0.19. Given a pair (X, Y) of locally square-integrable martingales, we call *predictable quadratic variation* $\langle X, Y\rangle$ the unique càdlàg predictable process of finite variation such that $X\,Y - \langle X, Y\rangle$ is a local martingale.

Definition 0.20. If X and Y are semimartingales, we call *quadratic covariation* of X and Y the process

$$[X, Y]_t = \langle X^c, Y^c\rangle_t + \sum_{s\le t} \Delta X_s \Delta Y_s.$$

If $X = Y$, then we call $[X, X]$ the *quadratic variation* of X and $\sum_{s\le t} \Delta X_s^2$ is its *discrete part*, that we will denote $[X, X]^d$.

Arbitrage Theory and Option Pricing

In this section we present one of the most relevant topic in Financial Mathematics: Option Pricing. This field tries to give an answer to one of the most important question in economy: what is a fair price for a financial product I want to sell in the market?

For the proves of all the theorems presented in this section we remind to [28] and [29].

Definition 0.21. We call $S_t^{(0)}$ a *risk-free asset* if its price is locally deterministic, that is if its price process is predictable with respect to the filtration $\mathcal{F}_t^S$. The dynamics of a risk-free asset is:

$$\mathrm{d}S_t^{(0)} = r_t S_t^{(0)} \mathrm{d}t,$$

with r_t any predictable process.

We call $S_t^{(i)}$, $i = 1, \dots, N$ *riskly assets* if their prices are not locally deterministic, that is if their price process are not predictable with respect to the filtration $\mathcal{F}_t^S$.

All along this chapter, we will denote with $\mathcal{M}$ the market, composed by N riskly traded assets and one non-riskly traded asset; we assume $S_t^{(0)} > 0$ for all t.

Definition 0.22. The *normalized economy* for the market $\mathcal{M}$ (also referred to as $Z-$market) is a market $\mathcal{M}_Z$ for which the price process Z is

$$Z_t = \left[1, Z_t^{(1)}, \dots, Z_t^{(N)}\right]^T = \left[\frac{S_t^{(0)}}{S_t^{(0)}}, \frac{S_t^{(1)}}{S_t^{(0)}}, \dots, \frac{S_t^{(N)}}{S_t^{(0)}}\right]^T;$$

we call Z the *discounted prices process.*

The normalized economy is strongly related to the market $\mathcal{M}$, as we will se in Lemma 0.31

Definition 0.23. A *portfolio strategy* (more often simply called a portfolio) is a predictable $(N+1)-$dimensional process $(h(t))_{t\geq 0}$ for which there exists a nonnegative real number α such that

$$\int_0^t h(u) \cdot \mathrm{d}Z_u \geq \alpha \quad \text{for all } t \in [0, T]. \tag{0.0.4}$$

The process $h_i(t)$ represents the number of shares of type i held at time t, and the condition 0.0.4 (called the *admissibility condition*) is necessary to avoid trivial strategies for making money without taking risk (see [29] pag. 138 for more details).

In financial terms, we will say that the portfolio h has a *long* position on the asset i at time t if $h_i(t) > 0$, while it has a *short* position if $h_i(t) < 0$. Remark that, from Definition 0.23, short position and in general real quantity holding are allowed, but unlimited short positions are not.

Throughout the thesis, we will make the following general assumptions for the Market:

- There is no *bid-ask spread*, that is the selling price of an asset is the same of its buying price.
- There are no transaction costs of trading.

These assumptions mean that we are dealing with *linear markets*. With these assumptions the value at time t of an arbitrary portfolio is

$$V_h(t) = h(t) \cdot S_t = \sum_{i=0}^{N} h_i(t) S_t^{(i)}. \tag{0.0.5}$$

Definition 0.24. A portfolio strategy is called *self-financing* if its value process satisfies

$$\mathrm{d}V_h(t) = \sum_{i=0}^{N} h_i(t) \mathrm{d}S_t^{(i)} \quad \text{for all } t \in [0, T].$$

To make more clear this definition, we will euristically derive it from the discrete time setting $\{t_0, \dots, t_{j-1}, t_j, \dots\}$; only for now, when we use the word portoflio we really mean portfolio (that is a porfolio strategy evalueted at a certain time t_j), and not the whole portfolio strategy. The idea is that, at time t_j, *i.e.* at the *start* of period t_j, we bring with us an "old" portfolio $h(t_{j-1})$ and we choose a "new" portfolio $h(t_j)$, to be held until time t_{j+1}. The portfolio is self-financing if the purchase of the new portfolio is financed solely by selling assets already in the portfolio; then, the value of the new portfolio $h(t_j)$ at time t_j must be equal to the value of the old portfolio $h(t_{j-1})$ at time t_j, that is

$$V_{h(t_j)}(t_j) = V_{h(t_{j-1})}(t_j).$$

Calling $t = t_j$ and $t_{j-1} = t - \Delta t$ and using 0.0.5 we obtain the *budget equation*

$$h(t) \cdot S_t = h(t - \Delta t) \cdot S_t. \tag{0.0.6}$$

With the natural notation $\Delta h(t) = h(t) - h(t - \Delta_t)$ and $\Delta S_t = S_t - S_{t-\Delta t}$, adding and subtracting the term $S_{t-\Delta t} \cdot \Delta h(t)$ to the left-hand side of equation 0.0.6 and letting $\Delta t \to 0$, we obtain

$$S_t \mathrm{d}h(t) + \mathrm{d}h(t)\mathrm{d}S_t = 0. \tag{0.0.7}$$

If we take the Itô differential of expression 0.0.5 we get

$$\mathrm{d}V_{h(t)}(t) = h(t)\mathrm{d}S_t + S_t \mathrm{d}h(t) + \mathrm{d}S_t \mathrm{d}h(t),$$

and substituing 0.0.7 we finally obtain

$$\mathrm{d}V_{h(t)}(t) = h(t)\mathrm{d}S_t.$$

Definition 0.25. An *arbitrage possibility* on a financial market is a self-financing portfolio h for which exist two times $0 \leq t < T$ such that

- $V_h(t) = 0$
- $P(V_h(T) \geq 0) = 1$
- $P(V_h(T) > 0) > 0$.

Essentially, an arbitrage possibility is a possibility of making a positive amount of money without taking any risk. A market is said to be *arbitrage-free* if there are no arbitrage possibilities.

Definition 0.26. A *contingent claim* with date of maturity (exercise date) T, also called a $T-$claim, is any stochastic variable $X \in \mathcal{F}_T^S$.

A contingent claim is called a *simple claim* if it is of the form $X = \Phi(S_T)$ and Φ is called the *contract function*.

Example 0.27. Two important examples of simple claims are the *European Call Option* and the *European Put Option*. A Euorepean Call Option on the *underlying asset* $S^{(i)}$, with *strike price* K and time of maturity T, is a claim

$$X = (S_T^{(i)} - K)^+;$$

this option gives the right (but not the obligation) to buy at time T the asset i at price K instead of paying it $S_T^{(i)}$.

Simmetrically, a European Put Option is the claim

$$X = (K - S_T^{(i)})^+;$$

it gives the right (but not the obligation) to sell at time T the asset i at price K instead that $S_T^{(i)}$.

Definition 0.28. An *hedging portfolio* for a claim X with date of maturity T is a self-financed portfolio h such that

$$V_h(T) = X.$$

We say that X is *hedgeable* if there exists an hedging portfolio for X.

Definition 0.29. A probability measure Q on $\mathcal{F}_T$ is called an *equivalent martingale measure* for the market model $\mathcal{M}$, the numeraire S_0 and the time interval $[0, T]$, if it has the following properties:

- Q is equivalent to P on $\mathcal{F}_T$.
- All price processes $Z^{(0)}, Z^{(1)}, \ldots, Z^{(N)}$ are martingales under Q on the time interval $[0, T]$.

Instead, we call Q a *local martingale measure* for the market if the discounted price processes $Z^{(i)}$ are Q-local martingales (and Q is equivalent to P on $\mathcal{F}_T$.)

Now we will state a really important theorem for Financial Mathematics in a "rough", but useful, version: even if this statement is not completely precise (but almost it is), it helps to undestand its deep meaning.

Theorem 0.30 (The First Foundamental Theorem, preliminary version)**.** *The market model $\mathcal{M}$ is arbitrage free if and only if there exists a martingale measure Q.*

So, the arbitrage possibilities are then deeply related to the martingale measures. Before passing to state precisely the First Foundamental Theorem, we need some more tecnical tools: first of all, we need to complete the connection between a market and his normalized economy:

Lemma 0.31 (Invariance Lemma)**.** *With Z_t defined as in Definition 0.23, define*

$$V_h^Z(t) = \sum_{i=0}^{N} h_i(t) Z_t^{(i)}.$$

The following hold:

- *a portfolio h is self-financing if and only if*

$$\mathrm{d}V_h^Z(t) = \sum_{i=0}^{N} h_i(t) \mathrm{d}Z_t^{(i)} \quad \textit{for all } t \in [0, T]$$

(that is h is Z-self-financing).

- *The value processes V_h and V_h^Z are connected by*

$$V_h^Z(t) = \frac{1}{S_t^{(0)}} \cdot V_h(t).$$

- *A claim X is hedgeable if and only if the claim $\frac{X}{S_t^{(0)}}$ is Z-hedgeable, that is there exists a self-financing portfolio h such that*

$$V_h^Z(T) = \frac{X}{S_t^{(0)}}.$$

- *The market model is arbitrage-free with respect to the S processes if and only if it is arbitrage-free with respect to the Z processes.*

The Invariance Lemma extabilishes then a natural correspondence between a S-market and the respective Z-market, and allows to do all the pricing theory under the assumption (that we will do in the following) $S_t^{(0)} \equiv 1$.

Lemma 0.32. *For any adapted process $h_S = [h_1, \dots, h_N]$ satisfying the admissibility condition 0.0.4, and for any real number x, there exists a unique adapted process h_0 such that*

- *the portfolio $h = [h_0, h_S]$ is self financing;*
- *its value process is given by*

$$V_h(t) = x + \int_0^t h(u)\mathrm{d}S_u.$$

By the previous Lemma we have that the set $\mathcal{K}_0$,

$$\mathcal{K}_0 = \left\{ \int_0^T h(u)\mathrm{d}S_u \middle| \ h \text{ is admissible} \right\},$$

is the space of portfolio values reachable at time T by means of a self-financin portfolio with zero initial value.

Now, define:

- $\mathcal{K} = \mathcal{K}_0 \cap \mathcal{L}^\infty(\Omega, \mathcal{F}_T, P)$.
- $\mathcal{C} = \mathcal{K} - \mathcal{L}^\infty_+(\Omega, \mathcal{F}_T, P)$,
  that is $v \in \mathcal{C} \Leftrightarrow v = v_{\mathcal{K}} + v_{L^\infty_+}$, where $v_{\mathcal{K}} \in \mathcal{K}$ and $v_{L^\infty_+} \in L^\infty_+(\Omega, \mathcal{F}_T, P)$.

Remark 0.33. We can now rewrite the arbitrage-free condition as following: a market model is arbitrage-free if $\mathcal{C} \cap \mathcal{L}^\infty_+(\Omega, \mathcal{F}_T, P) = \{0\}$.

Definition 0.34. A market model admits *No Free Lunch with Vanishing Risk* (NFLVR) if

$$\overline{\mathcal{C}} \cap \mathcal{L}^\infty_+(\Omega, \mathcal{F}_T, P) = \{0\},$$

where $\overline{\mathcal{C}}$ is the closure of $\mathcal{C}$ in $\mathcal{L}^\infty(\Omega, \mathcal{F}_T, P)$ with the weak topology.

If a market model is arbitrage-free but doesn't satisfies the (NFLVR) condition, then there will exist a nonzero claim $X \in L^{\infty}_{+}(\Omega, \mathcal{F}_T, P)$ and a sequence $X_n \in \mathcal{C}$ such that $|X - X_n| < 1/n$ for all n; thus there exists a self-financing portfolio with zero initial cost generating a claim which is closer than $1/n$ to the arbitrage claim X, while the risk is less than $1/n$. Even if this is not an arbtrage, it almost is.

This more general (and complicated) extension of the arbitrage concept is the right one to state precisely the First Foundamental Theorem:

Theorem 0.35 (The First Foundamental Theorem)**.** *Assume that the asset price process S is bounded (resp. locally bounded). Then the market model $\mathcal{M}$ admits No Free Lunch with Vanishing Risk if and only if there exists a martingale (resp. local martingale) measure Q.*

The First Foundamental Theorem is the building-block for the option pricing; in fact, assume now a market is arbitrage-free. Given a contingent claim X, two main approaches has been developed to give it a "fair" price process $\Pi_X(t)$:

- the first aprroach is to price the claim in such a way that the market remains arbitrage-free once we introduce these new prices in it; in other words, we want that the market $S^{(1)}, \dots, S^{(N)}, \Pi_X$ remains arbitrage-free.
- the second approach is to set $\Pi_X(t) = V_h(t)$, with h an hedging portfolio for X; this is completely natural, since for an investor it's identical to have X or h at time T.

Let's begin analyzing the first approach: the market $S^{(0)}, S^{(1)}, \dots, S^{(N)}$ is supposed to be arbitrage-free, then there exists a (not necessary unique) equivalent martingale measure Q. Thus, if we set

$$\Pi_X(t) = S^{(0)}_t \mathbb{E}^Q \left[\frac{X}{S^{(0)}_T} \middle| \mathcal{F}_t \right],$$

the discounted price process $\frac{\Pi_X}{S^{(0)}}$ automatically satisfies the martingale condition under Q. Again by The First Foundamental Theorem, we have then that the market $S^{(1)}, \dots, S^{(N)}, \Pi_X$ remains arbitrage free.

Theorem 0.36 (General Pricing Formula)**.** *The arbitrage-free price process for a given $T-$claim X is*

$$\Pi_X(t) = S^{(0)}_t \mathbb{E}^Q \left[\frac{X}{S^{(0)}_T} \middle| \mathcal{F}_t \right] \tag{0.0.8}$$

Note that, from one hand, this formula is really general; to work, in fact, it requires only that the market is arbitrage-free, that is a condition linked to the efficiency of the market and almost always assumed. On the other hand, it can be hard to use it. First of all, many times the dynamics for the prices are given under the measure P and to find an equivalent measure Q could be not easy; moreover, even if we have such dynamics, rarely the conditional expectation has a closed formula, even for easy claims such the call option.

Finally, it has to be noticed that if there exist more than one equivalent martingale measure Q, we have many possible prices $\Pi_X(t)$, one for each measure, and they are all prices that respect the arbitrage-free condition. Which one, then, is the right price for X?

Partially, an answer is given analyzing the second approach:

Proposition 0.37. *If a claim is hedgeable, then:*

- *all its hedging portfolio strategies have the same value at any time $t \leq T$;*
- *Using the General Pricing Formula, all the equivalent martingale measures give the same arbitrage-free price process $\Pi_X(t)$;*
- *if h is an hedging portfolio for X and Q is an equivalent martingale measure, then $V_h(t) = \Pi_X(t)$ for all $t \leq T$.*

What proposition 0.37 states is that, as long as a claim X is hedgeable, whatever of the two approaches we choose. we will find the same, unique, price. Moreover, we have now another way to find this price process: in fact, assume $\Pi_X(t) = \psi(t, S_t)$, with ψ smooth enough. Then, applying the Ito's Formula to the equation

$$V_h(t) = \Pi_X(t)$$

we obtain partial differential equation on ψ, that we can try and solve anlitically or, more often, numerically.

Note, however, that to use this approach one condition should be primarly satisfied: the claim has to be hedgeable, a condition that we didn't have for the General Pricing Formula. A particular case in which we can always apply this method are the *complete markets*:

Definition 0.38. A market is called *complete* if any $T-$claim is hedgeable.

Theorem 0.39 (The Second Foundamental Theorem)**.** *Assume that the market is arbitrage-free and consider a fixed numeraire asset $S^{(0)}$. Then, the market is complete if and only if the martingale measure Q, corresponding to the numeraire $S^{(0)}$, is unique.*

With the Second Foundamental Theorem we have then completely solved (at least theoretically) the option pricing problem in complete markets. In fact, we know that the price process is unique and we have two possible methods to find it. For uncomplete markets instead the problem is still open. Many solutions has been presented in literature, but this is still an open problem.

Market Models

In this section we will finally present some classical models for financial markets.

We start presenting the most famous financial model, Black & Scholes:

Example 0.40. In *Black & Scholes Model* the prices of riskly assets are represented by a geometric Brownian Motion, that is a Lévy process with only the drift and the diffusive part. The Stochastic differential equation for this process is

$$\mathrm{d}S_t = S_t\,(r\mathrm{d}t + \sigma \mathrm{d}W_t),$$

where r (the *drift*) and σ (the *volatility*) are both constant.

The main strenghtpoint of this model is its simplicity: in fact it is easy (both from a theoretical and pratical point of view) is fast to simulate and allows for some closed formulas when it comes to the option pricing.

We report here the celebrated Black & Scholes formula for a Call Option, that will be our benchmark in Chapter 4. Consider a market $\mathcal{M}$ composed of only two assets,

$$S^{(0)} \equiv 1$$

that it is the risk-free asset, and the riskly one that follows the Black & Scholes dynamics

$$\mathrm{d}S_1^{(1)} = S_t^{(1)} \sigma \mathrm{d}W_t.$$

In this case, the price of an European Call option $X = (S_T - K)^+$ is given by the formula

$$\Pi_t = S_t N(d_1) - K N(d_2),$$

where

$$d_1 = \frac{\ln \frac{S_t}{K} + \frac{1}{2}\sigma^2 (T-t)}{\sigma\sqrt{T-t}}, \quad d_2 = d_1 - \sigma\sqrt{T-t}$$

and N is the $N(0,1)$ cumulative distribution function.

Despite these appealing features, Black & Scholes model is not really versatile and there are many empirical evidences against it, as we will see in Chapter 1. Then, to overcome Black & Scholes limits, many more general models has been proposed, both in discrete and continuous time, such as:

- stochastic volatility models,
- more general Lévy models, with jumps in trajectories,
- generalized autoregressive conditional heteroskedasticity (GARCH) models.

All these models are more accurate than Black & Scholes in the description of empirical time series, but they pay this precision in term of a greater number of parameters to be estimated (and then a more difficult calibration) and with hard computational and theoretical issues. Moreover, as we will see in Chapter 1, they are not consinstent with some stylized facts of financial time series.

In the following we give a short review of these models.

Example 0.41. The *Heston model* is a stochastic volatility model that was presented in 1993, see [4]. In this model, the stock prices has the following dynamics:

$$\mathrm{d}S_t = S_t(r\mathrm{d}t + \sigma_t \mathrm{d}W_t^{(1)})$$

$$\mathrm{d}\sigma_t = k(\theta - \sigma_t)\mathrm{d}t + v\sigma_t \mathrm{d}W_t^{(2)},$$

with $W_t^{(1)}$ and $W_t^{(2)}$ are ρ-correlated Brownian Motion, that is

$$\mathbb{E}\left[W_t^{(1)} W_t^{(2)}\right] = \rho t.$$

This model does not allow for jumps, neither in the returns nor in the volatility, but more general (and complicate!) versions of it does.

Example 0.42. The *Variance-Gamma Model* (see [5] for references) is a Lévy provess where S_t has a Variance-Gamma law $VG(\sigma, \nu, \theta)$. It can be characterized as a time-changed Brownian Motion with drift:

$$S_t = \theta\gamma_t + \sigma W_{\gamma_t},$$

where W is a Brownian Motion and γ_t is a Gamma process, with parameters $a = \frac{1}{\nu}$ and $\beta = \frac{1}{\nu}$, independent of W. The Variance-Gamma process is a pure jump, finite variation process and is the difference between two increasing Lévy process:

$$S_t = G_t^{(1)} - G_t^{(2)},$$

where the $G^{(i)}$ are Gamma process with parameters $a = \frac{1}{\nu}$, $\beta = \nu_i$ such that:

$$\nu_1 = \frac{2}{\theta\nu + \sqrt{\theta^2\nu^2 + 2\nu\sigma^2}} > 0$$

$$\nu_2 = \frac{2}{-\theta\nu + \sqrt{\theta^2\nu^2 + 2\nu\sigma^2}} > 0.$$

The Variance-Gamma process has infinite activity and in the risk-neutral world, stock prices driven by a Variance-Gamma process have dynamics

$$Z_t = Z_0 \exp\left(rt + S_t + \frac{t}{\nu}\ln\left(1 - \theta\nu - \frac{\sigma^2\nu}{2}\right)\right);$$

the parameters ν and σ control skewness and kurthosis.

Example 0.43. The *tempered stable process* is a Lévy process introduced by Koponen and used by Bouchaud and Potters for financial modelling. It's a process without Gaussian component and with Lévy density

$$\frac{C_+}{x^{1+Y_+}}e^{-xM_+}\mathbf{1}_{x>0} + \frac{C_-}{|x|^{1+Y_-}}e^{-xM_-}\mathbf{1}_{x<0}$$

where $C_\pm > 0$, $M_\pm \geq 0$ and $Y_\pm < 2$. A particular case of this model is the *CGMY model*, introduced by Carr & others: in this case $C_+ = C_- = C$, $M_+ = M_- = M$, $Y_+ = Y_- = Y$ and the Lévy density of the process is

$$\frac{C}{x^{1+Y}}e^{-xM}\mathbf{1}_{x>0} + \frac{C}{|x|^{1+Y}}e^{-xM}\mathbf{1}_{x<0}$$

with $C > 0$, $M \geq 0$ and $Y < 2$. If $Y < 0$, there is a finite number of jumps in any finite interval; if not, the process has infinite activity. If $Y \in [1, 2[$, the process has infinite variation.

Example 0.44. The *Autoregressive Conditional Heteroskedasticity (ARCH)* models were introduced by Robert Engle in 1982 (see [7]); they are models in which the actual return of an asset is the sum of the mean return m_n (given by past information) plus an error term ϵ_n that depends on the previous q error terms. In other words, the ARCH(q) model is the following:

$$X_n = m_n + \epsilon_n$$

$$\epsilon_n = \sigma_n z_n$$
$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{n-i}^2,$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i = 1, \ldots, q$ and z_n are $N(0,1)$ independent random variables. An extension of the ARCH models proposed by Bollerslev are the $GARCH(p,q)$ models, in which the error terms depends in a more complicated way on the previous error terms; A GARCH(p,q) specification is

$$X_n = m_n + \epsilon_n$$
$$\epsilon_n = \sigma_n z_n$$
$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{n-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{n-i}^2,$$

where $\alpha_0 > 0$, $\alpha_i, \beta_j \geq 0$ for $i = 1, \ldots, q$, $j = 1, \ldots, p$ and z_n are $N(0,1)$ independent random variables. Due to simplicity reasons, the most widely used GARCH models are the GARCH$(1,1)$. Anyway, many variants has been introduced, such as GARCH models where the z_n are non gaussian random variables, for example we have GARCH models with tempered-stable innovations and many others.

Finally, we conclude this chapter presenting a continuous version of GARCH model, called the COGARCH model. The reason behind such an extension is twofold: on one hand, due to weekend and holiday effects, or in tickby- tick data, as well to other reasons, many financial time series are irregularly spaced; on the other, option pricing is mainly done in a continuous contest.

Example 0.45. The *COGARCH model*, proposed by Klüppelberg, Lindner and Maller, is an attempt to extend the GARCH methods to continuous time, see [21] and [22].

The one-dimensional COGARCH$(1,1)$ process is given as the solution of

$$\mathrm{d}G_t = \sqrt{\sigma_{t^-}}\mathrm{d}L_t,$$
$$\mathrm{d}\sigma_t = -\beta(\sigma_{t^-} - c)\mathrm{d}t + \alpha\sigma_{t^-}\mathrm{d}[L,L]_t^d,$$

with $[L,L]_t^d$ the discrete part of the quadratic variation of a univariate Lévy process L, $\alpha, \beta, c > 0$ and initial values $G_0 = 0$, $\sigma_0 \geq 0$. For this process, we have that an unique Lévy process drives both the returns and the volatility, which is the basic idea of GARCH models; in this sense, COGARCH is a continuous version of GARCH.

Chapter 1

The Model

The basic example of financial model for the prices of stocks or indices is Black & Scholes, that is used as a first benchmark in data anlysis. In this chapter we present the main statistichal features of financial time series (we will refer in particular to the DJIA index, daily open prices from 1935 to 2010) that are not well-fitted by the Black & Scholes model (and many other more complex models too).

In the following, if not differentely specified, whenever we will consider a time serie $\hat{X}$ or a stochastic process X modelling a time serie, it will be representing the logarithmes of the prices and not the prices themselves. We will consider both empirical time series $\hat{X} = (\hat{x}_t)_{0\leq t\leq T}$ (necessarly at discrete time) and theoretical models $X = (x_t)_{0\leq t\leq T}$, represented by (discrete or continuous) stochastic processes. Given an empirical (or simulated) time series $\hat{X}$, we define its empirical distribution over h days,

$$\hat{p}_h(\cdot) := \frac{1}{T-h}\sum_{t=1}^{T-h} \delta_{\hat{x}_{t+h}-\hat{x}_t}(\cdot)\,,$$

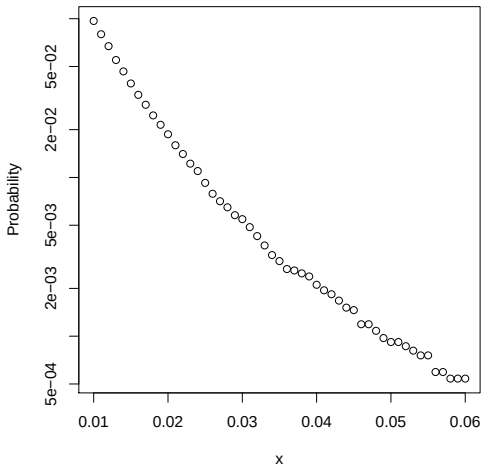
where $\delta_{\hat{x}}(\cdot)$ denotes the Dirac measure at $\hat{x} \in \mathbb{R}$.

In particular, we will use the symbol $\hat{\cdot}$ for any (empirical) probabilistic quantity computed under this empirical distribution, so $\hat{\mathbb{E}}$ will denote the empirical mean for the time series, $\widehat{Var}$ the empirical variance and so on.

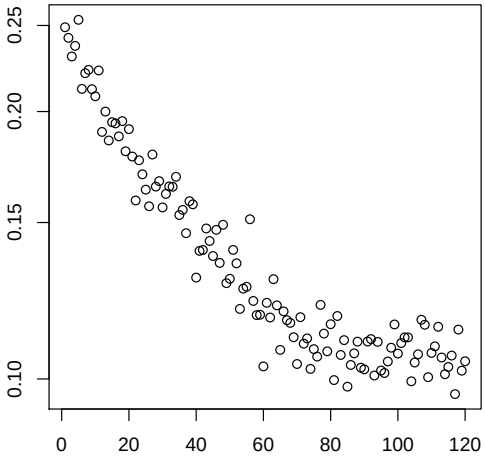
1.1 Stylized Facts

We begin this chapter summarizing some of the most remarkable stylized facts for financial time series:

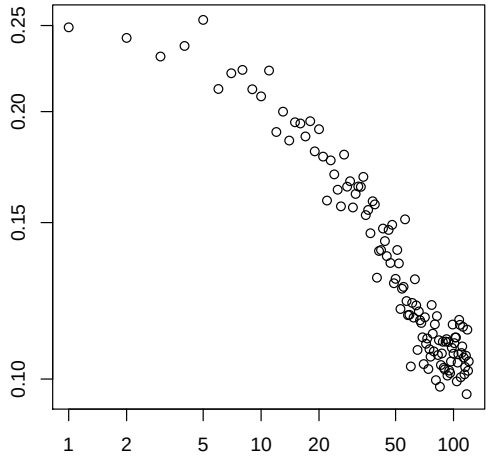
- *Heavy tails.* While the central part of the empirical distribution of log-returns is nearly Gaussian, the tails are remarkably fatter than in the Gaussian case, as we can see in Figure 1.1(A); the tails of the distribution represent rare events on the time series and are in particular related to *crisis*. A loss smaller or equal to 11% in log-return (20^{th} October 1987, the biggest loss in Dow Jones history from 1936 until now) would have a probability of about 10^{-33} in a Black & Scholes setting. Even less extreme events,



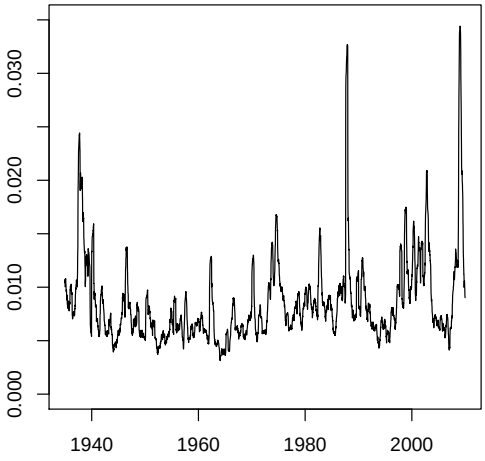
(A) *Right tail of daily log-return distribution.*



(B) *Volatility autocorrelation: log plot.*



(C) *Volatility autocorrelation: log-log plot.*



(D) *Empirical volatility.*

FIGURE 1.1: *Stylized facts for the DJIA time series (1935-2009).*
(A) Log plot of the right tail distribution for the daily log returns. Standard deviation $\simeq 0.01$.
(B) Log plot of the empirical 1-day volatility autocorrelation for the DJIA time series (1935-2009).
(C) Same as (B), but log-log plot instead of log plot.
(D) Empirical volatility for the daily log-returns of the DJIA time series in the period 1935-2009.
Note: The volatility at day i is evaluated as $\left[\frac{1}{2R+1}\sum_{j=i-R}^{i+R}(x_j - x_{j-1})^2\right]^{1/2}$, with $R = 50$.

such as losses of more than 5% in the log-return (there have been 17 of these crisis in about 80 years), would have a probability of 10^{-7} in the Gaussian case, while their empirical probability is about $\frac{1}{1000}$.

- *Clustering of volatility.* While the empirical log-returns are uncorrelated, they are not independent. In particular, denote with $\rho(X,Y) = \frac{Cov[X,Y]}{\sqrt{Var[X]\,Var[Y]}}$ the correlation coefficient between X and Y and, given $\hat{X}$, consider the quantity

$$\hat{\rho}_{h,t-s=}\hat{\rho}(|\hat{x}_{t+h} - \hat{x}_t|, |\hat{x}_{s+h} - \hat{x}_s|),$$

that is called the *empirical volatility autocorrelation.* It turns out that $\hat{\rho}_{h,t-s}$ has a slow decay on $|t-s|$, at least for modetate values of it. In Figure 1.1(C) we can see that for the Dow Jones the autocorrelation decay is faster than polynomial, while in Figure 1.1(B) we can see that it is slower than exponential.

- *Volatility peaks* Far from being constant, the volatility $\hat{\sigma}$ of many empirical time series has high peaks and in general present a great variability, as we can see in Figure 1.1(D)

These facts are very well known and a lot of literature can be found about. In particular, many models have been proposed to take them into account; as an example, we do a short analysis for the $GARCH(1,1)$ model. We fitted the $GARCH(1,1)$ model (see 0.44) to the DJIA time series (1935-2009) with the "garch()" function of the software "R" [25], package "tseries". In Figure 1.2 we can see that:

- the emprical distribution is reasonably fitted, both in the bulk (see Figure 1.2(B)) and int the tails (see Figure 1.2(A));
- the volatility autocorrelation is actually positive, even if its decay is *purely exponential*, while the one of the DJIA is slower, (see Figure 1.2(C));
- The local volatility of the $GARCH(1,1)$ model is qualitatively similar to the one of the DJIA time series, but *no peaks* are present (see Figure 1.2(D)).

1.2 Less-Standard Stylized Facts

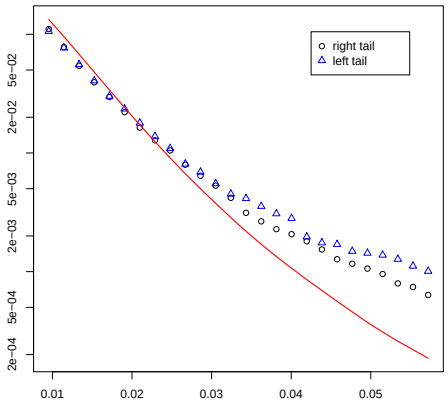
Beside these well known problems, more recently some remarkable scaling properties has been emphasized in papers such as [14],[9],[26]. These properties are:

- *Diffusive scaling of log-returns.* At least for h within a suitable time scale (that is about from 1 to 50 days for the DJIA, see Figure 1.3(A)), $\hat{p}_h$ obeys approximately this scaling relation:

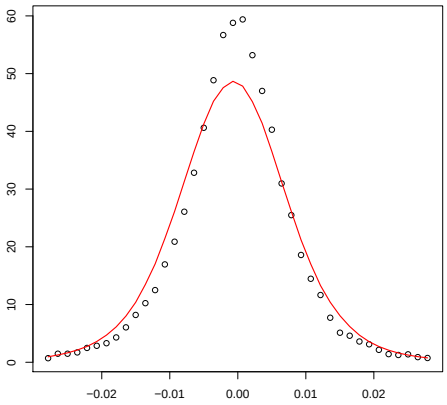
$$\hat{p}_h(\mathrm{d}r) \simeq \frac{1}{\sqrt{h}}\, g\left(\frac{r}{\sqrt{h}}\right)\mathrm{d}r, \tag{1.2.1}$$

where g is a probability density with tails heavier than Gaussian.

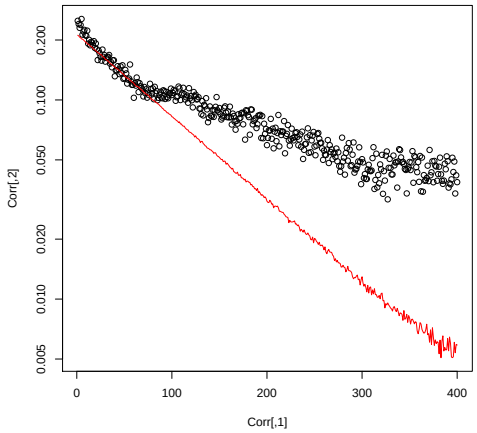
- *Multiscaling of the moments.* If one considers the q-th empirical moment $\hat{m}_q(h)$, from relation (1.2.1) it would be natural to guess that $\hat{m}_q(h)$ should scale as $h^{q/2}$, and this is



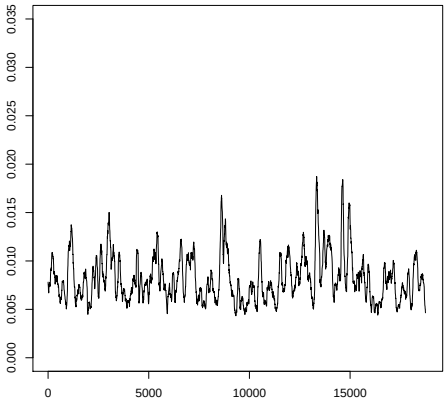
(A) *Right tail of daily log-return distribution.*



(B) *Bulk of daily log-return distribution.*

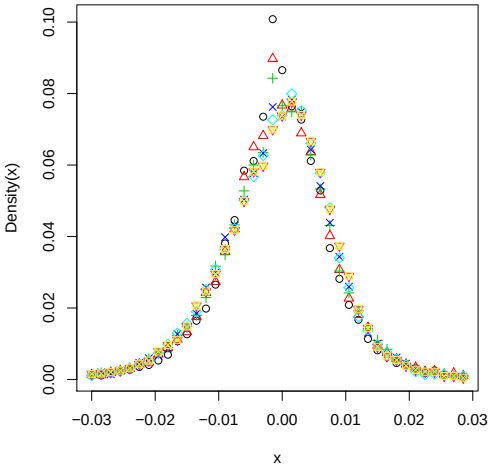


(C) *Volatility autocorrelation: log plot.*

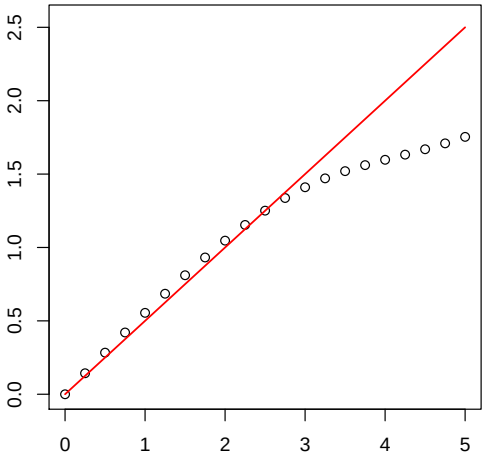


(D) *Local volatility for the* $GARCH(1,1)$.

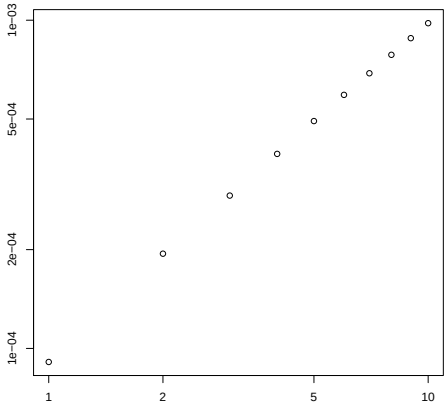
FIGURE 1.2: *A comparison between the DJIA time series and the* $GARCH(1,1)$ *model.*
(A) Log plot of the right tail distribution of the daily log returns of the DJIA (dots, see legend) and the theoretical tail distribution of the $GARCH(1,1)$ model (line). Standard deviation $\simeq 0.01$.
(B) Log plot of the bulk of the distribution of the daily log returns of the DJIA (dots) and the bulk of the theoretical distribution for the $GARCH(1,1)$ model (line). Standard deviation $\simeq 0.01$.
(C) Log plot of the empirical 1-day volatility autocorrelation for the DJIA time series (dots) and the theoretical 1-day volatility autocorrelation for the $GARCH(1,1)$ model (line).
(D) Empirical volatility for a simulation of 17789 data of the $GARCH(1,1)$ model.
Note: For the DJIA time series, we considered the daily opening prices from 2 Jan 1935 to 31 Dec 2009. For the theoretical quantities of the $GARCH(1,1)$ model we mean the empirical ones computed for a simulation of 5 million data.



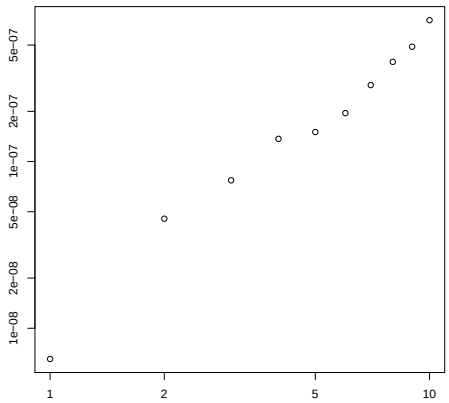
(A) *Rescaled empirical density of log-returns.*



(B) *Multiscaling of the moments.*



(C) *Empirical* 2^{nd} *moment* $\hat{m}_2(h)$.



(D) *Empirical* 6^{th} *moment* $\hat{m}_6(h)$.

FIGURE 1.3: *Less-standard stylized facts for the DJIA time series (1935-2009).*
(A) Plot of the rescaled empirical density of the h- log returns of the DJIA time series (see legend). Standard deviation $\simeq 0.01$.
(B) The empirical scaling exponent $\widehat{A}(q)$ in the period 1935-2009 (circles) as a function of q, and the straight line $\frac{q}{2}$ (red line).
(C) Log-log plot of the empirical 2^{nd} moment $\hat{m}_2(h)$ as a function of h, with h between 1 and 10.
(D) Log-log plot of the empirical 6^{th} moment $\hat{m}_5(h)$ as a function of h, with h between 1 and 10.

indeed what one observes for moments of small order. However, for moments of higher order, the different scaling relation $h^{A(q)}$, with $\hat{A}(q) < q/2$, takes place, as we can see in Figure 1.3(B)

Remark 1.1. While the diffusive scaling is really evident, it is much more subtle to detect the multiscaling of the moments. Consider the DJIA: for low moments (let say until about the 3^{th} moment) the relation $\hat{m}_q(h) \simeq h^{A(q)}$ is quite evidente for h between 1 and 10 days (see Figure 1.3(C)); however, when we consider higher moments (for example the 6^{th} moment) this relation is no more so clear, as we can see in Figure 1.3(D). Approximate linearity in the log-log plot of $\hat{m}_q(h)$ is still present, at least for small h, but one has to make a choice for the scaling esponent $A(q)$, because of a rather strong uncertainty. After some tries, we have chosen to set $A(q)$ as the best linear interpolation in the log-log plot for the points

$$\{(1, \hat{m}_q(1)), (2, \hat{m}_q(2)), (3, \hat{m}_q(3)), (4, \hat{m}_q(4)), (5, \hat{m}_q(5))\}.$$

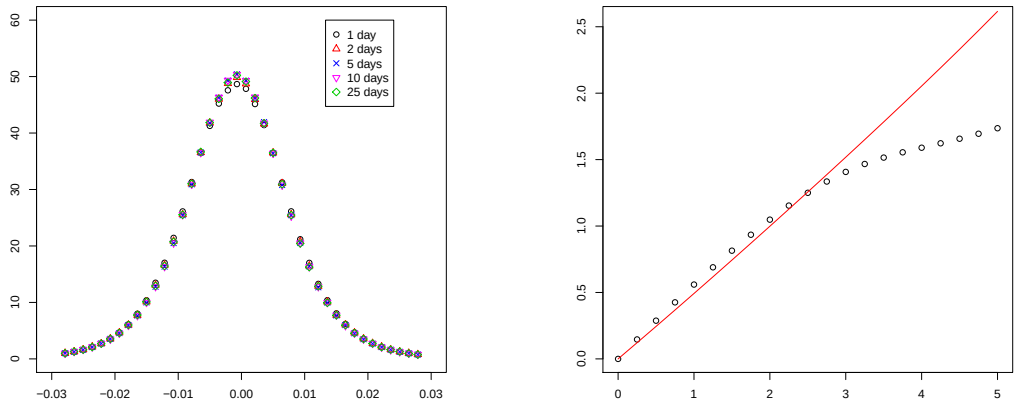
This choice seems reasonable for the following reasons:

- on one hand, the scaling is expected to hold only for h small, so higher values of h can be disregarded;
- on the other hand, simply consider the minimal case $h = 1$ seems too subject to randomness.

Anyway, slighty different choices do not change the scaling exponent too much, at least until the 5^{th}-moment. Moments of higher order are too dependent from rare events, and may fail to obey scaling.

The purpose of this thesis is to define a *simple* continuous-time stochastic model which agrees with *all* mentioned stylized facts. This is a non-trivial point, despite of the variety of models that can be found in the literature. For example, as we have seen the GARCH exhibits non-constant volatility (but not peaks!) and non-Gaussian distribution of log-returns, but the correlation of the absolute values of the log-return, although positive, decays exponentially fast, in contrast with empirical evidences. Moreover, while this model exhibits a diffusive scaling behaviour (see Figure 1.4(A)), the multiscaling of moments is not present, at least for the range of values of the parameters that most often occur in practice (see Figure 1.4(B)). We have also tested the model very recently proposed in [12]: this model is extremely accurate to fit the statistics of the empirical volatility and it does exhibit multiscaling of moments; however, the decay of correlations of absolute log-returns appears to be again purely exponential. Besides those specific aspects, we mention two features of this and related models, that should be compared with what we propose in this thesis.

- The model in [12], in its simplest version, requires the calibration of more than 30 parameters; in the model we propose, we calibrate 4 parameters. Clearly, we do not reach the same accuracy in some respect. However, the fact that for some fundamental features, such as clustering of volatility, we obtain an excellent fit to data, suggests that our model can be considered as a starting point for more refined generalizations.
- In [12], long memory effects are caused by including a suitable dependence on the past in the autoregressive equation for the volatility, while large price variations are



(A) *Diffusive scaling in the GARCH model.* (B) *No multiscaling in the GARCH.*

FIGURE 1.4: *Less-standard stylized facts for the* $GARCH(1,1)$ *model.*
(A) plot of the rescaled theoretical density for the h-log returns of the $GARCH(1,1)$ model (see legend).
(B) The empirical scaling exponent $\widehat{A}(q)$ for the DJIA in the period 1935-2009 (circles) and the theoretical scaling exponent $A(q)$ for the $GARCH(1,1)$ model (line).
Note. For the theoretical quantities of the $GARCH(1,1)$ model we mean the empirical ones computed for a simulation of 5 million data.

controlled by specific features of the driving noise. Here we propose a *single* simple mechanism, modeling the reaction of the market to shocks, which is the source of *all* features mentioned above: non-Gaussian distribution of log-returns, multiscaling of moments and clustering of volatility (with the correct decay).

Since our model is strongly inspired by the work of Baldovin and Stella, we begin with a short introduction on their approach.

1.3 Baldovin & Stella's Approach

Motivated by renormalization group arguments from statistical physics, Baldovin and Stella [9, 26] have recently proposed a new model in which scaling considerations play a major role. Their construction may be summarized as follows. They first introduce a process $(Y_t)_{t\geq 0}$ which satisfies the scaling relation (1.2.1) for a given function g, that is assumed to be even, so that its Fourier transform $\hat{g}(u) := \int_{\mathbb{R}} e^{iux} g(x)\mathrm{d}x$ is real (and even). The process $(Y_t)_{t\geq 0}$ is defined by specifying its finite dimensional laws: for $t_1 < t_2 < \cdots < t_n$ the joint density of $Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n}$ is given by

$$p(x_1,t_1;x_2,t_2;\ldots;x_n,t_n) = h\left(\frac{x_1}{\sqrt{t_1}}, \frac{x_2-x_1}{\sqrt{t_2-t_1}}, \ldots, \frac{x_n-x_{n-1}}{\sqrt{t_n-t_{n-1}}}\right), \tag{1.3.1}$$

where h is the function whose Fourier transform $\hat{h}$ is given by

$$\hat{h}(u_1, u_2, \ldots, u_n) := \hat{g}\Big(\sqrt{u_1^2 + \ldots + u_n^2}\Big). \tag{1.3.2}$$

Note that if g is the standard Gaussian density, then $(Y_t)_{t\geq 0}$ is the ordinary Brownian motion. For a non Gaussian g, the expression in (1.3.2) is not necessarily the Fourier transform of a probability on $\mathbb{R}^n$, so that some care is needed. We will see later for exactly what g's relations (1.3.1)-(1.3.2) define a consistent family of finite dimensional distributions, hence a stochastic process $(Y_t)_{t\geq 0}$, and it turns out that $(Y_t)_{t\geq 0}$ is always a *mixture* of Brownian motions. However, it is already clear from (1.3.1) that the increments of the process $(Y_t)_{t\geq 0}$ corresponding to time intervals of the same length (that is, for fixed $t_{i+1} - t_i$) have a *permutation invariant distribution* and therefore cannot exhibit any decay of correlations.

For this reason, Baldovin and Stella introduce what we believe is the most remarkable ingredient of their construction, namely a special form of time-inhomogeneity in the process $(Y_t)_{t\geq 0}$. Although they define it in terms of finite dimensional distributions, we find it simpler to give an equivalent pathwise construction. Given a sequence of times $0 < \tau_1 < \tau_2 < \cdots < \tau_n \uparrow +\infty$ and a fixed $0 < D \leq 1/2$, they introduce a new process $(X_t)_{t\geq 0}$ defined by

$$X_t := Y_{t^{2D}} \qquad \text{for } t \in [0, \tau_1), \tag{1.3.3}$$

and more generally

$$X_t := Y_{(t-\tau_n)^{2D} + \sum_{k=1}^{n} (\tau_k - \tau_{k-1})^{2D}} \qquad \text{for } t \in [\tau_n, \tau_{n+1})\,. \tag{1.3.4}$$

For $D = 1/2$ we have clearly $X_t \equiv Y_t$. On the other hand, for $D < 1/2$, $(X_t)_{t\geq 0}$ is obtained from $(Y_t)_{t\geq 0}$ by a nonlinear time-change, that is "refreshed" at each time τ_n. This transformation has the effect of amplifying the increments of the process for t immediately after the times $(\tau_n)_{n\geq 1}$, while the increments tend to become small for larger t.

In their numerical study Baldovin and Stella, besides considering the periodic case $\tau_n := nt_0$ for t_0 fixed, do not really simulate the process $(X_t)_{t\geq 0}$ but rather an autoregressive version of it. More precisely, given a small time step δ and a natural number N, they first simulate x_δ, $x_{2\delta}$, …, $x_{N\delta}$ according to the distribution of $(X_\delta, X_{2\delta}, \ldots, X_{N\delta})$. Then they compute the conditional distribution of $X_{(N+1)\delta}$ given $X_{2\delta} = x_{2\delta}$, $X_{3\delta} = x_{3\delta}$, …, $X_{N\delta} = x_{N\delta}$ — thus neglecting x_δ — and sample $x_{(N+1)\delta}$ from this distribution. Similarly, $x_{(N+2)\delta}$ is sampled from the conditional distribution of $X_{(N+2)\delta}$ given $X_{3\delta} = x_{3\delta}$, …, $X_{N\delta} = x_{N\delta}$, $X_{(N+1)\delta} = x_{(N+1)\delta}$, neglecting x_δ and $x_{2\delta}$, and so on; this autoregressive mechanism is the cause of the autocorrelation decay they found with simulations of their model.

After estimating the function g and the parameters t_0 and D on the DJIA time series, their simulated trajectories show a quite convincing agreement with both the basic scaling (1.2.1) and the multiscaling of moments, as well as with the clustering of volatility.

1.3.1 Some Comments on Baldovin & Stella Construction

Now we want to give a characterization for the class of functions g for which (1.3.1)–(1.3.2) provide a consistent family of finite dimensional distributions, with some more comments on the numerical results obtained by Baldovin & Stella. To do this, we first need some theory on exchangeable processes.

Definition 1.2. Set $\Delta Y_{(a,b)} := Y_b - Y_a$ for short. A stochastic process $(Y_t)_{t\geq 0}$ is said to have *exchangeable increments* if the distribution of the random vector

$$(\Delta Y_{I_1+y_1}, \ldots, \Delta Y_{I_n+y_n}),$$

where the I_j's are intervals and y_j's real numbers, does not depend on $y_1, \ldots, y_n$, as long as the intervals $y_1 + I_1, \ldots, y_n + I_n$ are disjoint.

Theorem 1.3. *If Y is a real-valued process with exchangeable increments and no fixed points of discontinuity, then Y is a mixture of processes with stationary, independent increments and no fixed points of discontinuity.*

For the proof of this theorem, see Theorem 3 in [16] (cf. also [1])

Definition 1.4. The real-valued random variables $\{X_1, ..., X_n\}$ will be called *isotropic* if their joint distribution has spherical symmetry (in n-dimensions).

Theorem 1.5. *If Y is a real-valued process with exchangeable, isotropic increments and no fixed points of discontinuity, then Y is a mixture of Brownian motions.*

For the proof of this theorem, see Theorem 4 in [16]

If a stochastic process $(Y_t)_{t\geq 0}$ is to satisfy relations (1.3.1)–(1.3.2), it must necessarily have exchangeable increments so, if we make the (very mild) assumption that $(Y_t)_{t\geq 0}$ has no fixed point of discontinuity, then Theorem 1.3 ensures that $(Y_t)_{t\geq 0}$ is a mixture of Lévy processes. Moreover, since by (1.2.1) the distribution of the increments of $(Y_t)_{t\geq 0}$ is *isotropic*, by Theorem 1.5 the process $(Y_t)_{t\geq 0}$ is a mixture of Brownian motions.

This means that we have the following representation for it:

$$Y_t = \sigma\, W_t, \tag{1.3.5}$$

where $(W_t)_{t\geq 0}$ is a standard Brownian motion and σ is a random variable (interpreted as a *random volatility*) taking values in $(0, \infty)$ and independent of $(W_t)_{t\geq 0}$. Viceversa, if a process $(Y_t)_{t\geq 0}$ satisfies (1.3.5),then, denoting by ν the law of σ, relations (1.3.1)–(1.3.2) hold with

$$g(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{x^2}{2\sigma^2}}\, \nu(\mathrm{d}\sigma)\,, \tag{1.3.6}$$

or, equivalently,

$$\hat{g}(u) = \int_{\mathbb{R}} e^{-\frac{\sigma^2 u^2}{2}}\, \nu(\mathrm{d}\sigma)\,,$$

as one easily checks. This shows that the functions g for which (1.3.1)–(1.3.2) provide a consistent family of finite dimensional distributions are exactly those that may be expressed as in (1.3.6) for some probability ν on $(0, +\infty)$.

Remark 1.6. A sample path of (1.3.5) is obtained by sampling independently σ from ν and $(W_t)_{t\geq 0}$ from the Wiener measure. Note that this path *cannot* be distinguished from the path of a Brownian motion with *constant* volatility, and therefore correlations of increment cannot be detected empirically, as we already mentioned. The same observation applies to the time-inhomogeneous process $(X_t)_{t\geq 0}$ obtained by $(Y_t)_{t\geq 0}$ through (1.3.3)–(1.3.4). The reason why Baldovin and Stella measure nonzero correlations from their samples lays only on the autoregressive mechanism they use to approximate their process.

1.4 Definition of the Model

The basic idea of this thesis is to define a simple continuous-time stochastic model $X = (X_t)_{t\geq 0}$ for the (detrended) log-prices of indexes which, on one hand, should captures the essential features of the model proposed by Baldovin and Stella and, on the other hand, should be easier to describe, to interpret and to simulate.

We propose a different mechanism, that replaces the autoregressive scheme by a random update of the volatility. Given two real numbers $D \in (0, 1/2]$, $\lambda \in (0, \infty)$ and a probability ν on $(0, \infty)$ (these may be viewed as our general parameters), our model is defined upon the following three sources of alea:

- a standard Brownian motion $W = (W_t)_{t\geq 0}$;
- a Poisson point process $\mathcal{T} = (\tau_n)_{n\in\mathbb{Z}}$ on $\mathbb{R}$ with intensity λ;
- a sequence $\Sigma = (\sigma_n)_{n\geq 0}$ of independent and identically distributed positive random variables. The marginal law of the sequence will be denoted by ν (so that $\sigma_n \sim \nu$ for all n) and for conciseness we denote by σ a variable with the same law ν.

We assume that $W, \mathcal{T}, \Sigma$ are defined on some probability space $(\Omega, \mathcal{F}, P)$ and that they are independent. By convention, we label the points of $\mathcal{T}$ so that $\tau_0 < 0 < \tau_1$. We will actually need only the points of $\mathcal{T} \cap [\tau_0, \infty)$, that is the variables $(\tau_n)_{n\geq 0}$. We recall that the random variables $-\tau_0$, τ_1, $(\tau_{n+1} - \tau_n)_{n\geq 1}$ are independent and identically distributed $Exp(\lambda)$, so that $1/\lambda$ is the mean distance between the points in $\mathcal{T}$. Although some of our results would hold for more general distributions of $\mathcal{T}$, we stick for simplicity to the (rather natural) choice of a Poisson process.

Moreover, we consider a more general time rescaling then the one proposed by Baldovin & Stella; in fact, our time rescaling function will be a map $\theta : [0, +\infty) \to [0, +\infty)$ that satisfies the following properties:

(1) θ is concave, strictly increasing, $\theta(0) = 0$, and $\theta(s) \to +\infty$ as $s \to +\infty$.

(2) θ is $\mathcal{C}^1$ on $(0, +\infty)$, and there exists a constant $0 < D < \frac{1}{2}$ such that

$$\lim_{s\downarrow 0} \frac{\theta'(s)}{2Ds^{2D-1}} = 1.$$

In particular then, $\theta(s)$ behaves as s^{2D} near $s = 0$.

Later on this thesis, for numerical applications of our model (see Chapter 3), we will make the following choices for θ :

- $\theta(t) = t^{2D}$
- $\theta(t) := \begin{cases} t^{2D} & \text{if } 0 < t < C \\ C^{2D} + 2\,D\,C^{2D-1}\,(t - C) & \text{if } t \geq C \end{cases}$,

with C a scalar parameter.

We are now ready to define our model $X = (X_t)_{t\geq 0}$. For $t \in [0, \tau_1]$ we set

$$X_t := \sigma_0 \big(W_{\theta(\lambda(t-\tau_0))} - W_{\theta(-\lambda\tau_0)} \big) , \tag{1.4.1}$$

while for $t \in [\tau_n, \tau_{n+1}]$ (with $n \geq 1$) we set

$$X_t := X_{\tau_n} + \sigma_n \left(W_{\theta(\lambda(t-\tau_n))+\sum_{k=1}^n \theta(\lambda(\tau_k-\tau_{k-1}))} - W_{\sum_{k=1}^n \theta(\lambda(\tau_k-\tau_{k-1}))} \right) . \tag{1.4.2}$$

In words: at the epochs τ_n the time inhomogeneity $t \mapsto \theta(\lambda t)$ is "refreshed" and the volatility is randomly updated: $\sigma_{n-1} \;;\; \sigma_n$. A possible financial interpretation of this mechanism is that jumps in the volatility correspond to shocks in the market. The reaction of the market is not homogeneous in time: if $D < 1/2$, the dynamics is fast immediately after the shock, and tends to slow down later, until a new jump occurs. For $\theta(\lambda t) = \lambda t$ (and then $D = 1/2$) our model reduces to a simple random volatility model $\mathrm{d}X_t = \sigma_t \, \mathrm{d}W_t$, where $\sigma_t := \sum_{k=0}^{\infty} \sigma_k \, \mathbf{1}_{[\tau_k, \tau_{k+1})}(t)$ is a (random) piecewise constant process.

Note that the pieces of the Brownian motion W used in each interval $[\tau_n, \tau_{n+1})$ (cf. (1.4.2)) are independent. Also observe that in (1.4.1) the instant $t = 0$ is not the beginning of the time inhomogeneity, as it was in (1.3.3): this is to ensure that the process X has stationary increments (see below).

Using the scale invariance of Brownian motion, we now give an alternative definition of our model X, that is equivalent in law with (1.4.2) but more convenient for the proofs in the following sections. For $t \geq 0$, define

$$i(t) \;:=\; \sup\{n \geq 0 : \, \tau_n \leq t\} \;=\; \#\{\mathcal{T} \cap (0, t]\} , \tag{1.4.3}$$

so that $\tau_{i(t)}$ is the location of the last point in $\mathcal{T}$ before t. Now we introduce the process $I = (I_t)_{t\geq 0}$ by

$$I_t \;:=\; \sigma_{i(t)}^2 \theta \left(\lambda(t - \tau_{i(t)}) \right) \;+\; \sum_{k=1}^{i(t)} \sigma_{k-1}^2 \theta \left(\lambda(\tau_k - \tau_{k-1}) \right) \;-\; \sigma_0^2 \theta \left(-\lambda\tau_0 \right) , \tag{1.4.4}$$

with the agreement that the sum in the right hand side is zero if $i(t) = 0$. In Figure 1.5 we plot a typical trajectory of I_t, where we can note that I_t has singularities at the points τ_i; these are clearly due to the singularity in 0 of the function θ.

We can then redefine our basic process $X = (X_t)_{t\geq 0}$ by setting

$$X_t := W_{I_t} . \tag{1.4.5}$$

Note that I is a strictly increasing process with absolutely continuous paths, and it is independent of the Brownian motion W; thus, our model may be viewed as an independent random time change of a Brownian motion. What is particularly interesting is that for this model all the stylized facts mentioned above can be rigorously proved (see Chapter 2) and that its computational simplicity allows, at least at a numerical level, to use this model as an alternative to Black & Scholes and related models in applications to finance, such as pricing of options (see Chapter 4).

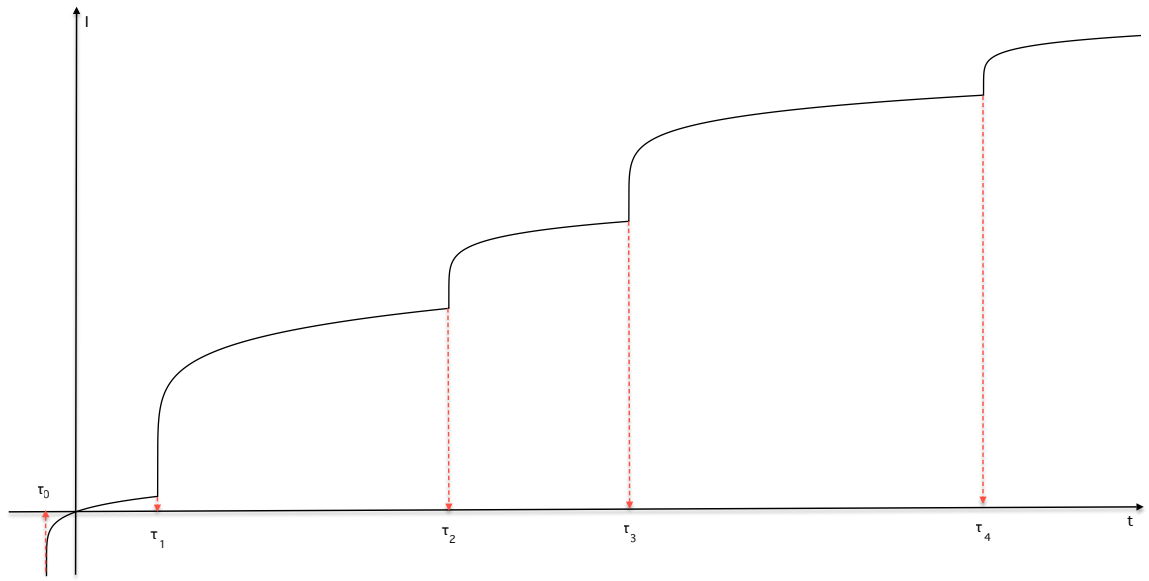


FIGURE 1.5: *A typical trajectory for* I_t.

1.5 Basic Properties

Some basic properties of our model can be easily established.

(A) *The process* X *has stationary increments.*

(B) *The process* X *can be represented as a stochastic volatility process*:

$$\mathrm{d}X_t = v_t \, \mathrm{d}B_t \,, \tag{1.5.1}$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion. More precisely, denoting $I'(s) := \frac{\mathrm{d}}{\mathrm{d}s} I(s)$, the variables B_t and v_t are defined by

$$B_t := \int_0^{I_t} \frac{1}{\sqrt{I'(I^{-1}(u))}} \, \mathrm{d}W_u \,, \qquad v_t := \sqrt{I'(t)} = \sigma_{i(t)} \sqrt{\lambda \, \sigma' \left(\lambda (t - \tau_{i(t)}) \right)} \,. \tag{1.5.2}$$

Note that, whenever $D < \frac{1}{2}$, the volatility v_t has singularities at the random times τ_n due to the singularity in 0 of the function θ.

(C) *The process* X *is a zero-mean, square-integrable martingale* (provided $\mathbb{E}[\sigma^2] < \infty$).

We now prove these three properties; in the following, we denote by $\mathcal{G}$ the σ-field generated by the whole process $(I_t)_{t\geq 0}$, which coincides with the σ-field generated by the sequences $\mathcal{T} = \{\tau_k\}_{k\geq 0}$ and $\Sigma = \{\sigma_k\}_{k\geq 0}$.

Proof of property (A). We first focus on the process $(I_t)_{t\geq 0}$, defined in (1.4.4). For $h > 0$ let $\mathcal{T}^h := \mathcal{T} - h$ and denote the points in $\mathcal{T}^h$ by $\tau_k^h = \tau_k - h$. As before, let $\tau^h_{i^h(t)}$ be the largest point of $\mathcal{T}^h$ smaller that t, i.e., $i^h(t) = i(t + h)$. Recalling the definition (1.4.4), we

can write

$$I_{t+h} - I_h = \sigma^2_{i^h(t)}\theta\left(\lambda(t - \tau^h_{i^h(t)})\right) + \sum_{k=i^h(0)+1}^{i^h(t)} \sigma^2_{k-1}\theta\left(\lambda(\tau^h_k - \tau^h_{k-1})\right) - \sigma^2_{i^h(0)}\theta\left(-\lambda\tau^h_{i^h(0)}\right),$$

where we agree that the sum in the right hand side is zero if $i^h(t) = i^h(0)$. This relation shows that $(I_{t+h} - I_h)_{t\geq 0}$ and $(I_t)_{t\geq 0}$ are *the same function* of the two random sets $\mathcal{T}^h$ and $\mathcal{T}$. Since $\mathcal{T}^h$ and $\mathcal{T}$ have the same distribution (both are Poisson point processes on $\mathbb{R}$ with intensity λ), the processes $(I_{t+h} - I_h)_{t\geq 0}$ and $(I_t)_{t\geq 0}$ have the same distribution too.

We recall that $\mathcal{G}$ is the σ-field generated by the whole process $(I_t)_{t\geq 0}$. From the definition $X_t = W_{I_t}$ and from the fact that Brownian motion has independent, stationary increments, it follows that for every Borel subset $A \subseteq \mathcal{C}([0, +\infty))$

$$\begin{aligned} P\left(X_{h+\cdot} - X_h \in A\right) &= \mathbb{E}\left[P\left(W_{I_{h+\cdot}} - W_{I_h} \in A \,|\, \mathcal{G}\right)\right] = \mathbb{E}\left[P\left(W_{I_{h+\cdot}-I_h} \in A \,|\, \mathcal{G}\right)\right] \\ &= P\left(W_{I_{h+\cdot}-I_h} \in A\right) = P\left(W_{I_\cdot} \in A\right) = P\left(X_\cdot \in A\right), \end{aligned}$$

where we have used the stationarity property of the process I. Thus the processes $(X_t)_{t\geq 0}$ and $(X_{h+t} - X_h)_{t\geq 0}$ have the same distribution, which implies stationarity of the increments. □

Proof of property (B). Observe first that $I'_s := \frac{\mathrm{d}}{\mathrm{d}s} I_s > 0$ a.s. and for Lebesgue–a.e. $s \geq 0$. By a change of variable, we can rewrite the process $(B_t)_{t\geq 0}$ defined in (1.5.2) as

$$B_t = \int_0^{I_t} \frac{1}{\sqrt{I'(I^{-1}(u))}}\,\mathrm{d}W_u = \int_0^t \frac{1}{\sqrt{I'_s}}\,\mathrm{d}W_{I_s} = \int_0^t \frac{1}{\sqrt{I'_s}}\,\mathrm{d}X_s\,,$$

which shows that relation (1.5.1) holds true. It remains to show that $(B_t)_{t\geq 0}$ is indeed a standard Brownian motion. Note that

$$B_t = \int_0^{I_t} \sqrt{(I^{-1})'(u)}\,\mathrm{d}W_u\,.$$

Therefore, conditionally on $\mathcal{I}$, $(B_t)_{t\geq 0}$ is a centered Gaussian process (it is a Wiener integral), with conditional covariance given by

$$Cov(B_s, B_t \,|\, \mathcal{I}) = \int_0^{\min\{I_s, I_t\}} (I^{-1})'(u)\,\mathrm{d}u = \min\{s, t\}\,.$$

This shows that, conditionally on $\mathcal{I}$, $(B_t)_{t\geq 0}$ is a Brownian motion. Therefore, it is *a fortiori* a Brownian motion without conditioning. □

Proof of property (C). The assumption $\mathbb{E}[\sigma^2] < \infty$ ensures that $\mathbb{E}[|X_t|^2] < \infty$ for all $t \geq 0$, as we prove in Proposition 1.8 below. Let us now denote by $\mathcal{F}^X_t = \sigma(X_s,\, s \leq t)$ the natural filtration of the process X. We recall that $\mathcal{G}$ denotes the σ-field generated by the whole process $(I_t)_{t\geq 0}$ and we denote by $\mathcal{F}^X_t \vee \mathcal{G}$ the smallest σ-field containing $\mathcal{F}^X_t$ and $\mathcal{G}$. Since $\mathbb{E}[W_{I_{t+h}} - W_{I_t} | \mathcal{F}^X_t \vee \mathcal{G}] = 0$ for all $h \geq 0$, by the basic properties of Brownian motion, recalling that $X_t = W_{I_t}$ we obtain

$$\mathbb{E}[X_{t+h} | \mathcal{F}^X_t \vee \mathcal{G}] = X_t + \mathbb{E}[W_{I_{t+h}} - W_{I_t} | \mathcal{F}^X_t \vee \mathcal{G}] = X_t.$$

Taking the conditional expectation with respect to $\mathcal{F}^X_t$ on both sides, we obtain the martingale property for $(X_t)_{t\geq 0}$. □

Remark 1.7. If we look at the process X for a *fixed* realization of the variables $\mathcal{T}$ and Σ, averaging only on W — that is, if we work under the conditional probability $P(\,\cdot\,|\mathcal{T},\Sigma)$ — the increments of X are no longer stationary, as it is clear from (1.4.2). We stress however that the properties (B) and (C) above continue to hold also under $P(\,\cdot\,|\mathcal{T},\Sigma)$. Of course, the condition $\mathbb{E}[\sigma^2]<\infty$ in (C) is not needed under $P(\,\cdot\,|\mathcal{T},\Sigma)$.

1.6 Some Properties on the Tails Distribution

The next property show some relations between the distribution ν of the σ_k's and the law of the random variable X_t.

Proposition 1.8. *For any $t>0$*

$$\mathbb{E}[|X_t|^q]<\infty \quad\Longleftrightarrow\quad \mathbb{E}[\sigma^q]<\infty\,.$$

Proof. Note that $\mathbb{E}[|X_t|^q]=\mathbb{E}[|W_{I_t}|^q]=\mathbb{E}[|I_t|^{q/2}]\,\mathbb{E}[|W_1|^q]$, by the independence of W and I and the scaling properties of Brownian motion. We are therefore left with showing that

$$\mathbb{E}[|I_t|^{q/2}]<\infty \quad\Longleftrightarrow\quad \mathbb{E}[\sigma^q]<\infty\,. \tag{1.6.1}$$

The implication "$\Rightarrow$" is easy: by the definition (1.4.4) of the process I we can write

$$\mathbb{E}[|I_t|^{q/2}]\geq\mathbb{E}[|I_t|^{q/2}\,\mathbf{1}_{\{i(t)=0\}}]=\mathbb{E}[\sigma_0^q]\,\mathbb{E}[|\theta(\lambda(t-\tau_0))-\theta(-\lambda\tau_0)|^{q/2}]P(i(t)=0)\,,$$

therefore if $\mathbb{E}[\sigma^q]=\infty$ then also $\mathbb{E}[|I_t|^{q/2}]=\infty$.

Next we prove the implication "$\Leftarrow$". Note that

$$\theta(a+b)-\theta(b)\leq\theta(a) \tag{1.6.2}$$

for all $a,b\geq 0$ because, from the fact that θ is $\mathcal{C}^1$ one has:

$$\theta(a+b)-\theta(b)=\int_0^a\theta'(b+s)\mathrm{d}s$$

and

$$\theta(a)=\int_0^a\theta'(s)\mathrm{d}s;$$

but $\theta'(\cdot)$ is decreasing by concavity, so $\theta'(b+s)\leq\theta'(s)$ and then $\theta(a+b)-\theta(b)\leq\theta(a)$.

If $i(t)\geq 1$, i.e. $\tau_1\leq t$ we can write

$$\begin{aligned} I_t&=\sigma^2_{i(t)}\theta(\lambda(t-\tau_{i(t)}))+\sum_{k=2}^{i(t)}\sigma^2_{k-1}\theta((\lambda(\tau_k-\tau_{k-1}))+\sigma_0^2\left[\theta(\lambda(\tau_1-\tau_0))-\theta(-\lambda\tau_0)\right]\\ &\leq\sigma^2_{i(t)}\theta(\lambda(t-\tau_{i(t)}))+\sum_{k=2}^{i(t)}\sigma^2_{k-1}\theta(\lambda(\tau_k-\tau_{k-1}))+\sigma_0^2\theta(\lambda\tau_1)\,, \end{aligned} \tag{1.6.3}$$

where we agree that the sum over k is zero if $i(t)=1$. Since $\tau_k\leq t$ for all $k\leq i(t)$, by the definition (1.4.3) of $i(t)$, and since θ is increasing by assumption, relation (1.6.3) yields the

bound $I_t \le \theta(\lambda t)\sum_{k=0}^{i(t)}\sigma_k^2$, which holds clearly also when $i(t)=0$. In conclusion, we have shown that for all $t, q > 0$

$$|I_t|^{q/2} \le \theta(\lambda t)^{\frac{q}{2}} \left(\sum_{k=0}^{i(t)} \sigma_k^2\right)^{q/2} . \tag{1.6.4}$$

Consider first the case $q > 2$: by Jensen's inequality we have

$$\left(\sum_{k=0}^{i(t)} \sigma_k^2\right)^{q/2} = (i(t)+1)^{q/2} \left(\frac{1}{i(t)+1}\sum_{k=0}^{i(t)} \sigma_k^2\right)^{q/2} \le (i(t)+1)^{q/2-1} \sum_{k=0}^{i(t)} \sigma_k^q \,. \tag{1.6.5}$$

Therefore, if $\mathbb{E}[\sigma^q] < \infty$, from (1.6.4) we can write

$$\mathbb{E}[|I_t|^{q/2}] \le \theta(\lambda t)^{\frac{q}{2}}\, \mathbb{E}[(i(t)+1)^{q/2}]\, \mathbb{E}[\sigma^q] < \infty \,, \tag{1.6.6}$$

because $i(t)$, being a Poisson random variable of mean λt, has finite moments of all orders and is independent of the variables $(\sigma_k)_{k\ge 0}$. Analogously, since $(\sum_{k=1}^{\infty} x_k)^\alpha \le \sum_{k=1}^{\infty} x_k^\alpha$ for $\alpha \in (0,1)$ and for every non-negative sequence $(x_n)_{n\in\mathbb{N}}$, for $q \le 2$ by (1.6.4) we can write

$$\mathbb{E}[|I_t|^{q/2}] \le \theta(\lambda t)^{\frac{q}{2}}\, \mathbb{E}\left[\sum_{k=0}^{i(t)} \sigma_k^q\right] \le \mathbb{E}[\theta(\lambda(t-\tau_0))^{\frac{q}{2}}]\, \mathbb{E}[i(t)+1]\, \mathbb{E}[\sigma^q] < \infty \,. \tag{1.6.7}$$

The proof of (1.6.1) is complete. □

As for most real financial indexes (like the DJIA, see Chapter 3) there is empirical evidence that $\mathbb{E}[|X_t|^q] < \infty$ for all $q > 0$, the main interest of our model is when σ *has finite moments of all orders*. In this case, a link between the exponential moments of σ and those of X_t is given in Proposition 1.9 below, in the case $\theta(t) = t^{(}2D)$. We believe that the same should hold for the general case too, but we didn't prove it yet; anyway, this fact has not much importance for the financial applications of the model, see Remark 1.12.

Proposition 1.9. *Regardless of the distribution of σ, for every $q > (1-D)^{-1}$ we have*

$$\mathbb{E}\left[\exp\left(\gamma|X_t|^q\right)\right] = \infty \,, \qquad \forall t > 0 \,,\ \forall \gamma > 0 \,. \tag{1.6.8}$$

On the other hand, for all $q < (1-D)^{-1}$ and $t > 0$ we have

$$\mathbb{E}\left[\exp\left(\gamma|X_t|^q\right)\right] < \infty \quad \forall \gamma > 0 \qquad \Longleftrightarrow \qquad \mathbb{E}\left[\exp\left(\alpha\sigma^{\frac{2q}{2-q}}\right)\right] < \infty \quad \forall \alpha > 0 \,, \tag{1.6.9}$$

and the same relation holds for $q = (1-D)^{-1}$ provided $D < \frac{1}{2}$.

Note that $(1-D)^{-1} \in (1,2]$, because $D \in (0, \frac{1}{2}]$, so that for $D < \frac{1}{2}$ the distribution of X_t has always tails heavier than Gaussian.

We first need two simple technical lemmas.

Lemma 1.10. *For $0 < q < 2$, consider the function $\varphi_q : [0,+\infty) \to [0,+\infty)$ define by*

$$\varphi_q(\beta) := \int_{-\infty}^{+\infty} e^{\beta|x|^q - \frac{1}{2}x^2}\, \mathrm{d}x.$$

Then there are constants $C_1, C_2 > 0$, that depend on q, such that for all $\beta > 0$

$$C_1 e^{C_1 \beta^{\frac{2}{2-q}}} \le \varphi_q(\beta) \le C_2 e^{C_2 \beta^{\frac{2}{2-q}}} \,. \tag{1.6.10}$$

Proof. We begin by observing that it is enough to establish the bounds in (1.6.10) for β large enough. Consider the function of positive real variable $f(r) := e^{\beta r^q - \frac{1}{2}r^2}$. It is easily checked that f is increasing for $0 \le r \le (\beta q)^{\frac{1}{2-q}}$. Thus

$$\varphi_q(\beta) \ge \int_{\frac{1}{2}(\beta q)^{\frac{1}{2-q}}}^{(\beta q)^{\frac{1}{2-q}}} f(r)\,\mathrm{d}r \ge \frac{1}{2}(\beta q)^{\frac{1}{2-q}} f\left(\frac{1}{2}(\beta q)^{\frac{1}{2-q}}\right) = \frac{1}{2}(\beta q)^{\frac{1}{2-q}} \exp\left[c(q)\beta^{\frac{2}{2-q}}\right],$$

with $c(q) := \frac{1}{2^q} q^{\frac{q}{2-q}} - \frac{1}{8} q^{\frac{2}{2-q}} > 0$. The lower bound in (1.6.10) easily follows for β large.

For the upper bound, by direct computation one observes that $f(r) \le e^{-\frac{1}{4}r^2}$ for $r > (4\beta)^{\frac{1}{2-q}}$. We have:

$$\varphi_q(\beta) \le \int_{|x|\le(4\beta)^{\frac{1}{2-q}}} f(|x|)\,\mathrm{d}x + \int_{|x|>(4\beta)^{\frac{1}{2-q}}} e^{-\frac{1}{4}x^2}\mathrm{d}x \le 2(4\beta)^{\frac{1}{2-q}}\|f\|_\infty + \int_{-\infty}^{+\infty} e^{-\frac{1}{4}x^2}\mathrm{d}x\,.$$

Since $\|f\|_\infty = f((\beta q)^{\frac{1}{2-q}}) = \exp\left(C(q)\beta^{\frac{2}{2-q}}\right)$ for a suitable $C(q)$, also the upper bound follows, for β large. □

Lemma 1.11. *Let $X_1, X_2, \ldots, X_n$ be independent random variables uniformly distributed in $[0,1]$, and $U_1 < U_2 < \ldots < U_n$ be the associated order statistics. For $n \ge 2$ and $k = 2, \ldots, n$, set $\xi_k := U_k - U_{k-1}$. Then, for every $\epsilon > 0$*

$$\lim_{n\to+\infty} P\left(\left|\left\{k \in \{2,\ldots,n\} :\ \xi_k > \frac{1}{n^{1+\epsilon}}\right\}\right| \ge n^{1-\epsilon}\right) = 1.$$

Proof. This is a consequence of the following stronger result: for every $x > 0$, as $n \to \infty$ we have the convergence in probability

$$\frac{1}{n}\left|\left\{k \in \{2,\ldots,n\} :\ \xi_k > \frac{x}{n}\right\}\right| \longrightarrow e^{-x}\,,$$

see [27] for a proof. □

Proof of Proposition 1.9. Without loss of generality, we set $\lambda = 0$. Since $X_t = W_{I_t}$ and $\sqrt{I_t}\,W_1$ have the same law, we can write

$$\mathbb{E}\left[e^{\gamma|X_t|^q}\right] = \mathbb{E}\left[\exp\left(\gamma I_t^{q/2}|W_1|^q\right)\right].$$

We begin with the proof of (1.6.9), hence we work in the regime $q < (1-D)^{-1}$, or $q = (1-D)^{-1}$ and $D < \frac{1}{2}$; in any case, $q < 2$. We start with the "$\Leftarrow$" implication. Since I_t and W_1 are independent, it follows by Lemma 1.10 that

$$\mathbb{E}\left[\exp\left(\gamma I_t^{q/2}|W_1|^q\right)\right] \le C\mathbb{E}\left[\exp\left(\delta I_t^{\frac{q}{2-q}}\right)\right], \tag{1.6.11}$$

for some $C, \delta > 0$. For the moment we work on the event $\{i(t) \ge 1\}$. It follows by the basic bound (1.6.3) that

$$I_t \le \sum_{k=0}^{i(t)} \xi_k^{2D}\,\sigma_k^2\,, \tag{1.6.12}$$

where we set

$$\xi_k := \begin{cases} \tau_1 & \text{for } k = 0 \\ \tau_{k+1} - \tau_k & \text{for } 1 \le k \le i(t) - 1 \\ 1 - \tau_{i(t)} & \text{for } k = i(t) \end{cases}$$

Note that $\sum_{k=0}^{i(t)} \xi_k = 1$. It is easily shown by Lagrange multipliers that the function $(x_0, x_1, \dots, x_n) \mapsto \sum_{k=0}^{n} x_k^{2D} \sigma_k^2$ subject to the constraint $\sum_{k=0}^{n} x_k = 1$ attains its maximum at $(x_0^*, x_1^*, \dots, x_n^*)$, with

$$x_k^* = \frac{\sigma_k^{\frac{2}{1-2D}}}{\sum_{h=0}^{n} \sigma_h^{\frac{2}{1-2D}}} .$$

Then it follows from (1.6.12) that

$$I_t \le \sum_{k=0}^{i(t)} (x_k^*)^{2D} \sigma_k^2 = \left(\sum_{k=0}^{i(t)} \sigma_k^{\frac{2}{1-2D}} \right)^{1-2D} .$$

By assumption $q \le \frac{1}{1-D}$, which is the same as $(1-2D)\frac{q}{2-q} \le 1$. Thus

$$I_1^{\frac{q}{2-q}} \le \left(\sum_{k=0}^{i(t)} \sigma_k^{\frac{2}{1-2D}} \right)^{(1-2D)\frac{q}{2-q}} \le \sum_{k=0}^{i(t)} \sigma_k^{\frac{2q}{2-q}} .$$

Now observe that if $i(t) = 0$ we have $I_t = (t-\tau_0)^{2D} - (-\tau_0)^{2D} \le t^{2D}$. Therefore, by (1.6.11)

$$\begin{aligned} \mathbb{E}\left[e^{\gamma |X_t|^q} \right] &\le C \left(\mathbb{E}\left[\exp\left(\delta \sum_{k=0}^{i(t)} \sigma_k^{\frac{2q}{2-q}} \right) \mathbf{1}_{\{i(t)\ge 1\}} \right] + \mathbb{E}\left[\exp\left(\delta\, t^{\frac{2Dq}{2-q}} \right) \mathbf{1}_{\{i(t)=0\}} \right] \right) \\ &\le C \left(\mathbb{E}\left[\rho^{i(t)+1} \right] + \exp\left(\delta\, t^{\frac{2Dq}{2-q}} \right) \right), \end{aligned} \tag{1.6.13}$$

where we have set

$$\rho := \mathbb{E}\left[\exp\left(\delta\, \sigma_0^{\frac{2q}{2-q}} \right) \right] .$$

Therefore, if $\rho < \infty$, the right hand side of (1.6.13) is finite, because $i(t) \sim Po(t)$ has finite exponential moments of all order. This proves the "$\Leftarrow$" implication in (1.6.9).

The "$\Rightarrow$" implication in (1.6.9) is simpler. By the lower bound in Lemma 1.10 we have

$$\mathbb{E}\left[e^{\gamma |X_t|^q} \right] = \mathbb{E}\left[\exp\left(\gamma I_t^{q/2} |W_1|^q \right) \right] \ge C \mathbb{E}\left[\exp\left(\delta I_t^{\frac{q}{2-q}} \right) \right], \tag{1.6.14}$$

for suitable $C, \delta > 0$. We note that

$$\begin{aligned} \mathbb{E}\left[\exp\left(\delta I_t^{\frac{q}{2-q}} \right) \right] &\ge \mathbb{E}\left[\exp\left(\delta I_t^{\frac{q}{2-q}} \right) \mathbf{1}_{\{i(t)=0\}} \right] \\ &= \mathbb{E}\left[\exp\left(\delta \left[(t-\tau_0)^{2D} - (-\tau_0)^{2D} \right] \sigma_0^{\frac{2q}{2-q}} \right) \right] P(i(t) = 0) . \end{aligned} \tag{1.6.15}$$

Under the condition

$$\mathbb{E}\left[\exp\left(\alpha\sigma^{\frac{2q}{2-q}}\right)\right] = +\infty \ \forall \alpha > 0,$$

the last expectation in (1.6.15) is infinite, since $\left[(t-\tau_0)^{2D} - (-\tau_0)^{2D}\right] > 0$ almost surely and is independent of σ_0. Looking back at (1.6.14), we have proved the "$\Rightarrow$" implication in (1.6.9).

Next we prove (1.6.8), hence we assume that $q > (1-D)^{-1}$. Consider first the case $q < 2$ (which may happen only for $D < \frac{1}{2}$). By (1.6.14)

$$\mathbb{E}\left[e^{\gamma|X_t|^q}\right] \geq C\mathbb{E}\left[\exp\left(\delta I_t^{\frac{q}{2-q}}\right)\right].$$

We note that, by the definition (1.4.3) of I_t, we can write

$$I_t \geq \sum_{k=2}^{i(t)} \sigma_{k-1}^2 (\tau_k - \tau_{k-1})^{2D}, \tag{1.6.16}$$

where we agree that the sum is zero if $i(t) < 2$. For $n \geq 0$, we let P_n to denote the conditional probability $P(\,\cdot\,|i(t) = n)$ and E_n the corresponding expectation. Note that, under P_n, the random variables $(\tau_k - \tau_{k-1})_{k=2}^n$ have the same law of the random variables $(\xi_k)_{k=2}^n$ in Lemma 1.11, for $n \geq 2$. Consider the following events:

$$A_n := \{\sigma_k^2 \geq a, \ \forall k = 2, \dots, n\}, \qquad B_n := \left\{\left|\left\{k = 2, \dots, n : \ \xi_k > \frac{1}{n^{1+\epsilon}}\right\}\right| \geq n^{1-\epsilon}\right\},$$

where $a > 0$ is such that $\nu([a, +\infty)) =: \rho > 0$ and $\epsilon > 0$ will be chosen later. Note that $P_n(A_n) = \rho^{n-1}$ while $P_n(B_n) \to 1$ as $n \to +\infty$, by Lemma 1.11. In particular, there is $c > 0$ such that $P_n(B_n) \geq c$ for every n. Plainly, A_n and B_n are independent under P_n. We have

$$\begin{aligned}\psi(n) := \mathbb{E}_n\left[\exp\left(\delta I_t^{\frac{q}{2-q}}\right)\right] &\geq \mathbb{E}_n\left[\exp\left(\delta I_t^{\frac{q}{2-q}}\right)\mathbf{1}_{A_n\cap B_n}\right] \\ &\geq c\,\rho^{n-1}\exp\left[\delta a^{\frac{q}{2-q}}\left(\frac{1}{n^{1+\epsilon}}\right)^{2D\frac{q}{2-q}} n^{(1-\epsilon)\frac{q}{2-q}}\right] \\ &= c\,\rho^{n-1}\exp\left[\delta a^{\frac{q}{2-q}} n^{(1-2D-\epsilon(1+2D))\frac{q}{2-q}}\right] \end{aligned} \tag{1.6.17}$$

Note that $q > \frac{1}{1-D}$ is equivalent to $(1-2D)\frac{q}{2-q} > 1$, therefore ϵ can be chosen small enough so that $b := (1-2D-\epsilon(1+2D))\frac{q}{2-q} > 1$. It then follows by (1.6.17) that $\psi(n) \geq d\,\exp(d\,n^b)$ for every $n \in \mathbb{N}$, for a suitable $d > 0$. Therefore

$$\mathbb{E}\left[\exp\left(\delta I_t^{\frac{q}{2-q}}\right)\right] = \mathbb{E}[\psi(i(t))] = +\infty,$$

because $i(t) \sim Po(t)$ and hence $\mathbb{E}[\exp(d\,i(t)^b)] = \infty$ for all $d > 0$ and $b > 1$.

Next we consider the case $q \geq 2$. Note that

$$\mathbb{E}\left[e^{\gamma|X_t|^q}\right] = \mathbb{E}\left[\exp\left(\gamma I_t^{q/2}|W_1|^q\right)\right], \tag{1.6.18}$$

hence if $q > 2$ we have $\mathbb{E}\left[e^{\gamma|X_t|^q}\right] = \infty$, because $\mathbb{E}[\exp(c|W_1|^q)] = \infty$ for every $c > 0$, $I_t > 0$ almost surely and I_t is independent of W_1. On the other hand, if $q = 2$ we must have $D < \frac{1}{2}$ (recall that we are in the regime $q > (1-D)^{-1}$) and the steps leading to (1.6.17) have shown that in this case I_t is unbounded. It then follows again from (1.6.18) that $\mathbb{E}[e^{\gamma|X_t|^2}] = \infty$. □

Remark 1.12. If ν has a bounded support, Proposition 1.9 implies $\mathbb{E}\left[\exp\left(\gamma|X_t|^q\right)\right] < +\infty$ for $q < (1-D)^{-1}$, and $\mathbb{E}\left[\exp\left(\gamma|X_t|^q\right)\right] = +\infty$ for $q > (1-D)^{-1}$. This means that the tails of the distribution of X_t are heavier than Gaussian but lighter than exponential, *seemingly* in contrast with the red curve in Figure 3.3(B), that shows a decay *slower* than exponential for the distribution of the daily log-returns of our model (with parameters chosen to fit the DJIA time series, see (3.2.4)). As a matter of fact, simulations show that the predicted super-exponential decay of the tails can actually be seen:

- at a much larger distance from the mean than six standard deviations (which is the range of Figure 3.3(B)) for *daily* log-returns;
- at a more reasonable distance from the mean for log-returns over longer periods, such as hundreds of days (cf. the red curve in Figure 3.5(B)).

As both cases appear to have little interest in applications, the *relevant* tails are heavier than exponential, despite of Proposition 1.9.

1.7 Ergodicity

We finally show that the distribution of the increments of the process is *ergodic*. The ergodicity of the process entails in particular that for every $\delta > 0$, $k \in \mathbb{N}$ and for every choice of the intervals $(a_1, b_1), \ldots, (a_k, b_k) \subseteq (0, \infty)$ and of the measurable function $F : \mathbb{R}^k \to \mathbb{R}$, we have almost surely

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} F(X_{n\delta+b_1} - X_{n\delta+a_1}, \ldots, X_{n\delta+b_k} - X_{n\delta+a_k}) \\ = \mathbb{E}\big[F(X_{b_1} - X_{a_1}, \ldots, X_{b_k} - X_{a_k})\big], \quad (1.7.1)$$

provided the expectation appearing in the right hand side is well defined. In words: the empirical average of a function of the increments of the process over a long time period is close to its expected value.

Thanks to this property, our main results concerning the moments and the correlation of the increments of the process X, that we state in the next subsection, are of direct relevance for the comparison of our model with real data series. Some care is needed, however, because the accessible time length N in (1.7.8) may not be large enough to ensure that the empirical averages are close to their limit. We elaborate more on this issue in Chapter 3, where we compare our model with the DJIA data from a numerical viewpoint.

We actually show two stronger results than ergodicity; a 0–1 law for the tail σ-field of X, and the fact that the increments of X are *exponentially mixing*. In what follows, given a continuous or discrete-time stochastic process $\eta = (\eta_t)_{t\geq 0}$, we set $\Theta_t(\eta) := \sigma\{\eta_s : s \geq t\}$ and

we define the so-called *tail* σ-field $\Theta(\eta) := \bigcap_{t\geq 0} \Theta_t(\eta)$. Also, for an interval $I \subseteq [0, +\infty)$, we let

$$\mathcal{F}_I^{\mathcal{D}} := \sigma\left(X_t - X_s : s, t \in I\right)$$

to denote the σ-field generated by the increments in I of the process X.

Proposition 1.13. *The following properties hold:*

(a) *The tail σ-field $\Theta(X)$ is* trivial, *i.e., its events have probability either zero or one.*

(b) *Let $I = [a, b)$, $J = [c, d)$, with $0 \leq a < b \leq c < d$. Then, for every $A \in \mathcal{F}_I^{\mathcal{D}}$ and $B \in \mathcal{F}_J^{\mathcal{D}}$*

$$|P(A \cap B) - P(A)P(B)| \leq 2e^{-\lambda(c-b)}. \tag{1.7.2}$$

We first need some preliminary results for the proof of Proposition 1.13.

Lemma 1.14. *Let X, Y be independent random vectors and let f be Borel such that $f(X, Y)$ is integrable. Define*

$$g(x) \;=\; \begin{cases} \mathbb{E}[f(x, Y)] & \text{if } |\mathbb{E}[f(x, Y)]| < \infty \\ 0 & \text{otherwise.} \end{cases}$$

then $\mathbb{E}[f(X, Y)|X] = g(X)$.

Theorem 1.15 (The Hewitt-Savage 0-1 Law)**.** *Consider $(\mathbb{R}^{\mathbb{N}}, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is the σ-algebra generated by cylinders. Denote by Z_n the n^{th} projection of $\mathbb{R}^{\mathbb{N}}$, let $\mathcal{S}_n$ be the class of $\mathcal{F}$-sets A that are invariant under permutations of the first n coordinates, and define $\mathcal{S} = \bigcap_{n=1}^{\infty} \mathcal{S}_n$. If the Z_n are independent and identically distributed under P, then $P(A)$ is 0 or 1 for each A in $\mathcal{S}$.*

For the proves of these results, we remind to [8] and [20].

Lemma 1.16. *Let η and θ be two independent stochastic processes, such that both $\Theta(\eta)$ and $\Theta(\theta)$ are trivial. Then the σ-field*

$$\mathcal{G} := \bigcap_{s,t\geq 0} \left[\Theta_t(\eta) \vee \Theta_s(\theta)\right]$$

is trivial, where $\Theta_t(\eta) \vee \Theta_s(\theta)$ denotes the smallest σ-field containing $\Theta_t(\eta) \cup \Theta_s(\theta)$.

Proof. We introduce the function space $E := \mathbb{R}^{[0,+\infty)}$, equipped with the σ-field generated by cylinder sets. If Z is a bounded $\mathcal{G}$-measurable random variable, then for every $s, t \geq 0$ we can write

$$Z = f_{s,t}\left(\eta_{s+\cdot}, \theta_{t+\cdot}\right), \tag{1.7.3}$$

where $\eta_{s+\cdot}$ denotes the process $(\eta_{s+u})_{u\geq 0}$, and $f_{s,t} : E \times E \to \mathbb{R}$ is bounded and measurable. We set $f := f_{0,0}$ and we denote by μ_η, μ_θ respectively the laws of η, θ on E.

By (1.7.3), for μ_η-a.e. $x \in E$ we have $f(x, \theta) = f_{0,t}(x, \theta_{t+\cdot})$, for every $t \geq 0$. Thus $f(x, \theta)$ is $\Theta(\theta)$-measurable and hence a.s. constant: $f(x, \theta) = f_1(x)$ a.s., where $f_1(x) := \mathbb{E}[f(x, \theta)]$. Therefore, for every measurable $B \subseteq E$ we can write

$$\int_E f(x, y)\, \mathbf{1}_B(y)\, \mu_\theta(\mathrm{d}y) = f_1(x) \cdot \mu_\theta(B)\,, \qquad \text{for } \mu_\eta\text{-a.e. } x \in E\,. \tag{1.7.4}$$

Observe now that, again by (1.7.3), we have a.s.

$$f_1(\eta) = \mathbb{E}[f(\eta,\theta)|\eta] = \mathbb{E}(f_{t,0}[\eta_{t+\cdot},\theta)|\eta] = \mathbb{E}[f_{t,0}(x,\theta)]|_{x=\eta_{t+\cdot}}\,.$$

This shows that $f_1(\eta)$ is $\Theta(\eta)$-measurable and hence a.s. constant: $f_1(\eta) = c$ a.s., where $c := \mathbb{E}[f_1(\eta)] = \mathbb{E}[f(\eta,\theta)]$. Therefore, for every measurable $A \subseteq E$ we can write

$$\int_E f_1(x)\,\mathbf{1}_A(x)\,\mu_\eta(\mathrm{d}x) = c\cdot\mu_\eta(A)\,. \tag{1.7.5}$$

Putting together (1.7.4), (1.7.5) and Fubini's theorem, it follows that for all measurable $A, B \subseteq E$

$$\int_{E\times E} f(x,y)\,\mathbf{1}_{A\times B}(x,y)\,(\mu_\eta\otimes\mu_\theta)(\mathrm{d}x,\mathrm{d}y) = c\,(\mu_\eta\otimes\mu_\theta)(A\times B)\,, \tag{1.7.6}$$

where $\mu_\eta\otimes\mu_\theta$ denotes of course the product measure of μ_η and μ_θ on $E\times E$.

Since the sets of the form $A\times B$ are a basis of the product σ-field of $E\times E$, relation (1.7.6) still holds true with $A\times B$ replaced by any measurable subset $C\subseteq E\times E$. Choosing $C_n^+ = \{(x,y):\, f(x,y) > c+\frac{1}{n}\}$ and $C_n^- = \{(x,y):\, f(x,y) < c-\frac{1}{n}\}$, it follows easily that $P(C_n^+) = P(C_n^-) = 0$ for every $n\in\mathbb{N}$, that is $f(x,y) = c$ for $(\mu_\eta\otimes\mu_\theta)$-a.e. $(x,y)\in E\times E$. Therefore $Z = f(\eta,\theta) = c$ a.s.. We have proved that every bounded $\mathcal{G}$-measurable random variable is almost surely constant, which is equivalent to the triviality of $\mathcal{G}$. □

Proof of Proposition 1.13. We begin with the proof of statement (a), the 0-1 law for $\Theta(X)$. If Y is a bounded $\Theta(X)$-measurable random variable, by definition for every $t\geq 0$ it can be written in the form

$$Y = \varphi_t\left(X_{t+\cdot}\right) = \varphi_t\left(W_{I_{t+\cdot}}\right),$$

where $\varphi_t : \mathcal{C}[0,+\infty)\to\mathbb{R}$ is bounded and measurable. Now, for $s\geq 0$, set

$$Y_{t,s} = \varphi_t\left(W_{I_{t+\cdot}}\right)\mathbf{1}_{\{I_t>s\}}.$$

Note that $Y_{t,s}$ is $\Theta_s(W)\vee\Theta_t(I)$-measurable. Plainly, $I_n\uparrow+\infty$ almost surely as $n\uparrow+\infty$, therefore for every $s\geq 0$

$$\lim_{n\to+\infty} Y_{n,s} = \lim_{n\to+\infty}\varphi_n\left(W_{I_{n+\cdot}}\right) = Y.$$

It follows that Y is $\bigcap_{s,t\geq 0}[\Theta_s(W)\vee\Theta_t(I)]$-measurable. In other words

$$\Theta(X)\subseteq\bigcap_{s,t\geq 0}\left[\Theta_s(W)\vee\Theta_t(I)\right],$$

and by Lemma 1.16 all we need to show is that both $\Theta(W)$ and $\Theta(I)$ are trivial.

The triviality of $\Theta(W)$ is well known: it follows, e.g., from Blumenthal 0-1 law (cf. Theorem 2.7.17 in [19]) applied to the Brownian motion $(sW_{1/s})_{s\geq 0}$. For the triviality of $\Theta(I)$ we proceed as follows. If $A\in\Theta(I)$, for each $n\in\mathbb{N}$

$$\mathbf{1}_A = \varphi_n(I_{n+\cdot})\,,$$

for a suitable measurable function φ_n. Recalling the definition (1.4.3) of the variable $i(t)$, it is clear that $i(t) \uparrow +\infty$ as $t \uparrow +\infty$, hence for every $m \in \mathbb{N}$

$$\mathbf{1}_A = \lim_{n \to +\infty} \varphi_n(I_{n+\cdot}) \mathbf{1}_{\{i(n)>m\}} .$$

Now define $\eta_k := (\sigma_k, \tau_{k+1} - \tau_k)$ and observe that the function

$$\mathbb{E}[\varphi_n(I_{n+\cdot}) \mathbf{1}_{\{i(n)>m\}} | \sigma(\eta_k : k \geq 0)]$$

is trivially measurable with respect to the σ-field $\sigma(\eta_k : k \geq 0)$, and is invariant under any permutation of $(\eta_0, \ldots, \eta_{m-1})$ by Lemma 1.14. This shows that $\mathbb{E}[\mathbf{1}_A | \sigma(\eta_k : k \geq 0)]$ is measurable for $\sigma(\eta_k : k \geq 0)$ and invariant by finite permutations of the η_k's. We have then, by Theorem 1.15, that $P(A|\sigma(\eta_k : k \geq 0))$ is either zero or one and then $P(A)$ too. The proof of (a) is now completed.

We now prove part (b). We recall that $\mathcal{T}$ denotes the set $\{\tau_k : k \in \mathbb{Z}\}$ and, for $I \subseteq \mathbb{R}$, $\mathcal{G}_I$ denotes the σ-algebra generated by the family of random variables $(\tau_k \mathbf{1}_{\{\tau_k \in I\}}, \sigma_k \mathbf{1}_{\{\tau_k \in I\}})_{k \geq 0}$, where $(\sigma_k)_{k \geq 0}$ is the sequence of volatilities. We are considering $I = [a, b)$, $J = [c, d)$, with $0 \leq a < b \leq c < d$. We introduce the $\mathcal{G}_{[b,c)}$-measurable event

$$\Gamma := \{\mathcal{T} \cap [b, c) \neq \emptyset\}.$$

We claim that, for $A \in \mathcal{F}_I^{\mathcal{D}}$, $B \in \mathcal{F}_J^{\mathcal{D}}$, we have

$$P(A \cap B|\Gamma) = P(A)P(B). \tag{1.7.7}$$

To see this, the key is in the following two remarks.

- $\mathcal{F}_I^{\mathcal{D}}$ and $\mathcal{F}_J^{\mathcal{D}}$ are independent *conditionally* to $\mathcal{G} = \mathcal{G}_{\mathbb{R}}$. This follows immediately from the independence of W and (I_t).
- Conditionally to $\mathcal{G}$, the process $(X_t)_{t \in I}$ is a Gaussian process whose covariances are all $\mathcal{G}_I$-measurable. It follows that $P(A|\mathcal{G})$ is $\mathcal{G}_I$-measurable. Similarly, $P(B|\mathcal{G})$ is $\mathcal{G}_J$-measurable.

Thus

$$P(A \cap B|\Gamma) = \mathbb{E}[P(A \cap B|\mathcal{G})|\Gamma] = \mathbb{E}[P(A|\mathcal{G})P(B|\mathcal{G})|\Gamma] = \frac{\mathbb{E}[P(A|\mathcal{G})\mathbf{1}_\Gamma P(B|\mathcal{G})]}{P(\Gamma)},$$

where the first of the remarks above has been used. Now, by the second remark, $P(A|\mathcal{G})\mathbf{1}_\Gamma$ is $\mathcal{G}_{[a,c)}$-measurable, and $P(B|\mathcal{G})$ is $\mathcal{G}_{[c,d)}$-measurable. Therefore they are independent, and we obtain

$$P(A \cap B|\Gamma) = P(A|\Gamma)P(B) = P(A)P(B)$$

where, in the last equality, we have used the fact that, since $\mathcal{G}_I$ and Γ are independent, $P(A|\Gamma) = \mathbb{E}[P(A|\mathcal{G})|\Gamma] = \mathbb{E}[P(A|\mathcal{G})] = P(A)$. To complete the proof of part (b) we observe that, by what just shown,

$$|P(A \cap B) - P(A)P(B)| = |P(A \cap B) - P(A \cap B|\Gamma)|.$$

But, letting $C := A \cap B$

$$|P(C) - P(C|\Gamma)| = \left| \mathbb{E}\left[\mathbf{1}_C\right] - \frac{\mathbb{E}\left[\mathbf{1}_C \mathbf{1}_\Gamma\right]}{P(\Gamma)} \right| = \frac{|\mathbb{E}\left[(\mathbf{1}_\Gamma - P(\Gamma))\mathbf{1}_C\right]|}{P(\Gamma)} \leq$$

$$\leq \frac{\mathbb{E}\left[|(\mathbf{1}_\Gamma - P(\Gamma))|\right]}{P(\Gamma)} = \frac{(1-P(\Gamma))\mathbb{E}\left[\mathbf{1}_\Gamma\right] + P(\Gamma)\mathbb{E}\left[\mathbf{1}_{\overline{\Gamma}}\right]}{P(\Gamma)} = 2(1-P(\Gamma)).$$

Since $1 - P(\Gamma) = e^{-\lambda(c-b)}$, we obtain

$$|P(A \cap B) - P(A)P(B)| \leq 2(1-P(\Gamma)) = 2e^{-\lambda(c-b)}$$

as desired. □

As a consequence of Proposition 1.13, equation (1.7.8) holds true almost surely and in L^1, for every measurable function $F : \mathbb{R}^k \to \mathbb{R}$ such that $\mathbb{E}[|F(X_{b_1} - X_{a_1}, \ldots, X_{b_k} - X_{a_k})|] < +\infty$. To show this, we first need some theory on ergodicity.

Definition 1.17. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A mapping $T : \Omega \to \Omega$ is a *measure-preserving trasformation* if it is measurable and $P(T^{-1}A) = P(A)$ for all $A \in \mathcal{F}$ (and then $P(T^{-n}A) = P(A)$ for $n \geq 0$). If, further, T is a one-to-one mapping onto Ω and the point mapping T^{-1} is measurable, then T is *invertible* and T^{-1} automatically preserves measures: $P(A) = P(T^{-1}TA) = P(TA)$.

Example 1.18. *The Bernoulli Shift.* Let $\mathbb{R}$ be equipped with probabilities P_u, and set $\Omega = \mathbb{R}^{\mathbb{N}}$ the space of all infinite sequences

$$\omega = (z_1(\omega), z_2(\omega), \ldots)$$

of elements of $\mathbb{R}$: $z_k(\omega) \in \mathbb{R}$ for all $\omega \in \mathbb{R}^{\mathbb{N}}$ and $k \geq 1$. Consider on Ω the σ-field $\mathcal{C}$ of cylinders and the product probability $P = \otimes_u P_u$.

Define the *shift* T by

$$T\omega = (z_2(\omega), z_3(\omega), \ldots),$$

that is $z_k(T\omega) = z_{k+1}(\Omega)$ for $k \geq 1$.

The shift T is then measurable from $(\Omega, \mathcal{C})$ to $(\Omega, \mathcal{C})$ and preserves the measure P.

Definition 1.19. The $\mathcal{F}$-set A is *invariant* under the measure-preserving transformation T if $T^{-1}A = A$; it is a *nontrivial invariant set* if $0 < P(A) < 1$. T is said to be *ergodic* if in $\mathcal{F}$ the only invariant sets are the trivial ones (that is $T^{-1}A = A \Rightarrow P(A) = 0$ or $P(A) = 1$).

A measurable function f is *invariant* if $f(T\omega) = f(\omega)$ for all ω.

Theorem 1.20 (Ergodic Theorem)**.** *Suppose that T is a measure-preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$ and that f is measurable and integrable. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(T^{k-1}\omega) = \widetilde{f}(\omega)$$

with probability 1*, where $\widetilde{f}$ is invariant and integrable and $\mathbb{E}[\widetilde{f}] = \mathbb{E}[f]$. If T is ergodic, then $\widetilde{f} = \mathbb{E}[f]$ with probability* 1*.*

Definition 1.21. Let T denotes the shift (see Example 1.18) and $Y = (Y_n)_{n\in\mathbb{N}}$ a real valued, stationary and discrete stochastic process. The process Y naturally induces on $\mathbb{R}^{\mathbb{N}}$ a probability P_Y in this way: $P_Y(A) = P(Y \in A)$, for every A in the σ-field of cylinders of $\mathbb{R}^{\mathbb{N}}$. The process Y is said to be *ergodic* if under the probability P_Y the shift T is ergodic.

We can now turn back to consider our model process X and proof an ergodicity property for its increments.

Proposition 1.22. *For every $\delta > 0$, $k \in \mathbb{N}$ and for every choice of the intervals (a_1, b_1), $\ldots$, $(a_k, b_k) \subseteq (0, \infty)$ and of the measurable function $F : \mathbb{R}^k \to \mathbb{R}$, we have almost surely*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} F(X_{n\delta+b_1} - X_{n\delta+a_1}, \ldots, X_{n\delta+b_k} - X_{n\delta+a_k}) = \mathbb{E}\big[F(X_{b_1} - X_{a_1}, \ldots, X_{b_k} - X_{a_k})\big], \quad (1.7.8)$$

provided that $\mathbb{E}[|F(X_{b_1} - X_{a_1}, \ldots, X_{b_k} - X_{a_k})|] < +\infty$.

Proof. Consider the $\mathbb{R}^k$-valued stochastic process $\xi = (\xi_n)_{n\in\mathbb{N}}$ defined by

$$\xi_n := (X_{n\delta+b_1} - X_{n\delta+a_1}, \ldots, X_{n\delta+b_k} - X_{n\delta+a_k}),$$

for fixed $\delta > 0$, $k \in \mathbb{N}$ and (a_1, b_1), $\ldots$, $(a_k, b_k) \subseteq (0, \infty)$. The process ξ is stationary, because we have proven in section 1.5 that X has stationary increments. Furthermore, the tail triviality of X implies that ξ is *ergodic*, because every invariant event of ξ is contained in the tail σ-field of X: in fact, if A is invariant then, for all $k \geq 1$

$$\{\xi \in A\} = \{\xi \in T^{-k}A\} = \{T^k\xi \in A\} \in \sigma\{\xi_n : n \geq k\} \subseteq \sigma\{X_n : n \geq k\}.$$

The existence of the limit in (1.7.8) is then a consequence of Theorem 1.20 □

Chapter 2

Main Results

In this chapter, we state the main results for the process X. They concern some of the basic stylized facts that we have mentioned in Chapter 1, in particular: the diffusive scaling of the distributions of log-returns (Theorem 2.2), the multiscaling of moments (Theorem 2.3 and Corollary 2.4) and the clustering of volatility (Theorem 2.5 and Corollary 2.6).

Before of proving these results, we state an easy Lemma that will be often useful in the following.

Lemma 2.1. *As $h \downarrow 0$ one has*

$$\frac{\theta(s+h)-\theta(s)}{h} \uparrow \theta'(s).$$

Proof. First, we prove that $\frac{\theta(s+h)-\theta(s)}{h}$ is increasing as h decreases. In fact, fix $\widetilde{h} \in [0,h]$; by concavity of θ one has

$$\theta(s+\widetilde{h}) \geq \theta(s) + \frac{\theta(s+h)-\theta(s)}{h} \cdot \widetilde{h},$$

that is

$$\frac{\theta(s+\widetilde{h})-\theta(s)}{\widetilde{h}} \geq \frac{\theta(s+h)-\theta(s)}{h}.$$

Moreover

$$\lim_{h\to 0} \frac{\theta(s+h)-\theta(s)}{h} = \theta'(s).$$

□

2.1 Diffusive Scaling

The first main result for X states that for small h the increments $(X_{t+h} - X_t)$ have an approximate diffusive scaling, in agreement with (1.2.1).

Theorem 2.2. *As $h \downarrow 0$ we have the convergence in distribution*

$$\frac{(X_{t+h}-X_t)}{\sqrt{h}} \xrightarrow[h\downarrow 0]{d} f(x)\,\mathrm{d}x\,, \tag{2.1.1}$$

where f is a mixture of centered Gaussian densities, namely

$$f(x) = \int_0^\infty \nu(\mathrm{d}\sigma) \int_0^\infty \mathrm{d}t\, \lambda e^{-\lambda t}\, \frac{1}{\sigma\sqrt{2\pi\theta'(\lambda t)}}\, e^{-\frac{x^2}{2\sigma^2\theta'(\lambda t)}}\,. \tag{2.1.2}$$

Proof. We observe that for all fixed $t, h > 0$ we have the equality in law $X_{t+h} - X_t \sim \sqrt{I_h}\, W_1$, as it follows by the definition of our model $(X_t)_{t\geq 0} = (W_{I_t})_{t\geq 0}$.

We also observe that $i(h) = \#\{\mathcal{T} \cap (0,h]\} \sim Po(\lambda h)$, as it follows from (1.4.3) and the properties of the Poisson process.

Since $P(i(h) \geq 1) = 1 - e^{-\lambda h} \to 0$ as $h \downarrow 0$, we may focus on the event $\{i(h) = 0\} = \{\mathcal{T} \cap (0,h] = \emptyset\}$, on which we have $I_h = \sigma_0^2(\theta(\lambda(h-\tau_0)) - \theta(-\lambda\tau_0))$, with $-\tau_0 \sim Exp(\lambda)$. Since $X_{t+h} - X_t \sim \sqrt{I_h}\, W_1$, for every Borel set A we can write as $h \downarrow 0$

$$\begin{aligned} P(X_{t+h} - X_t \in A) &= P(\sqrt{I_h}\, W_1 \in A; [0,h] \cap \mathcal{T} = \emptyset) + o(1) \\ &= \int_0^\infty \nu(\mathrm{d}\sigma) \int_0^\infty \mathrm{d}s\, \lambda e^{-\lambda s}\, P\Big(\sigma\sqrt{\theta(\lambda(h+s)) - \theta(\lambda s)}\, W_1 \in A\Big) + o(1). \end{aligned}$$

Now observe that, by Lemma 2.1, $\frac{\theta(\lambda(h+s))-\theta(\lambda s)}{\lambda h} \uparrow \theta'(\lambda s)$ as $h \downarrow 0$, for every $s > 0$, hence, plainly,

$$\frac{1}{\sqrt{\lambda h}}\, \sigma\sqrt{\theta((h+s)) - \theta(\lambda s)}\, W_1 \;\xrightarrow[h\downarrow 0]{d}\; \sigma\sqrt{\theta'(\lambda s)} W_1\,.$$

Therefore, if we choose $A = \sqrt{\lambda h}\, C$, where C is a Borel set whose boundary has zero Lebesgue measure, as $h \downarrow 0$ we have

$$P\left(\frac{X_{t+h} - X_t}{\sqrt{\lambda h}} \in C\right) = \int_0^\infty \nu(\mathrm{d}\sigma) \int_0^\infty \mathrm{d}s\, \lambda e^{-\lambda s}\, P\left(W_1 \in \frac{C}{\sigma\sqrt{\theta'(\lambda s)}}\right) + o(1).$$

Since $P(W_1 \in dx) = (2\pi)^{-1} \exp(-x^2/2)\, \mathrm{d}x$, the proof is completed. □

We stress that the function f appearing in (2.1.1)–(2.1.2), which describes the asymptotic rescaled law of the increment $(X_{t+h} - X_t)$ in the limit of small h, has a different tail behavior from the density of $(X_{t+h} - X_t)$ for fixed h. For instance, when σ has finite moments of all orders, it follows by Proposition 1.8 that the same holds for $(X_{t+h} - X_t)$. However, from (2.1.2) and a simple change of variables we get

$$\int_{\mathbb{R}} |x|^q\, f(x)\, \mathrm{d}x = \mathbb{E}[|W_1|^q]\mathbb{E}[\sigma^q] \int_0^\infty \lambda e^{-\lambda s}\, \theta'(\lambda s)^{\frac{q}{2}}\, \mathrm{d}s\,,$$

which is finite if and only if $q < q^* := (1/2 - D)^{-1}$. Therefore, independently of the law ν of σ, the density f *has always polynomial tails*: $\int_{\mathbb{R}} |x|^q f(x)\, \mathrm{d}x = \infty$ for $q \geq q^*$.

2.2 Multiscaling of the Moments

The last feature of f we underlined in the previous section, has striking consequences on the scaling behavior of the moments of the increments our model. If we set for $q \in (0, \infty)$

$$m_q(h) := \mathbb{E}[|X_{t+h} - X_t|^q]\,, \tag{2.2.1}$$

from the convergence result (2.1.1) it would be natural to guess that $m_q(h) \approx h^{q/2}$ as $h \downarrow 0$, in analogy with the Brownian motion case. However, this turns out to be true only for $q < q^*$. For $q \geq q^*$, the faster scaling $m_q(h) \approx h^{Dq+1}$ holds instead, the reason being precisely the fact that the q-moment of f is infinite for $q \geq q^*$. This transition in the scaling behavior of $m_q(h)$ goes under the name of *multiscaling of moments* and is discussed in detail, e.g., in [14]; Let us now state our result.

Theorem 2.3 (Multiscaling of moments)**.** *Let $q > 0$, and assume $\mathbb{E}[\sigma^q] < +\infty$. Then the quantity $m_q(h) := \mathbb{E}[|X_{t+h} - X_t|^q] = \mathbb{E}[|X_h|^q]$ is finite and has the following asymptotic behavior as $h \downarrow 0$:*

$$m_q(h) \sim \begin{cases} C_q\,(\lambda h)^{\frac{q}{2}} & \text{if } q < q^* \\ C_q\,(\lambda h)^{\frac{q}{2}} \log(\frac{1}{h}) & \text{if } q = q^* \\ C_q\,(\lambda h)^{Dq+1} & \text{if } q > q^* \end{cases}, \qquad \text{where } \; q^* := \frac{1}{(\frac{1}{2} - D)}\,.$$

The constant $C_q \in (0, \infty)$ is given by

$$C_q := \begin{cases} \mathbb{E}[|W_1|^q]\,\mathbb{E}[\sigma^q]\,\int_0^\infty (\theta'(s))^{\frac{q}{2}} e^{-s}\mathrm{d}s & \text{if } q < q^* \\ E[|W_1|^q]\,\mathbb{E}[\sigma^q]\,(2D)^{q/2} & \text{if } q = q^* \\ \mathbb{E}[|W_1|^q]\,\mathbb{E}[\sigma^q]\,\big[\int_0^\infty ((1+x)^{2D} - x^{2D})^{\frac{q}{2}}\,\mathrm{d}x \,+\, \frac{1}{Dq+1}\big] & \text{if } q > q^* \end{cases}, \tag{2.2.2}$$

where $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x$ denotes Euler's Gamma function.

Proof. Since $X_{t+h} - X_t \sim \sqrt{I_h}\,W_1$, we can write

$$\mathbb{E}[|X_{t+h} - X_t|^q] = \mathbb{E}[|I_h|^{q/2}|W_1|^q] = \mathbb{E}[|W_1|^q]\,\mathbb{E}[|I_h|^{q/2}] = c_q\,\mathbb{E}[|I_h|^{q/2}]\,, \tag{2.2.3}$$

where we set $c_q := \mathbb{E}[|W_1|^q]$. We therefore focus on $\mathbb{E}[|I_h|^{\frac{q}{2}}]$, that we write as the sum of three terms, that will be analyzed separately:

$$\mathbb{E}[|I_h|^{\frac{q}{2}}] = \mathbb{E}[|I_h|^{\frac{q}{2}}\,\mathbf{1}_{\{i(h)=0\}}] + \mathbb{E}[|I_h|^{\frac{q}{2}}\,\mathbf{1}_{\{i(h)=1\}}] + \mathbb{E}[|I_h|^{\frac{q}{2}}\,\mathbf{1}_{\{i(h)\geq 2\}}]\,. \tag{2.2.4}$$

For the first term in the right hand side of (2.2.4), we note that $P(i(h) = 0) = e^{-\lambda h} \to 1$ as $h \downarrow 0$ and that $I_h = \sigma_0^2(\theta(\lambda(h - \tau_0)) - \theta(-\lambda\tau_0))$ on the event $\{i(h) = 0\}$. Setting $-\tau_0 =: \lambda^{-1}S$ with $S \sim Exp(1)$, we obtain as $h \downarrow 0$

$$\mathbb{E}[|I_h|^{\frac{q}{2}}\,\mathbf{1}_{\{i(h)=0\}}] = \mathbb{E}[\sigma^q]\,\mathbb{E}\big[(\theta(S + \lambda h) - \theta(S))^{\frac{q}{2}}\big]\,\big(1 + o(1)\big)\,. \tag{2.2.5}$$

Therefore, Lemma 2.1 and the monotone convergence theorem yeld

$$\lim_{h\downarrow 0} \frac{\mathbb{E}\big[(\theta(S + \lambda h) - \theta(S))^{\frac{q}{2}}\big]}{\lambda^{\frac{q}{2}}\,h^{\frac{q}{2}}} \;=\; \mathbb{E}[(\theta'(S))^{\frac{q}{2}}] \;=\; \int_0^\infty (\theta'(s))^{\frac{q}{2}} e^{-s}\mathrm{d}s. \tag{2.2.6}$$

Recalling that $q^* := (\frac{1}{2} - D)^{-1}$, we have

$$q \;\mathsf{T}\; q^* \quad\Longleftrightarrow\quad \frac{q}{2} \;\mathsf{T}\; Dq + 1 \quad\Longleftrightarrow\quad -1 \;\mathsf{T}\; \left(D - \frac{1}{2}\right) q\,.$$

We show now that $\mathbb{E}[\theta'(S)]$ is finite for $q < q^*$: in fact, $\theta'(s)$ is asymptotic near 0 to s^{2D-1}, so

$$\int_0^\epsilon \theta'(s)e^{-s}\mathrm{d}s$$

is finite if and only if $q < q^*$, while (using the fact that θ' is decreasing since θ is concave and $\mathcal{C}^1$) we have

$$\int_\epsilon^\infty \theta'(s)e^{-s}\mathrm{d}s \leq \int_\epsilon^\infty \theta'(\epsilon)e^{-s}\mathrm{d}s < \infty,$$

that is finite.

Next we want to show that, for $q > q^*$, $\mathbb{E}\big[(\theta(S+\lambda h) - \theta(S)^{\frac{q}{2}}\big]$ gives a contribution of order h^{2D+1}, in particular

$$\lim_{h\to 0} \frac{\mathbb{E}\big[(\theta(S+\lambda h) - \theta(S)^{\frac{q}{2}}\big]}{(\lambda h)^{Dq+1}} = \int_0^\infty \big((1+x)^{2D} - x^{2D}\big)^{\frac{q}{2}}.$$

Fix $\epsilon > 0$, then

$$\int_0^\infty \frac{(\theta(s+\lambda h) - \theta(s))^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s = \int_0^\epsilon \frac{(\theta(s+\lambda h) - \theta(s))^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s + \int_\epsilon^\infty \frac{(\theta(s+\lambda h) - \theta(s))^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s,$$

and consider the second term. For $q > q^*$ it goes to zero when h goes to zero because

$$\int_\epsilon^\infty \frac{(\theta(s+\lambda h) - \theta(s))^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s = (\lambda h)^{\frac{q}{2}-(Dq+1)} \int_\epsilon^\infty \left(\frac{\theta(s+\lambda h) - \theta(s)}{\lambda h}\right)^{\frac{q}{2}} e^{-s}\,\mathrm{d}s =$$

$$(\lambda h)^{\frac{q}{2}-(Dq+1)} \int_\epsilon^\infty \left(\frac{\int_0^{\lambda h} \theta'(s+x)\mathrm{d}x}{\lambda h}\right)^{\frac{q}{2}} e^{-s}\,\mathrm{d}s \leq (\lambda h)^{\frac{q}{2}-(Dq+1)} \int_\epsilon^\infty \theta'(s)^{\frac{q}{2}}\, e^{-s}\,\mathrm{d}s \leq$$

$$(\lambda h)^{\frac{q}{2}-(Dq+1)} \theta'(\epsilon)^{\frac{q}{2}},$$

that goes to zero because $\frac{q}{2} - (Dq+1) > 0$ exactly for $q > q^*$.

For the term $\displaystyle\int_0^\epsilon \frac{(\theta(s+\lambda h) - \theta(s))^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s$ observe that, for every $\epsilon > 0$, there exist δ such that

$$(1-\delta)\, 2Ds^{2D-1} \leq \theta'(s) \leq (1+\delta)\, 2Ds^{2D-1};$$

then

$$\int_0^\epsilon \frac{(\theta(s+\lambda h) - \theta(s))^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s = \int_0^\epsilon \frac{\big(\int_0^{\lambda h} \theta'(s+x)\mathrm{d}x\big)^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s \leq$$

$$\int_0^\epsilon \frac{\big(\int_0^{\lambda h} (1+\delta)\, 2D(s+x)^{2D-1}\mathrm{d}x\big)^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s = (1+\delta) \int_0^\epsilon \frac{((s+\lambda h)^{2D} - s^{2D})^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s.$$

Analogously

$$\int_0^\epsilon \frac{((s+\lambda h)^{2D} - s^{2D})^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s \geq (1-\delta) \int_0^\epsilon \frac{((s+\lambda h)^{2D} - s^{2D})^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\, e^{-s}\,\mathrm{d}s,$$

so it only remains to show that

$$\lim_{h\to 0}\int_0^{\epsilon}\frac{((s+\lambda h)^{2D}-s^{2D})^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\,e^{-s}\,\mathrm{d}s=\int_0^{\infty}((1+x)^{2D}-x^{2D})^{\frac{q}{2}}\mathrm{d}x.$$

By the change of variables $s=(\lambda h)x$, we can write

$$\int_0^{\epsilon}((s+\lambda h)^{2D}-s^{2D})^{\frac{q}{2}}\,e^{-s}\mathrm{d}s=(\lambda h)^{Dq+1}\int_0^{\frac{\epsilon}{\lambda h}}((1+x)^{2D}-x^{2D})^{\frac{q}{2}}\,e^{-\lambda h x}\,\mathrm{d}x,\qquad(2.2.7)$$

and note that $((1+x)^{2D}-x^{2D})^{\frac{q}{2}}\sim(2D)^{\frac{q}{2}}x^{(D-\frac{1}{2})q}$ as $x\to+\infty$ and $(D-\frac{1}{2})q<-1$ if and only if $q>q^*$. Therefore, by the monotone convergence theorem we obtain

$$\text{for } q>q^*:\quad\lim_{h\to 0}\int_0^{\epsilon}\frac{((s+\lambda h)^{2D}-s^{2D})^{\frac{q}{2}}}{(\lambda h)^{Dq+1}}\,e^{-s}\,\mathrm{d}s=\int_0^{\infty}((1+x)^{2D}-x^{2D})^{\frac{q}{2}}\mathrm{d}x.\qquad(2.2.8)$$

Finally, in the case $q=q^*$ we have $((1+x)^{2D}-x^{2D})^{q^*/2}\sim(2D)^{q^*/2}\,x^{-1}$ as $x\to+\infty$ and we want to study the integral in the second term of (2.2.7). Fix an arbitrary (large) $M>0$ and note that, integrating by parts and performing a change of variables, as $h\downarrow 0$ we have

$$\begin{aligned}\int_M^{\infty}\frac{e^{-\lambda h x}}{x}\,\mathrm{d}x\;&=\;-\log M e^{-\lambda h M}\;+\;\lambda h\int_M^{\infty}(\log x)\,e^{-\lambda h x}\,\mathrm{d}x\;=\;O(1)+\int_{\lambda h M}^{\infty}\log\Big(\frac{y}{\lambda h}\Big)\,e^{-y}\,\mathrm{d}y\\&=\;O(1)+\int_{\lambda h M}^{\infty}\log\Big(\frac{y}{\lambda}\Big)\,e^{-y}\,\mathrm{d}y\;+\;\log\Big(\frac{1}{h}\Big)\int_{\lambda h M}^{\infty}e^{-y}\,\mathrm{d}y\;=\;\log\Big(\frac{1}{h}\Big)\,\big(1+o(1)\big).\end{aligned}$$

From this it is easy to see that as $h\downarrow 0$

$$\int_0^{\infty}((1+x)^{2D}-x^{2D})^{\frac{q^*}{2}}\,e^{-\lambda h x}\,\mathrm{d}x\;\sim\;(2D)^{\frac{q^*}{2}}\log\Big(\frac{1}{h}\Big)\,.$$

Moreover, noting that $Dq+1=\frac{q}{2}$ for $q=q^*$ and reasoning exactly as for the case $q>q^*$, it holds that

$$\int_{\epsilon}^{\infty}\frac{(\theta(s+\lambda h)-\theta(s))^{\frac{q}{2}}}{(\lambda h)^{Dq+1}\log(\frac{1}{h})}\,e^{-s}\,\mathrm{d}s\le\frac{1}{\log\frac{1}{h}}\,\theta'(\epsilon)^{\frac{q}{2}}\underset{h\to 0}{\longrightarrow}0,$$

and then it follows that

$$\lim_{h\downarrow 0}\frac{\mathbb{E}\big[(\theta(S+\lambda h)-\theta(S))^{\frac{q^*}{2}}\big]}{\lambda^{Dq^*+1}\,h^{\frac{q^*}{2}}\,\log(\frac{1}{h})}=(2D)^{\frac{q^*}{2}}\,.\qquad(2.2.9)$$

Recalling (2.2.3) and (2.2.5), the relations (2.2.6), (2.2.8) and (2.2.9) show that the first term in the right hand side of (2.2.4) has the same asymptotic behavior as in the statement of the theorem, except for the regime $q>q^*$ where the constant does not match (the missing contribution will be obtained in a moment).

We now focus on the second term in the right hand side of (2.2.4). Note that, conditionally on the event $\{i(h)=1\}=\{\tau_1\le h,\tau_2>h\}$, we have

$$I_h=\sigma_1^2\theta(\lambda(h-\tau_1))+\sigma_0^2\big(\theta((\lambda(\tau_1-\tau_0))-\theta(-\lambda\tau_0)\big)\sim\sigma_1^2\theta(\lambda(h-hU))+\sigma_0^2(\theta(\lambda hU+S)-\theta(S))\,,$$

where $S \sim Exp(1)$ and $U \sim U(0,1)$ (uniformly distributed on the interval $(0,1)$) are independent of σ_0 and σ_1.

Now,

$$\lim_{h\to 0} \frac{\mathbb{E}\left[|I_h|^{\frac{q}{2}} \middle| i(h)=1\right]}{(\lambda h)^{Dq}} = \int_0^1 \mathrm{d}u \int_0^\infty \left(\frac{\theta(\lambda(h-hU))}{(\lambda h)^{2D}} + \frac{\theta(\lambda hU+s)-\theta(s)}{(\lambda h)^{2D}} \right)^{\frac{q}{2}} e^{-s} \mathrm{d}s, \tag{2.2.10}$$

and we show that we can take the limit inside the expectation. In fact, since by the hypotheses $\theta(a+b)-\theta(b) \le \theta(a)$, we have

$$\frac{\theta(\lambda(h-hU))}{(\lambda h)^{2D}} + \frac{\theta(\lambda hU+s)-\theta(s)}{(\lambda h)^{2D}} \le 2\frac{\theta(\lambda h)}{(\lambda h)^{2D}},$$

and, for $h \in [0, \widetilde{h}]$,

$$\frac{\theta(\lambda h)}{(\lambda h)^{2D}} \le 1+\epsilon;$$

we can then pass the limit inside the expectation using the dominated convergence theorem.

Moreover:

$$\lim_{h\to 0} \frac{\theta(\lambda h(1-u))}{(\lambda h)^{2D}} = \lim_{h\to 0} \frac{\theta'(\lambda h(1-u))}{2D\,(\lambda h(1-u))^{2D-1}} \cdot (1-u)^{2D} = (1-u)^{2D}$$

and

$$\lim_{h\to 0} \frac{\theta(\lambda hu+s)-\theta(s)}{(\lambda h)^{2D}} = \lim_{h\to 0} \frac{\theta(\lambda hu+s)-\theta(s)}{\lambda hu} \cdot u(\lambda h)^{1-2D} = \lim_{h\to 0} u\,\theta'(s)\,(\lambda h)^{1-2D} = 0.$$

Since $P(i(h)=1) = \lambda h + o(h)$ as $h \downarrow 0$, it follows (for every $q \in (0,\infty)$)

$$\lim_{h\downarrow 0} \frac{\mathbb{E}[|I_h|^{\frac{q}{2}}\,\mathbf{1}_{\{i(h)=1\}}]}{(\lambda h)^{Dq+1}} \;=\; \mathbb{E}[\sigma_1^q]\,\mathbb{E}\big[(1-U)^{Dq}\big] \;=\; \mathbb{E}[\sigma_1^q]\,\frac{1}{Dq+1}\,. \tag{2.2.11}$$

This shows that the second term in the right hand side of (2.2.4) gives a contribution of the order h^{Dq+1} as $h \downarrow 0$. This is relevant only for $q > q^*$, because for $q \le q^*$ the first term gives a much bigger contribution of the order $h^{q/2}$ (see (2.2.6) and (2.2.9)). Recalling (2.2.3), it follows from (2.2.11) and (2.2.8) that the contribution of the first and the second term in the right hand side of (2.2.4) matches the statement of the theorem (including the constant).

It only remains to show that the third term in the right hand side of (2.2.4) gives a negligible contribution. Consider first the case $q > 2$: from the basic bound (1.6.4) and Jensen's inequality, in analogy with (1.6.5) and (1.6.6), we obtain

$$\begin{aligned}\lim_{h\to 0} \frac{\mathbb{E}\big[|I_h|^{q/2}\mathbf{1}_{\{i(h)\ge 2\}}\big]}{(\lambda h)^{Dq}} &\le \lim_{h\to 0} \frac{\theta(\lambda h)^{\frac{q}{2}}}{(\lambda h)^{Dq}}\,\mathbb{E}[\sigma^q]\mathbb{E}\big[(i(h)+1)^{q/2}\mathbf{1}_{\{i(h)\ge 2\}}\big] \\ &= \mathbb{E}[\sigma^q]\,\mathbb{E}\big[(i(h)+1)^{q/2}\mathbf{1}_{\{i(h)\ge 2\}}\big]\,, \end{aligned} \tag{2.2.12}$$

where in the last equality we have used the fact that $\lim_{h\to 0} \dfrac{\theta(\lambda h)}{(\lambda h)^{2D}} = 1$.

On the other hand, since for $\alpha \in (0,1)$ we have $(\sum_{k=1}^{\infty} x_k)^\alpha \le \sum_{k=1}^{\infty} x_k^\alpha$, when $q \le 2$ from (1.6.4) we can write, in analogy with (1.6.7),

$$\lim_{h\to 0} \frac{\mathbb{E}\big[|I_h|^{q/2}\mathbf{1}_{\{i(h)\ge 2\}}\big]}{(\lambda h)^{Dq}} \le \lim_{h\to 0} \frac{\theta(\lambda h)^{\frac{q}{2}}}{(\lambda h)^{Dq}} \mathbb{E}[\sigma^q]\mathbb{E}\big[(i(h)+1)\mathbf{1}_{\{i(h)\ge 2\}}\big] = \mathbb{E}[\sigma^q]\mathbb{E}\big[(i(h)+1)\mathbf{1}_{\{i(h)\ge 2\}}\big]. \quad (2.2.13)$$

For any fixed $a > 0$, by the Hölder inequality with $p = 3$ and $p' = 3/2$ we can write for $h \le 1$

$$\begin{aligned}\mathbb{E}\big[(i(h)+1)^a\,\mathbf{1}_{\{i(h)\ge 2\}}\big] &\le \mathbb{E}\big[(i(h)+1)^{3a}\big]^{1/3} P(i(h)\ge 2)^{2/3} \\ &\le \mathbb{E}\big[(i(1)+1)^{3a}\big]^{1/3} (1-e^{-\lambda h}-e^{-\lambda h}\lambda h)^{2/3} \le (const.)\, h^{4/3},\end{aligned}$$

because $\mathbb{E}\big[(i(1)+1)^{3a}\big] < \infty$ (recall that $i(h) \sim Po(\lambda)$) and $(1-e^{-\lambda h}-e^{-\lambda h}\lambda h) \sim \frac{1}{2}(\lambda h)^2$ as $h \downarrow 0$. Then it follows from (2.2.12) and (2.2.13) that

$$\mathbb{E}\big[|I_h|^{q/2}\,\mathbf{1}_{\{i(h)\ge 2\}}\big] \le (const.')\, h^{Dq+4/3}.$$

This shows that the contribution of the third term in the right hand side of (2.2.4) is always negligible with respect to the contribution of the second term (recall (2.2.11)). □

Corollary 2.4. *The following relation holds true:*

$$A(q) := \lim_{h\downarrow 0} \frac{\log m_q(h)}{\log h} = \begin{cases} \frac{q}{2} & \text{if } q \le q^* \\ Dq+1 & \text{if } q \ge q^* \end{cases}. \quad (2.2.14)$$

The explicit form (2.2.2) of the multiplicative constant C_q will be used in Chapter 3 for the estimation of the parameters of our model on the DJIA time series.

2.3 Volatility Autocorrelation Decay

Our last theoretical result concerns the correlations of the absolute value of two increments, a quantity which is usually called *volatility autocorrelation*. We start determining the behavior of the covariance.

Theorem 2.5. *Assume that $\mathbb{E}[\sigma^2] < \infty$. The following relation holds as $h \downarrow 0$, for all $s, t > 0$:*

$$Cov\big(|X_{s+h}-X_s|, |X_{t+h}-X_t|\big) = \frac{2}{\pi}\, e^{-\lambda|t-s|}\,\big(\phi(\lambda|t-s|)\,\lambda h + o(h)\big), \quad (2.3.1)$$

where

$$\phi(x) := Cov\Big(\sigma\sqrt{\theta'(S)}\,,\, \sigma\sqrt{\theta'(S+x)}\Big) \quad (2.3.2)$$

and $S \sim Exp(1)$ is independent of σ.

Proof. Given a Borel set $I \subseteq \mathbb{R}$, we let $\mathcal{G}_I$ denote the σ-algebra generated by the family of random variables $(\tau_k \mathbf{1}_{\{\tau_k \in I\}}, \sigma_k \mathbf{1}_{\{\tau_k \in I\}})_{k \geq 0}$. Informally, $\mathcal{G}_I$ may be viewed as the σ-algebra generated by the variables τ_k, σ_k for the values of k such that $\tau_k \in I$. From the basic property of the Poisson process and from the fact that the variables $(\sigma_k)_{k \geq 0}$ are independent, it follows that for disjoint Borel sets I, I' the σ-algebras $\mathcal{G}_I$, $\mathcal{G}_{I'}$ are independent. We set for short $\mathcal{G} := \mathcal{G}_{\mathbb{R}}$, which is by definition the σ-algebra generated by all the variables $(\tau_k)_{k \geq 0}$ and $(\sigma_k)_{k \geq 0}$, which coincides with the σ-algebra generated by the process $(I_t)_{t \geq 0}$.

We have to prove (2.3.1). Plainly, by translation invariance we can set $s = 0$ without loss of generality. We also assume that $h < t$. We start writing

$$\begin{aligned} &Cov(|X_h|, |X_{t+h} - X_t|) \\ &\quad = Cov\big(\mathbb{E}\big[|X_h| \big| \mathcal{G}\big]\,, \mathbb{E}\big[|X_{t+h} - X_t| \big| \mathcal{G}\big]\big) \;+\; \mathbb{E}\big[Cov\big(|X_h|, |X_{t+h} - X_t| \big| \mathcal{G}\big)\big]\,. \end{aligned} \tag{2.3.3}$$

We recall that $X_t = W_{I_t}$ and the process $(I_t)_{t \geq 0}$ is $\mathcal{G}$-measurable and independent of the process $(W_t)_{t \geq 0}$. It follows that, conditionally on $(I_t)_{t \geq 0}$, the process $(X_t)_{t \geq 0}$ has independent increments, hence the second term in the right hand side of (2.3.3) vanishes, because $Cov(|X_h|, |X_{t+h} - X_t| | \mathcal{G}) = 0$ a.s.. For fixed h, from the equality in law $X_h = W_{I_h} \sim \sqrt{I_h} W_1$ it follows that $\mathbb{E}[|X_h| | \mathcal{G}] = c_1 \sqrt{I_h}$, where $c_1 = \mathbb{E}[|W_1|] = \sqrt{2/\pi}$. Analogously $\mathbb{E}[|X_{t+h} - X_t| | \mathcal{G}] = \sqrt{I_{t+h} - I_t}$ and (2.3.3) reduces to

$$Cov(|X_h|, |X_{t+h} - X_t|) \;=\; \frac{2}{\pi}\, Cov\big(\sqrt{I_h}, \sqrt{I_{t+h} - I_t}\big)\,. \tag{2.3.4}$$

Recall the definitions (1.4.3) and (1.4.4) of the variabes $i(t)$ and I_t. We now claim that we can replace $\sqrt{I_{t+h} - I_t}$ by $\sqrt{I_{t+h} - I_t}\, \mathbf{1}_{\{\mathcal{T} \cap (h,t] = \emptyset\}}$ in (2.3.4). In fact from (1.4.4) we can write

$$I_{t+h} - I_t = \sigma^2_{i(t+h)} \theta((\lambda(t + h - \tau_{i(t+h)})) + \sum_{k=i(t)+1}^{i(t+h)} \sigma^2_{k-1} \theta(\lambda(\tau_k - \tau_{k-1})) - \sigma^2_{i(t)} \theta(\lambda(t - \tau_{i(t)}))\,,$$

where we agree that the sum in the right hand side is zero if $i(t + h) = i(t)$. This shows that $(I_{t+h} - I_t)$ is a function of the variables τ_k, σ_k with index $i(t) \leq k \leq i(t + h)$. Since $\{\mathcal{T} \cap (h, t] \neq \emptyset\} = \{\tau_{i(t)} > h\}$, this means that $\sqrt{I_{t+h} - I_t}\, \mathbf{1}_{\{\mathcal{T} \cap (h,t] \neq \emptyset\}}$ is $\mathcal{G}_{(h,t+h]}$-measurable, hence independent of $\sqrt{I_h}$, which is clearly $\mathcal{G}_{(-\infty,h]}$-measurable. This shows that $Cov(\sqrt{I_h}, \sqrt{I_{t+h} - I_t}\, \mathbf{1}_{\{\mathcal{T} \cap (h,t] \neq \emptyset\}}) = 0$, therefore from (2.3.4) we can write

$$Cov(|X_h|, |X_{t+h} - X_t|) \;=\; \frac{2}{\pi}\, Cov\big(\sqrt{I_h}, \sqrt{I_{t+h} - I_t}\, \mathbf{1}_{\{\mathcal{T} \cap (h,t] = \emptyset\}}\big)\,. \tag{2.3.5}$$

It is convenient to define

$$j(t) \;:=\; \inf\{k \in \mathbb{Z} : \tau_k > t\} \;=\; i(t) + 1\,,$$

and note that the variables $\tau_{j(t)}$, $\tau_{j(t)+1}$, ... are $\mathcal{G}_{(t,\infty)}$-measurable. We now introduce a variable $A_{t,h}$, which is essentially $I_{t+h} - I_t$ with $\tau_{i(t)}$ replaced by $\tau_{i(h)}$. More precisely, we set

$$A_{t,h} \;:=\; \sigma^2_{i(h)} \big[\theta((\lambda t + h - \tau_{i(h)})) - \theta(\lambda(t - \tau_{i(h)}))\big] \qquad \text{if } \mathcal{T} \cap (t, t + h] = \emptyset\,, \tag{2.3.6}$$

and

$$\begin{aligned} A_{t,h} \;:=\; & \sigma^2_{i(t+h)}\theta(\lambda(t+h-\tau_{i(t+h)})) + \sum_{k=j(t)+1}^{i(t+h)} \sigma^2_{k-1}\theta(\lambda(\tau_k-\tau_{k-1})) \\ & + \sigma^2_{i(h)}\big[\theta(\lambda(\tau_{j(t)}-\tau_{i(h)})) - \theta(\lambda(t-\tau_{i(h)}))\big] \qquad \text{if } \mathcal{T}\cap(t,t+h]\neq\emptyset\,, \end{aligned} \tag{2.3.7}$$

where we agree that the sum in the right hand side of (2.3.7) is zero unless $i(t+h) > j(t)$, that is $\#(\mathcal{T}\cap(t,t+h]) \geq 2$. We stress that $A_{t,h}$ is $\mathcal{G}_{(-\infty,h]\cup(t,t+h]}$-measurable: in fact, $A_{t,h}$ is nothing but $I_{t+h} - I_t$ when we erase from $\mathcal{T}$ the points τ_k (if any) falling in $(h,t]$. In particular, on the event $\{\mathcal{T}\cap(h,t]=\emptyset\}$ we can replace $I_{t+h}-I_t$ by $A_{t,h}$, rewriting (2.3.5) as

$$\begin{aligned} Cov(|X_h|,|X_{t+h}-X_t|) \;&=\; \frac{2}{\pi}\,Cov\big(\sqrt{I_h},\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap(h,t]=\emptyset\}}\big) \\ &=\; \frac{2}{\pi}\,\mathbb{E}\big[\big(\sqrt{I_h}-\mathbb{E}[\sqrt{I_h}]\big)\cdot\sqrt{A_{t,h}}\cdot\mathbf{1}_{\{\mathcal{T}\cap(h,t]=\emptyset\}}\big] \\ &=\; \frac{2}{\pi}\,e^{-\lambda(t-h)}\,\mathbb{E}\big[\big(\sqrt{I_h}-\mathbb{E}[\sqrt{I_h}]\big)\cdot\sqrt{A_{t,h}}\,\big]\,, \end{aligned} \tag{2.3.8}$$

where in the last equality we have used the independence of the σ-algebras $\mathcal{G}_{(-\infty,h]\cup(t,\infty)}$ and $\mathcal{G}_{(h,t]}$ and the fact that $P(\mathcal{T}\cap(h,t]=\emptyset) = P(i(t)-i(h)=0) = e^{-\lambda(t-h)}$.

We now focus on

$$\mathbb{E}\big[\big(\sqrt{I_h}-\mathbb{E}[\sqrt{I_h}]\big)\cdot\sqrt{A_{t,h}}\,\big] \;=\; Cov\big(\sqrt{I_h}\,,\,\sqrt{A_{t,h}}\,\big)\,, \tag{2.3.9}$$

that we can write as the sum of four terms:

$$\begin{aligned} Cov\big(\sqrt{I_h}\,,\,\sqrt{A_{t,h}}\,\big) \;=\; & Cov\big(\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,,\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]=\emptyset\}}\big) \\ & +\; Cov\big(\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]\neq\emptyset\}}\,,\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]=\emptyset\}}\big) \\ & +\; Cov\big(\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,,\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big) \\ & +\; Cov\big(\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]\neq\emptyset\}}\,,\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big)\,. \end{aligned} \tag{2.3.10}$$

The first term in the right hand side gives the leading contribution:

$$\begin{aligned} & Cov\big(\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,,\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]=\emptyset\}}\big) \;= \\ & Cov\big(\sigma_0\sqrt{\theta(\lambda(h-\tau_0))-\theta(-\lambda\tau_0)}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,,\,\sigma_0\sqrt{\theta(\lambda(t+h-\tau_0))-\theta(\lambda(t-\tau_0))}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]=\emptyset\}}\big) \\ & =\; e^{-2\lambda h}\,Cov\big(\sigma_0\sqrt{\theta(\lambda h+S)-\theta(S)}\,,\,\sigma_0\sqrt{\theta(\lambda t+\lambda h+S)-\theta(\lambda t+S)}\big)\,, \end{aligned}$$

where $S := \lambda(-\tau_0) \sim Exp(1)$ and we have used the fact that the variables τ_0, $\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]=\emptyset\}}$ and $\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}$ are independent. Since by Lemma 2.1 $\delta^{-1}(\theta(\delta+x)-\theta(x)) \uparrow \theta'(x)$ as $\delta\downarrow 0$, by monotone convergence we obtain

$$\frac{1}{\lambda h}\,Cov\big(\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,,\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]=\emptyset\}}\big) \;\xrightarrow{h\downarrow 0}\; Cov\big(\sigma\,\sqrt{\theta'(S)},\sigma\,\sqrt{\theta'(S+\lambda t)}\big)\,,$$

in agreement with (2.3.1) and (2.3.2).

It remains to show that the second, the third and the fourth term in the right hand side of (2.3.10) give a negligible contribution.

For the term $Cov\big(\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,,\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big)$ we can procede as follows: by triangular inequality

$$|(Cov\big(\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,,\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big)|\leq\mathbb{E}\big[\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big]$$
$$+\mathbb{E}\big[\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\big]\cdot\mathbb{E}\big[\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big]$$

and both terms are negligible as we now prove.
Recalling (1.6.2), $\theta(a+b)-\theta(b)\leq\theta(a)$ for all $a,b\geq 0$, we have

$$\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\leq\sqrt{\theta(\lambda h)}\,\sigma_0\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}$$

and

$$\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\leq\sqrt{\theta(\lambda h)}\,\sqrt{\sum_{k=j(t)}^{i(t+h)}\sigma_{k-1}^2}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}.$$

Using the independence of $\mathcal{T}$ and the σ_k it follows that

$$\mathbb{E}\big[\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big]\leq e^{-\lambda h}\,\theta(\lambda h)\,\mathbb{E}\big[\sigma_0\big]\,\mathbb{E}\left[\sqrt{\sum_{k=j(t)}^{i(t+h)}\sigma_{k-1}^2}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\right]$$
$$\leq e^{-\lambda h}\,\theta(\lambda h)\,\mathbb{E}\big[\sigma_0\big]\,\mathbb{E}\left[\sum_{k=j(t)}^{i(t+h)}\sigma_{k-1}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\right]$$
$$=e^{-\lambda h}\,\theta(\lambda h)\,\mathbb{E}\big[\sigma_0\big]^2\,\mathbb{E}\left[(\eta_{[t,t+h]}+1)\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\right],$$

where $\eta_{[t,t+h]}$ is the cardinality of $\mathcal{T}\cap[t,t+h]$, that is strictly positive on the event $\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}$. By the fact that

$$\mathbb{E}\big[(\eta_{[t,t+h]}+1)\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big]\leq\mathbb{E}\big[\eta_{[t,t+h]}\big]+P\big(\mathcal{T}\cap[t,t+h]\neq\emptyset\big)\leq 2\lambda h$$

we have that, for $h\downarrow 0$,

$$\mathbb{E}\big[\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big]\leq(const.)\;h^{2D+1}.$$

A totally similar reasonament can be apply to $\mathbb{E}\big[\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\big]\cdot\mathbb{E}\big[\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big]$ and so we proved that $|Cov\big(\sqrt{I_h}\,\mathbf{1}_{\{\mathcal{T}\cap[0,h]=\emptyset\}}\,,\,\sqrt{A_{t,h}}\,\mathbf{1}_{\{\mathcal{T}\cap[t,t+h]\neq\emptyset\}}\big)|$ is negligible. In a similar way one can prove that the other terms too give a negligible contribution. □

We recall that $\rho(Y,Z):=Cov(Y,Z)/\sqrt{Var(Y)Var(Z)}$ is the correlation coefficient of two random variables Y,Z. As Theorem 2.3 yields

$$\lim_{h\downarrow 0}\frac{1}{\lambda h}\,Var\big(|X_{t+h}-X_t|\big)\;=\;Var\big(\sigma\,|W_1|\,\sqrt{\theta'(S)}\big)\,,$$

where $S\sim Exp(1)$ is independent of σ,W_1, we easily obtain the following result.

Corollary 2.6 (Volatility autocorrelation)**.** *Assume that* $\mathbb{E}[\sigma^2] < \infty$*. The correlation of the increments of the process* X *has the following asymptotic behavior as* $h \downarrow 0$*:*

$$\lim_{h\downarrow 0} \rho(|X_{s+h} - X_s|, |X_{t+h} - X_t|) \;=\; \rho(t-s) \;:=\; \frac{2}{\pi\, Var\big[\sigma\,|W_1|\,\sqrt{\theta'(S)}\big]}\, e^{-\lambda|t-s|}\, \phi(\lambda|t-s|)\,, \tag{2.3.11}$$

where $\phi(\cdot)$ *is defined in* (2.3.2) *and* $S \sim Exp(1)$ *is independent of* σ *and* W_1*.*

This shows that the volatility autocorrelation of our process decays exponentially fast for time scales greater than the mean distance $1/\lambda$ between the epochs τ_k. For shorter time scales, a relevant contribution is given by the function $\phi(\cdot)$.

Note that, in the case $\theta(t) = t^{2D}$ (that will be the basic case we will study for the numerical applications of the model), by (2.3.2) we can write

$$\phi(x) \;=\; 2D\big[Var(\sigma)\,\mathbb{E}[S^{D-1/2}\,(S+x)^{D-1/2}] + \mathbb{E}[\sigma]^2\, Cov(S^{D-1/2}, (S+x)^{D-1/2})\big]\,, \tag{2.3.12}$$

where $S \sim Exp(1)$. As $x \to \infty$, the two terms in the right hand side decay as

$$\mathbb{E}[S^{D-1/2}\,(S+x)^{D-1/2}] \;\approx\; x^{D-1/2}\,, \qquad Cov(S^{D-1/2}, (S+x)^{D-1/2}) \;\approx\; x^{D-3/2}\,,$$

while for $x = O(1)$ the decay of both terms is faster than polynomial but slower than exponential (see Figures 3.2(A) and 3.2(B) in Chapter 3).

Chapter 3

Parameter Estimation & Data Fitting

We now consider some aspects of our model from a numerical viewpoint. We compare the theoretical predictions and the simulated data of our model with the time series of the Dow Jones Industrial Average (DJIA) and of the Standard & Poor's 500 (S&P 500), the first over a period of 75 years and the second over a period of 61 years. Particurlarly, in our analysis we focused on the DJIA index, for which we have considered the opening prices from 2 Jan 1935 to 31 Dec 2009, for a total of 18849 daily data. Of course, other indexes would be interesting as well, but we have decided to stick to the DJIA and the S&P both for their importance and their considerable time lengths.

The data analysis, the simulations and the plots have been obtained with the software R [25]. The code we have used is available on the web page `http://www.matapp.unimib.it/~fcaraven/c.html`.

3.1 Preliminary Considerations

For the numerical comparison of our process $(X_t)_{t\geq 0}$ with the empirical time series, we have decided to focus on the following quantities:

(a) The *multiscaling of moments*, cf. Corollary 2.4.

(b) The *volatility autocorrelation decay*, cf. Corollary 2.6.

(c) The *distribution* of X_t.

(d) The *local volatility*.

Roughly speaking, the idea is to compute *empirically* these quantities on the time series and then to compare the results with the *theoretical* predictions of our model. This is justified by the ergodicity of the increments of our process $(X_t)_{t\geq 0}$, cf. equation (1.7.8).

For the numerical analysis, we used the following two easy versions of the model:

- the first with $\theta(t) = t^{2D}$
- the second with $\theta(t) := \begin{cases} t^{2D} & \text{if } 0 < t < C \\ C^{2D} + 2\,D\,C^{2D-1}\,(t - C) & \text{if } t \geq C \end{cases}$,

with C a scalar parameter. Note that for $C = \infty$ we obtain just the first version of the model.

The first problem that one faces is the *estimation of the parameters* of our model: the three scalars $\lambda \in (0, \infty)$, $D \in (0, \frac{1}{2}]$, $C \in (0, \infty]$ (where for $C = \infty$ we mean to consider the first version of the model) and the *distribution* ν *of* σ. This last one, in principle, belongs to an infinite dimensional space, but we initially focus on the moments $\mathbb{E}[\sigma]$ and $\mathbb{E}[\sigma^2]$ only. We will later show that, indeed, just to determine these quantities is enough for our purposes.

In order to estimate $(D, \lambda, \mathbb{E}[\sigma], \mathbb{E}[\sigma^2], C)$, we take into account four significant quantities that depend only on these parameters: the multiscaling coefficients C_1 and C_2 (see (2.2.2)), the multiscaling exponent $A(q)$ (see (2.2.14)) and the volatility autocorrelation function $\rho(t)$ (see (2.3.11)).

Our first try was, as in [12], to determine these parameters equation by equation, in order to obtain a really easy and fast algorithm for the estimation of the model. This method, really fast indeed, has the drawback of producing an estimated model whose fitting with data is not fully satisfactory; so, our second try was to consider a natural loss functional $\mathcal{L} = \mathcal{L}(D, \lambda, \mathbb{E}[\sigma], \mathbb{E}[\sigma^2], C)$, see (3.2.3) below, which describes some *global* distance between these theoretical quantities and the corresponding empirical ones (evaluated on the chosen time series). We then define the estimator for $(D, \lambda, \mathbb{E}[\sigma], \mathbb{E}[\sigma^2], C)$ as the point at which $\mathcal{L}$ attains its overall minimum, subject to the constraint $\mathbb{E}[\sigma^2] \geq (\mathbb{E}[\sigma])^2$. The algorithm obtained in this way gives a much better estimation then the prevoius one, although it is considerably slower and raises some numerical problems. Since fixing a value for C greatly simplify this task and produced more reliable results, we decided to content ourselves fixing some values for C and then solving the optimization problem for C fixed.

Before describing in detail the procedure we have just sketched, let us fix some notation: we will focus only on the DJIA time series, that will be denoted by $(s_i)_{0 \leq i \leq N}$ (where $N = 18848$) and the corresponding detrended log-DJIA time series will be denoted by $(x_i)_{0 \leq i \leq N}$:

$$x_i := \log(s_i) - \overline{d}(i)\,,$$

where $\overline{d}(i) := \frac{1}{250} \sum_{k=i-250}^{i-1} \log(s_i)$ is the mean log-DJIA price on the previous 250 days. Of course, other reasonable choices for $\overline{d}(i)$ are possible, but they affect the analysis only in a minor way.

3.2 Estimation of the Parameters

As said above, we consider C fixed, and in particular we will do the numerical analysis for the values $C = 0.4, 0.6, \infty$. Values of C much smaller than 0.4 turn out to be unsatisfactory, since in this regime the model behaves nearly as a Brownian motion with piecewise constant volatility. On the other hand, values of C much higher than 0.6 are nearly indistinguishable from the $C = \infty$ case. Of course, a more refined mesh of values would have produced better

results, but it would have required an eccessively high computational time. The theoretical scaling exponent $A(q)$ is defined in (2.2.14) while the multiscaling constants C_1 and C_2 are given by (2.2.2) for $q = 1$ and $q = 2$. Since $q^* = (\frac{1}{2} - D)^{-1} > 2$ (we recall that $0 \le D \le \frac{1}{2}$), we can write more explicitly, for $C \neq \infty$

$$C_1 = \frac{2\sqrt{D}\,\mathbb{E}[\sigma]\,\left(\gamma\left(\frac{1}{2}+D, C\right) + C^{D-\frac{1}{2}}e^{-C}\right)}{\sqrt{\pi}},$$

$$C_2 = 2D\,\mathbb{E}[\sigma^2]\,\left(\gamma(2D, C) + C^{2D-1}e^{-C}\right),$$

where $\gamma(a, b)$ is the lower incomplete gamma function (that is, $\gamma(a, b) = \int_0^b t^{a-1}e^{-t}\mathrm{d}t$); for $C = \infty$ these expressions become

$$C_1 = \frac{2\sqrt{D}\,\Gamma(\frac{1}{2}+D)\,\mathbb{E}[\sigma]}{\sqrt{\pi}}, \qquad C_2 = 2D\,\Gamma(2D)\,\mathbb{E}[\sigma^2]\,. \tag{3.2.1}$$

Defining the corresponding empirical quantities requires some care, because the DJIA data are in discrete-time and therefore no $h \downarrow 0$ limit is possible. We first evaluate the empirical q-moment $\widehat{m}_q(h)$ of the DJIA log-returns over h days, namely

$$\widehat{m}_q(h) := \frac{1}{N+1-h}\sum_{i=0}^{N-h} |x_{i+h} - x_i|^q\,.$$

By Theorem 2.3, the relation $\log \widehat{m}_q(h) \sim A(q)(\log h) + \log(C_q)$ should hold for h small; plotting $(\log \widehat{m}_q(h))$ versus $(\log h)$ one finds indeed an approximate linear behavior, for moderate values of h and when q is not too large ($q \;.\; 5$). By a standard linear regression of $(\log \widehat{m}_q(h))$ versus $(\log h)$ for $h = 1, 2, 3, 4, 5$ days we therefore determine the empirical values of $A(q)$ and C_q on the DJIA time series, that we call $\widehat{A}(q)$ and $\widehat{C}_q$.

For what concerns the theoretical volatility autocorrelation, Corollary 2.6 and the stationarity of the increments of our process $(X_t)_{t\ge 0}$ yield, for the case $C = \infty$,

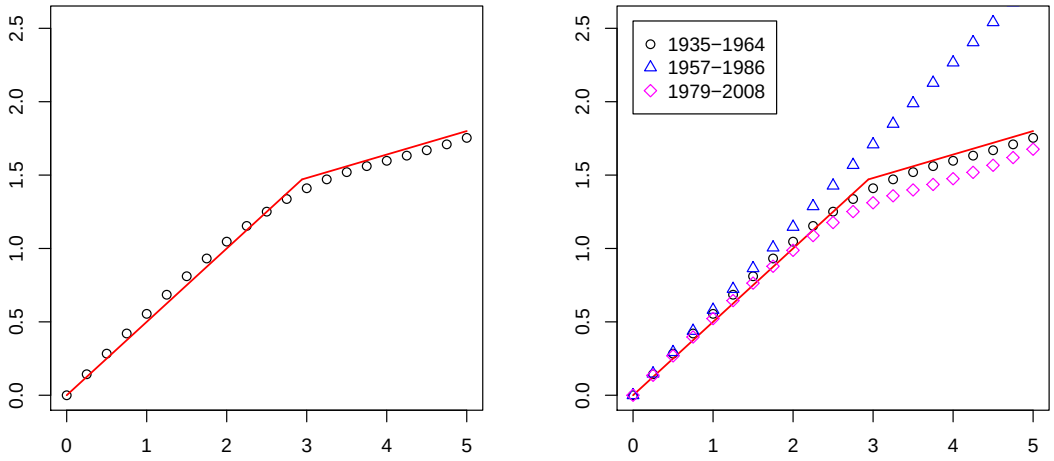
$$\rho(t) \;:=\; \lim_{h\downarrow 0}\rho(|X_h|, |X_{t+h} - X_t|) \;=\; \frac{2}{\pi\, Var(\sigma\,|W_1|\,S^{D-1/2})}\,e^{-\lambda t}\,\phi(\lambda t)\,, \tag{3.2.2}$$

where $S \sim Exp(1)$ is independent of σ and W_1 and where the function $\phi(\cdot)$ is given by

$$\phi(x) \;=\; Var(\sigma)\,E[S^{D-1/2}\,(S+x)^{D-1/2}] \;+\; \mathbb{E}[\sigma]^2\,Cov(S^{D-1/2}, (S+x)^{D-1/2})\,,$$

cf. (2.3.12). Note that, although $\phi(\cdot)$ does not admit an explicit expression, it can be quite easily evaluated numerically; for the case $C \neq \infty$, $\rho(t)$ has a more complicated expression, but still quite easily numerically evaluable (see Remark 3.1 for more details).

For the analogous empirical quantity, we define the empirical DJIA volatility autocorrelation $\widehat{\rho}_h(t)$ over h-days as the *sample correlation coefficient* of the two sequences $(|x_{i+h} - x_i|)_{0\le i\le N-h-t}$ and $(|x_{i+h+t} - x_{i+t}|)_{0\le i\le N-h-t}$. Since no $h \downarrow 0$ limit can be taken on discrete data, we are going to compare $\rho(t)$ with $\widehat{\rho}_h(t)$ for $h = 1$ day.



(A) *Multiscaling in the DJIA (1935-2009).* (B) *Multiscaling in subperiods of 30 years.*

FIGURE 3.1: *Multiscaling of moments in the DJIA time series (1935-2009).*
(A) The empirical scaling exponent $\widehat{A}(q)$ in the period 1935-2009 (circles) and the theoretical scaling exponent $A(q)$ (line) as a function of q.
(B) The empirical scaling exponent $\widehat{A}(q)$ evaluated in subperiods of 30 years (see legend) and the theoretical scaling exponent $A(q)$ (line) as a function of q.
Note. The theoretical scaling exponent $A(q)$ is evaluated for $D = 0.16$, cf. (3.2.4),(3.2.5),(3.2.6).

We can now define, for C fixed, a *loss functional* $\mathcal{L}_C$ as follows:

$$\begin{aligned}\mathcal{L}_C(D,\lambda,\mathbb{E}[\sigma],\mathbb{E}[\sigma^2]) \;&= \alpha\frac{1}{2}\left\{\left(\frac{\widehat{C}_1}{C_1}-1\right)^2+\left(\frac{\widehat{C}_2}{C_2}-1\right)^2\right\}+\beta\frac{1}{20}\sum_{k=1}^{20}\left(\frac{\widehat{A}(k/4)}{A(k/4)}-1\right)^2\\ &\quad+\delta\sum_{n=1}^{400}\frac{e^{-n/T}}{\left(\sum_{m=1}^{400}e^{-m/T}\right)}\left(\frac{\widehat{\rho}_1(n)}{\rho(n)}-1\right)^2,\end{aligned} \tag{3.2.3}$$

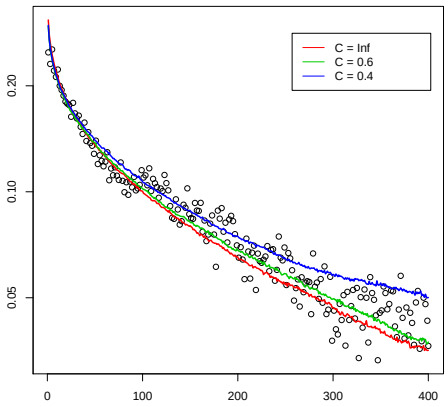
where α,β,δ are weights to be decided and the constant T controls a discount factor in long-range correlations. We fix $T = 40$ (days) and we define the estimator $(\widehat{D},\widehat{\lambda},\widehat{\mathbb{E}[\sigma]},\widehat{\mathbb{E}[\sigma^2]})$ of the parameters of our model as the point where the functional $\mathcal{L}_C$ attains its overall minimum, that is

$$(\widehat{D},\widehat{\lambda},\widehat{\mathbb{E}[\sigma]},\widehat{\mathbb{E}[\sigma^2]}) := \underset{\substack{D\in(0,\frac{1}{2}],\ \lambda,\mathbb{E}[\sigma],\mathbb{E}[\sigma^2]\in(0,\infty)\\ \text{such that } \mathbb{E}[\sigma^2]\geq(\mathbb{E}[\sigma])^2}}{\arg\min}\left\{\mathcal{L}_C(D,\lambda,\mathbb{E}[\sigma],\mathbb{E}[\sigma^2])\right\},$$

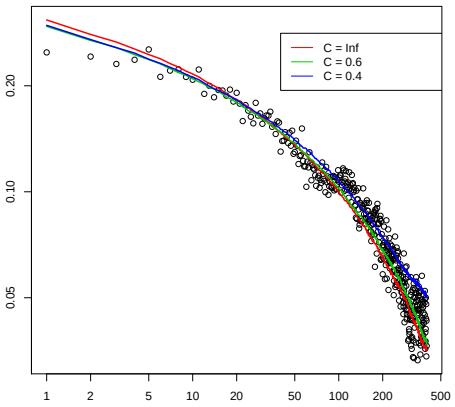
where the constraint $\mathbb{E}[\sigma^2] \geq (\mathbb{E}[\sigma])^2$ is due to $Var(\sigma) = \mathbb{E}[\sigma^2] - (\mathbb{E}[\sigma])^2 \geq 0$. We expect that such an estimator has good properties, such as asymptotic consistency and normality.

We have then proceeded to the numerical study of the functional $\mathcal{L}_C$, which appears to be quite regular.

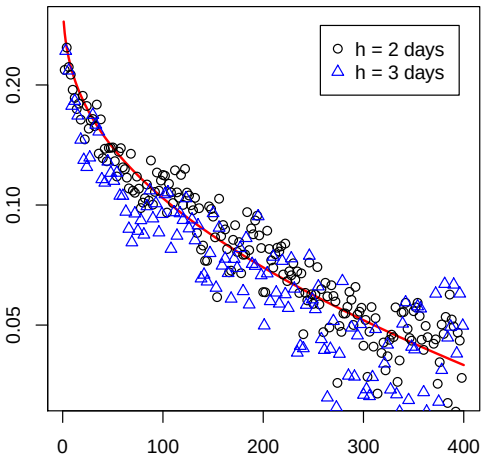
Remark 3.1. The numerical study of $\mathcal{L}_C$ is anyway a non-trivial problem. First of all, this functional is still too hard to be numerically minimized by a software like Mathematica. So,



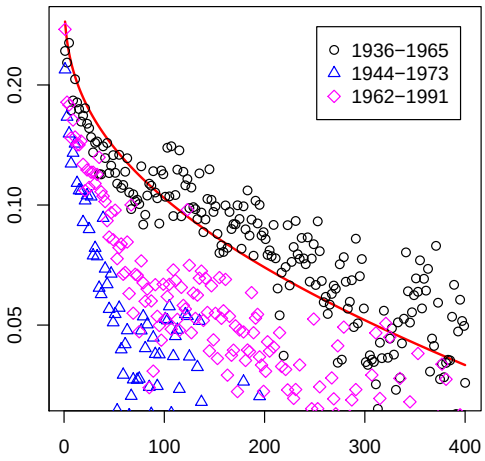
(A) *Volatility autocorrelation in the DJIA (1935-2009): log plot.*



(B) *Volatility autocorrelation in the DJIA (1935-2009): log-log plot.*



(C) *Volatility autocorrelation over h days.*



(D) *Volatility autocorrelation in subperiods of 30 years.*

FIGURE 3.2: *Volatility autocorrelation decay in the DJIA time series (1935-2009).*
(A) Log plot for the empirical 1-day volatility autocorrelation $\widehat{\rho}_1(t)$ in the period 1935-2009 (circles) and the theoretical prediction $\rho(t)$ (line), as functions of t (days).
(B) Same as (A), but log-log plot instead of log plot.
(C) The empirical h-day volatility autocorrelation $\widehat{\rho}_h(t)$ in the period 1935-2009 for $h = 2$ and $h = 3$ days (see legend) and the theoretical prediction $\rho(t)$ (line) as a function of t (days).
(D) The empirical 1-day volatility autocorrelation $\widehat{\rho}_1(t)$ evaluated in subperiods of 30 years (see legend) and the theoretical prediction $\rho(t)$ (line) as a function of t (days). For clarity, only one data out of two is plotted.
Note. The theoretical curve $\rho(t)$ is evaluated for the parameters values given in (3.2.4), (3.2.5), (3.2.6) in (A) and (B), while is evaluated only for the parameters given in (3.2.4) for (C) and (D)

we constructed a mesh of values for D, not so refined (we chose a step of 0.01, that due to the regularity of $\mathcal{L}_C$ is small enough for obtaining a good precision) and we used the software Mathematica [23] to minimize this function for each value D and we finally took the minimum overall D. It has to be remarked that, besides the regularity of the function and all these simplifications, we detected many false minima given by the program. The term that gives more problems is clearly $\rho(t)$, that has not an explicit expression; so, while a single value of $\rho(t)$ needs only some istants to be obtained, the calculation of the whole curve requires up to a minute, and this greatly increases the time needed to obtain a minimum. Moreover, the method used by the software to find the minimum is quite instable, giving very different results using slightly different weights. To try and overcome this last problem, we gave more weight to the curve $\rho(t)$, so that the program uses more precision handling this term; however, to give great importance to the correlation decay is a rather natural choice, because being a curve and not a single scalar parameter, it gives a lot of information about the empirical time series.

An alternative method for the minimization of $\mathcal{L}_C$ is to directly use a mesh of values for all the parameters and do a global minimization over this grid; again thanks to the regularity of $\mathcal{L}_C$, this mesh need not to be too refined for obtaining good results. Anyway, one should have at least an idea of the right range for the parameters, otherwise the mesh would be too large, so the best thing is probably to combine this two methods: first to use a software to have an idea of the right range for the parameters, and then to use a grid to obtain their right estimation.

With the help of the software Mathematica [23], we have finally obtained the following estimates for the parameters.

$$C = \infty :$$

$$\widehat{D} \simeq 0.16, \quad \widehat{\lambda} \simeq 0.00097, \quad \widehat{\mathbb{E}[\sigma]} \simeq 0.328, \quad \widehat{\mathbb{E}[\sigma^2]} \simeq 0.1078, \quad \frac{\sqrt{\widehat{Var(\sigma)}}}{\widehat{\mathbb{E}[\sigma]}} \simeq 0.045. \quad (3.2.4)$$

$$C = 0.6 :$$

$$\widehat{D} \simeq 0.16, \quad \widehat{\lambda} \simeq 0.0009, \quad \widehat{\mathbb{E}[\sigma]} \simeq 0.313, \quad \widehat{\mathbb{E}[\sigma^2]} \simeq 0.1049, \quad \frac{\sqrt{\widehat{Var(\sigma)}}}{\widehat{\mathbb{E}[\sigma]}} \simeq 0.263. \quad (3.2.5)$$

$$C = 0.4 :$$

$$\widehat{D} \simeq 0.16, \quad \widehat{\lambda} \simeq 0.0013, \quad \widehat{\mathbb{E}[\sigma]} \simeq 0.243, \quad \widehat{\mathbb{E}[\sigma^2]} \simeq 0.067, \quad \frac{\sqrt{\widehat{Var(\sigma)}}}{\widehat{\mathbb{E}[\sigma]}} \simeq 0.353. \quad (3.2.6)$$

Remark 3.2. Note that for all the choices of C we made, we obtain the same estimation for D; this is in agreement with the fact that the theoretical multiscaling exponent depends on the choice of θ only for this parameter, so all these versions of the model will have the same curve for $A(q)$.

Moreover, the estimated values for λ also have relatively small variations in the three cases, so that the estimated mean distance between the successive epochs $(\tau_n)_{n\geq 0}$ of the Poisson process τ is $\frac{1}{\widehat{\lambda}} \simeq 1000$ days. As a consequence, detailed properties of the distribution of σ are not expected to be detectable from data. In fact, in a time period of the length of the DJIA time series we are considering, only $18849/1000 \simeq 19$ variables σ_k are expected

to be sampled, which is certainly not enough to allow more than a rough estimation of the distribution of σ. This should be viewed more as a robustness than a limitation of our model: even when $\frac{\sqrt{\widehat{Var(\sigma)}}}{\widehat{\mathbb{E}[\sigma]}}$ is significantly greater than 0, the first two moments of σ contain the information which is relevant for application to real data.

Remark 3.3. For the case $C = \infty$ we can note that the estimated σ is nearly constant, while this is no more true for $C \neq \infty$. This is linked to the fact that for the model with $C \neq \infty$ the correlations decay faster than for the model with $C = \infty$ (with the same parameters), so a non-constant σ is needed to balance this behaviour, because as the variance of σ increases the theoretical volatility autocorrelation decay slows down. This will have important consequences for the volatility curve of the process, as we will see in subsection 3.4.

Using these estimated values of the parameters, we find a very good agreement between the empirical and the theoretical quantities we have considered:

- the empirical values of the multiscaling constants are $C_1 \simeq 6.62 \cdot 10^{-3}$, $C_2 \simeq 9.23 \cdot 10^{-5}$, while the theoretical values are respectevely:

  (1) $\widehat{C}_1 \sqrt{\widehat{\lambda}} \simeq 6.30 \cdot 10^{-3}$, $\quad \widehat{C}_2 \, \widehat{\lambda} \simeq 9.36 \cdot 10^{-5}$, $\quad$ for the case $C = \infty$,

  (2) $\widehat{C}_1 \sqrt{\widehat{\lambda}} \simeq 6.44 \cdot 10^{-3}$, $\quad \widehat{C}_2 \, \widehat{\lambda} \simeq 9.37 \cdot 10^{-5}$, $\quad$ for the case $C = 0.6$,

  (3) $\widehat{C}_1 \sqrt{\widehat{\lambda}} \simeq 6.43 \cdot 10^{-3}$, $\quad \widehat{C}_2 \, \widehat{\lambda} \simeq 9.42 \cdot 10^{-5}$, $\quad$ for the case $C = 0.4$;

- the theoretical and empirical multiscaling exponents $A(q)$ and $\widehat{A}(q)$ are remarkably close, see Figure 3.1(A) for a graphical comparison;

- the fit is very good also for the volatility autocorrelation functions $\rho(t)$ and $\widehat{\rho}_1(t)$, see Figures 3.2(A) (log plot) and 3.2(B) (log-log plot) for a graphical comparison, which also shows that the decay is faster than polynomial but slower than exponential. From Figure 3.2(C) one sees that the agreement between $\rho(t)$ and $\widehat{\rho}_h(t)$ is still good also for $h = 2$ and $h = 3$ days.

We point out that, evaluating the empirical multiscaling exponent $\widehat{A}(q)$ and the empirical volatility autocorrelation $\widehat{\rho}_1(t)$ over subperiods or 7500 data (roughly 30 years) of the DJIA time series, one finds a considerable amount of variability, cf. Figure 3.1(B) and Figure 3.2(D). This indicates that some uncertainty in the estimation of the parameters is unavoidable, at least for time series of the given length. We discuss this issue in subsection 3.5 below, where we show that *a comparable variability* is observed in data series simulated from our model.

Remark 3.4. Baldovin and Stella [9, 26] find $D \simeq 0.24$ for their (different) model.

Remark 3.5. The multiscaling of empirical moments has been observed in several financial indexes in [14], where it is claimed that data provide solid arguments *against* model with linear or piecewise linear scaling exponents. Note that the theoretical scaling exponent $A(q)$ of our model is indeed piecewise linear, cf. (2.2.14). However, Figure 3.6(A) shows that the empirical scaling exponent $\widehat{A}(q)$ evaluated on data simulated from our model "smooths out"

the change of slope, yielding graphs that are analogous to those obtained for the DJIA time series, cf. Figure 3.1(B). This shows that the objection against models with piecewise linear $A(q)$, raised in [14], cannot apply to the model we have proposed.

3.3 The Distribution of Log-Returns

To completely specify the model, all it remains to do is to choose the distribution of σ according to the estimated mean and variance. For the case $C = \infty$, since the variance of σ is small compared to its mean, we made the choice of a constant σ; for the other two cases, we chose a gamma distribution with the right mean and variance. Anyway, the choice of the distribution of σ seems not to be relevant: in fact, we obtained nearly the same fitting with every other distribution we tried, according to Remark 3.2.

We have then compared the theoretical distribution $p_t(\cdot) := P(X_t \in \cdot) = P(X_t - X_0 \in \cdot)$ of our model for $t = 1$ (daily log-return) with the analogous quantity evaluated on the DJIA time series, i.e., the empirical distribution $\widehat{p}_t(\cdot)$ of the sequence $(x_{i+t} - x_i)_{0 \le i \le N-t}$:

$$\widehat{p}_t(\cdot) := \frac{1}{N+1-t} \sum_{i=0}^{N-t} \delta_{x_{i+t}-x_i}(\cdot)\,.$$

Although we do not have an analytic expression for the theoretical distribution $p_t(\cdot)$, it can be evaluated numerically via Monte Carlo simulations. In Figure 3.3(A) we have plotted the bulk of the distributions $p_1(\cdot)$ and $\widehat{p}_1(\cdot)$ or, more precisely, the corresponding densities in the range $[-3\hat{s}, +3\hat{s}]$, where $\hat{s} \simeq 0.0095$ is the standard deviation of $\widehat{p}_1(\cdot)$ (i.e., the empirical standard deviation of the daily log returns evaluated on the DJIA time series). In Figure 3.3(B) we have plotted the integrated tail of $p_1(\cdot)$, that is the function $z \mapsto P(X_1 > z) = P(X_1 < -z)$ (note that $X_t \sim -X_t$ for our model) and the right and left integrated empirical tails $\widehat{R}(z)$ and $\widehat{L}(z)$ of $\widehat{p}_1(\cdot)$, defined for $z \ge 0$ by

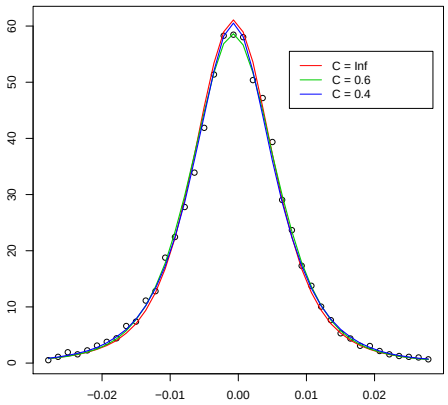
$$\widehat{L}(z) := \frac{\#\{1 \le i \le N :\ x_i - x_{i-1} < -z\}}{N}\,, \quad \widehat{R}(z) := \frac{\#\{1 \le i \le N :\ x_i - x_{i-1} > z\}}{N}\,,$$

in the range $z \in [\hat{s}, 6\hat{s}]$. As one can see, the agreement between the empirical and theoretical distributions is remarkably good for both figures, especially if one considers that *no parameter has been estimated using these curves.*

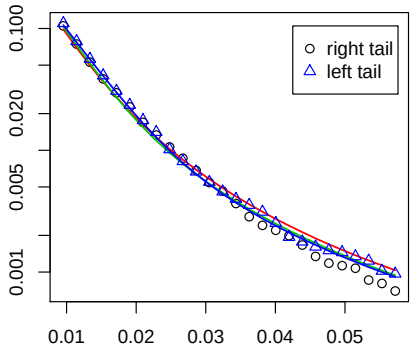
Figures 3.3(C) and 3.3(D) show that some amount of variability — less pronounced than for the multiscaling of moments and for the correlations — is present also in the empirical distribution $\widehat{p}_1(\cdot)$ evaluated over subperiods of 7500 data (roughly 30 years).

Remark 3.6. At this point, a fundamental difference between our model and the one proposed by Baldovin and Stella is fully revealed. Baldovin and Stella's *starting point* is the non-Gaussian density $g(\cdot)$ of the log-return distribution: as a first step, they estimate $g(\cdot)$ on the DJIA time series and build a basic process having $g(\cdot)$ as marginal density, cf. (1.3.1) and (1.3.2); afterwards, they build a new process via nonlinear time change, in order to break exchangeability and produce decay of correlations.

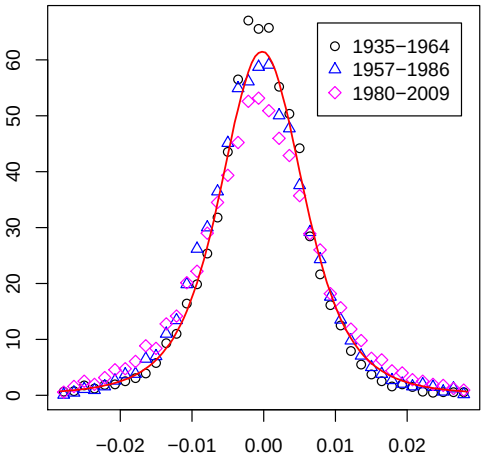
On the other hand, when σ is constant (as for the DJIA in the case $C = \infty$), the basic process on which our model is built is just Brownian motion with constant volatility



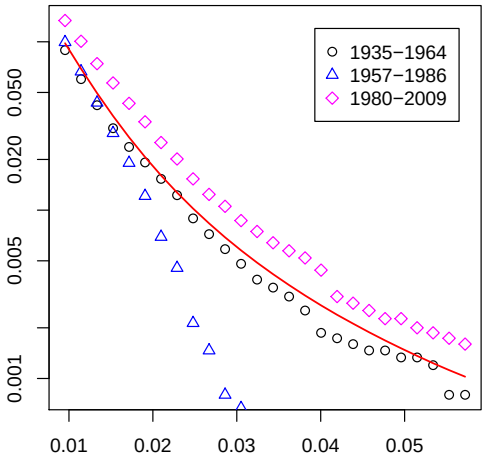
(A) *Density of the empirical distribution of the DJIA log-returns (1935-2009).*



(B) *Tails of the empirical distribution of the DJIA log-returns (1935-2009).*

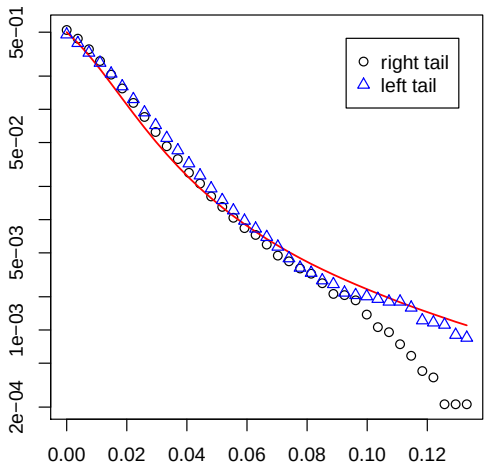


(C) *Density in subperiods of 30 years.*

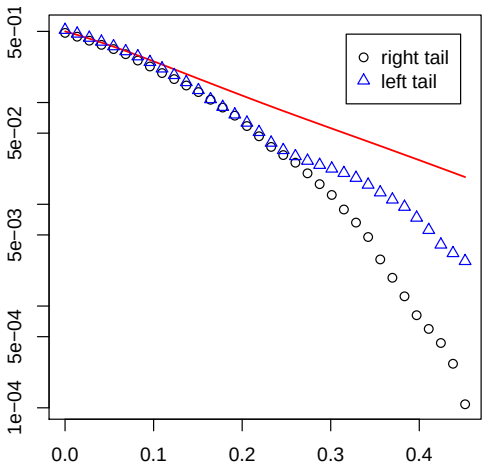


(D) *Left tail in subperiods of 30 years.*

FIGURE 3.3: *Distribution of 1-day log-returns in the DJIA time series (1935-2009).*
(A) The density of the log-return empirical distribution $\hat{p}_1(\cdot)$ in the period 1935-2009 (circles) and the theoretical prediction $p_1(\cdot)$ (line).
(B) The integrated right and left tails of the log-return empirical distribution $\hat{p}_1(\cdot)$ in the period 1935-2009 (see legend) and the theoretical prediction $p_1(\cdot)$ (line).
(C) The density of the log-return empirical distribution $\hat{p}_1(\cdot)$ evaluated in subperiods of 30 years (see legend) and the theoretical prediction $p_1(\cdot)$ (line).
(D) The integrated left tail of the log-return empirical distribution $\hat{p}_1(\cdot)$ evaluated in subperiods of 30 years (see legend) and the theoretical prediction $p_1(\cdot)$ (line).
Note. The theoretical distribution $p_1(\cdot)$ is evaluated for the parameters values given in (3.2.4),(3.2.5), (3.2.6) in (A) and (B), while is evaluated only for the parameters given in (3.2.4) for (C) and (D). The empirical standard deviation is $\hat{s} = \int x\, \hat{p}_1(\mathrm{d}x) \simeq 0.0095$; the range of plots (A) and (C) is $(-3\hat{s}, 3\hat{s})$ while the range of plots (B) and (D) is $(\hat{s}, 6\hat{s})$.



(A) *Tails of the empirical distribution of the DJIA log-returns over 5 days.*



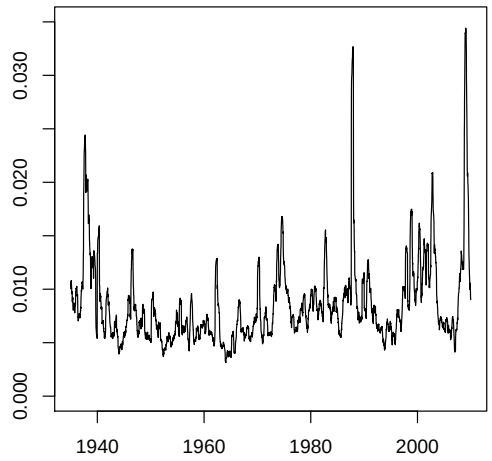
(B) *Tails of the empirical distribution of the DJIA log-returns over 400 days.*

FIGURE 3.4: *Tails of the distribution for different range log-returns.*
• Integrated right and left tails (see legend) of the log-return empirical distribution $\hat{p}_h(\cdot)$ for the DJIA time series in the period 1935-2009 and the theoretical prediction $p_h(\cdot)$ (line) for $h = 5$ days (A) and $h = 400$ days (B).
Note: The parameters of our model are fixed as in (3.2.4).

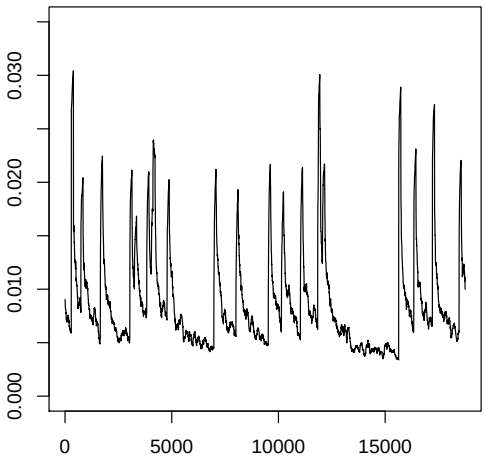
$(\sigma W_t)_{t\geq 0}$. Thus, the non-Gaussianity of the law of X_1 in our model is not imposed *a priori*, but is produced by the randomness of the time change, i.e. by the distribution of the shocks times $\mathcal{T}$. Having made the most natural choice for this distribution, namely that of a Poisson process, the fact that *a posteriori* we get an excellent fit with the empirical law of X_1 is quite remarkable.

Remark 3.7. While for short range log-returns (over 1-5-10-50 days) the tails of the empirical distribution are close to the ones predicted by our model, cf. Figure 3.5(A), for longer ranges (such as 400 ore more days) some discrepancies emerge, cf. Figure 3.5(B). In this range, the predicted super-exponential decay of the tails for the model with $\theta(t) = t^{2D}$, cf. Proposition 1.9, can actually be seen; anyway, this fact appears to have little interest in applications.

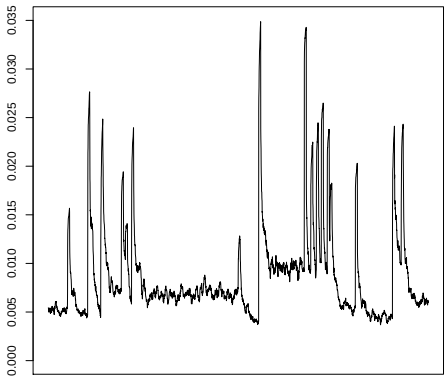
Remark 3.8. Our model predicts an even distribution for X_t. On the other hand, it is known that DJIA data exhibit a small but nonzero skewness. At the discrete-time level, and therefore at the level of simulations, skewness could be easily introduced in the framework of our model. In fact, recalling that *conditionally* on $\mathcal{T}$ the increments ΔX of our model are centered Gaussian random variables, it would suffice to replace the distribution of the increments which follow closely the events of $\mathcal{T}$ — those that are most likely to be large — by a skew "deformation" of a Gaussian.



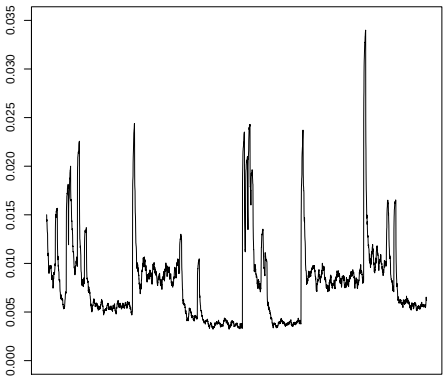
(A) *Empirical volatility of the DJIA (1935-2009).*



(B) *Empirical volatility of a simulation of our model with* $C = \infty$.



(C) *Empirical volatility of a simulation of our model with* $C = 0.6$.



(D) *Empirical volatility of a simulation of our model with* $C = 0.4$.

FIGURE 3.5: *Local volatility for the DJIA and for simulations of the model.*
(A) Empirical volatility for the daily log-returns of the DJIA time series in the period 1935-2009
(B) Empirical volatility for a simulation of our model with $C = \infty$ over 75 years.
(C) Empirical volatility for a simulation of our model with $C = 0.6$ over 75 years.
(D) Empirical volatility for a simulation of our model with $C = 0.4$ over 75 years.
Note: The volatility at day i is evaluated as $\left[\frac{1}{2R+1}\sum_{j=i-R}^{i+R}(x_j - x_{j-1})^2\right]^{1/2}$, with $R = 50$. The parameters of our model are fixed as in (3.2.4), (3.2.5), (3.2.6).

3.4 Volatility

For what concerns the features we have analyzed by now, the three versions of the model have almost the same fit with the empirical data, with maybe a very mild preference for the case $C = 0.6$, that effectively gives the minimal value for $\mathcal{L}_C$ among all the C values we tried. On the other hand, the version with $C = \infty$ is much simpler, both for its easier theoretical formulas and, more important, for its nearly constant σ that, as we will see in the next chapter, helps a lot in the option pricing setting. Anyway, the versions with $C \neq \infty$ are much more performant when it comes to consider the *volatility* of the process, particularly with $C = 0.6$. In fact, as we can see in Figure 3.5(B), the peaks of the empirical volatility evaluated on time series simulated according to our model with $C = \infty$ are all of about the same height and of sharply asymmetric shape, while the graphic of the empirical volatility for the DJIA time series shows clearly peaks of various heights and of quite symmetric shape, cf. Figure 3.5(A). While the asymmetric shape is a direct consequence of the definition of the process $(I_t)_{t\geq 0}$, cf. (1.4.4), the limited variability in the height of the peaks is mainly due to the fact that σ is constant.

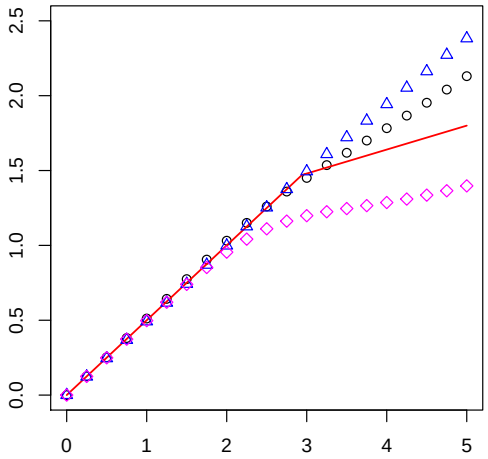
For the other versions of the model instead, a higher variability in the heights of the peaks of the volatility is obtained (cf. Figure 3.5(C) and Figure 3.5(D)), mainly thanks to the greater variance of σ. Moreover, the volatility decay due to the sublinearity of $t \mapsto t^{2D}$ stops after time $\frac{C}{\lambda}$, and for $\frac{C}{\lambda}$ sufficiently small this greatly reduces the asymmetries in the peaks profiles; in particular, for $C = 0.6$ we obtain a volatility that is quite comparable to the empirical one and in general this version has the best fit with respect to all the features considered.

A loss function that would take into account the volatility too, should certainly better distinguish between version of this model with different C, while without considering this feature the loss function assumes similar minimal values for very different fixed C; a loss function of this sort seems to lead, unfortunately, to a substancial increase of computational complexity.

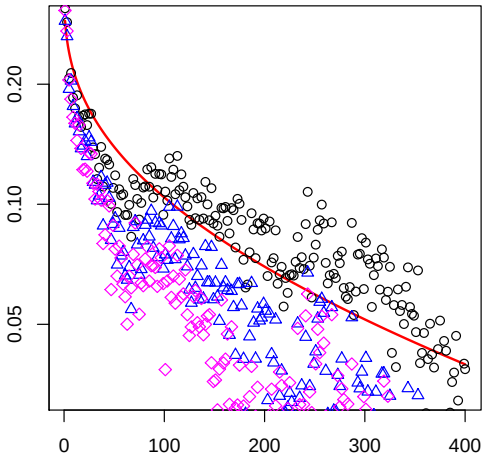
Remark 3.9. The current literature contains models whose simulated empirical volatility shows a better fit with real data; however, the level of accuracy reached, for example, in [12], has required many parameters and various sources of noise. One of our aim has been to keep low the number of parameters, and high the analytic tractability.

3.5 Variability of Estimators

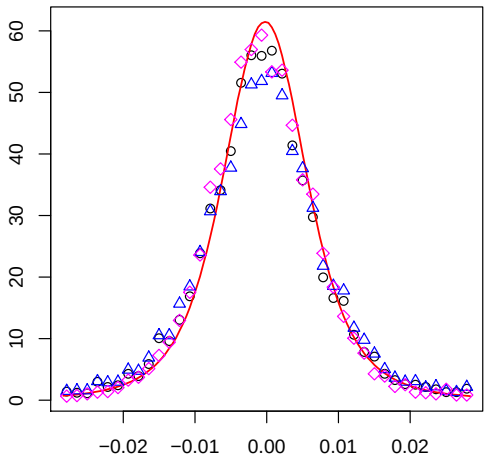
Generally speaking, it would be desirable to have "good" estimators, whose value does not change much if we compute them in sufficiently long but different time windows of the time series. On the contrary, we have already noticed that *considerable fluctuations* in different time windows of the DJIA time series are observed for the estimators we have used, namely the empirical scaling exponent $\widehat{A}(q)$, cf. Figure 3.1(A), the empirical volatility autocorrelation $\widehat{\rho}_1(t)$, cf. Figure 3.2(D), and — to a less extent — the empirical distribution $\widehat{p}_1(\cdot)$ of the 1-day log-return, cf. Figure 3.3(C) and 3.3(D). This is linked to the fact that relation (1.7.8) is an equality only for $N \to \infty$, therefore there is a priori no guarantee that the DJIA time series over a few decades is *close to the ergodic limit*, i.e., long enough to



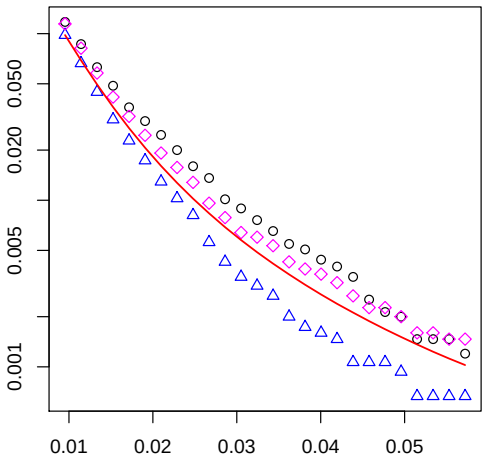
(A) *Multiscaling of moments.*



(B) *Volatility autocorrelation.*

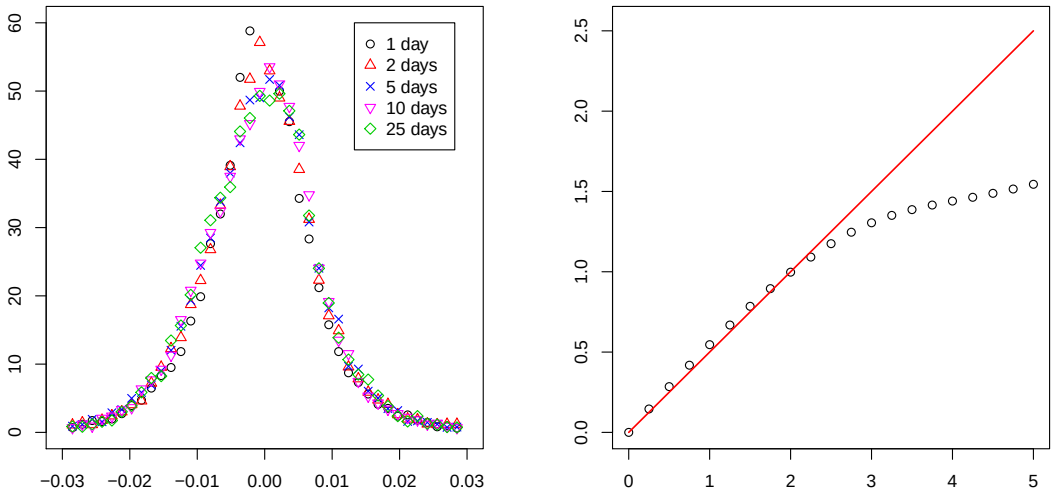


(C) *Density of the log-return empirical distribution.*



(D) *Tails of the log-return empirical distribution.*

FIGURE 3.6: *Simulations over 75 years: variability in subperiods of 30 years.*
Empirical evaluation of some observables in subperiods or 30 years for a 75-years-long time series sampled from our model $(X_t)_{t\geq 0}$ (with parameters fixed as in (3.2.4)): the scaling exponent $\widehat{A}(q)$ (A), the volatility autocorrelation $\widehat{\rho}_1(t)$ (B), the density (C) and the integrated tails (D) of the daily log-return distribution $\widehat{p}_1(\cdot)$.



(A) *Rescaled empirical density of log-returns.* (B) *Multiscaling of the moments.*

FIGURE 3.7: *Less-standard stylized facts for the S&P500 time series (1950-2010).*
(A) Plot of the rescaled empirical density of the h- log returns of the S&P500 time series (see legend). Standard deviation $\simeq 0.01$.
(B) The empirical scaling exponent $\widehat{A}(q)$ in the period 1950-2010 (circles) as a function of q, and the straight line $\frac{q}{2}$ (red line).

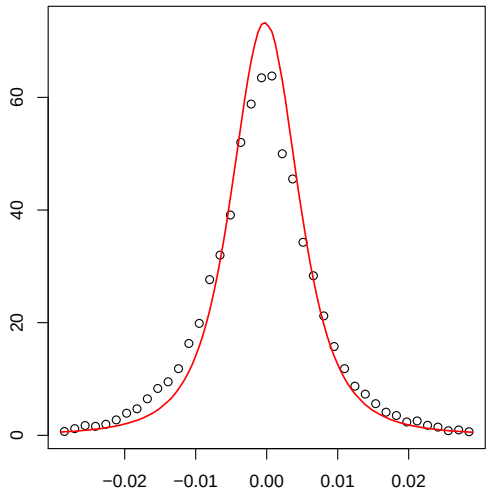
guarantee a reasonably approximate equality.

It is relevant to show that also this aspect is consistent with our model. In fact, it could happen that, unlike the DJIA, samples from our model over a few decades are long enough to damp out fluctuations. We have therefore simulated time series distributed according to our model, of the same length as the DJIA time series we are considering (18849 daily data), and we have evaluated the relevant quantities $\widehat{A}(q)$, $\widehat{\rho}_1(t)$ and $\widehat{p}_1(\cdot)$ over subperiods of 7500 days. Figure 3.6 shows that the variability displayed in such a sample is *quantitatively consistent* with what is observed on the DJIA time series. We believe that this is a crucial point in showing the agreement of our model with the DJIA time series.

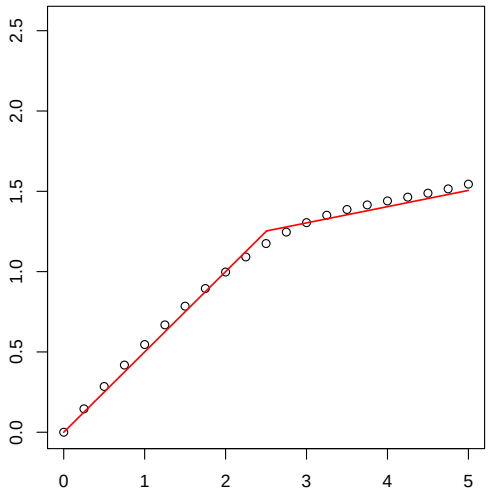
Remark 3.10. We point out that, among the different quantities that we have considered, the scaling exponent $\widehat{A}(q)$ appears to be the most sensitive. For instance, if instead of the opening prices one took the closing prices of the DJIA time series (over the same time period 1935-2009), one would obtain a different (though qualitatively similar) graph of $\widehat{A}(q)$.

3.6 S&P

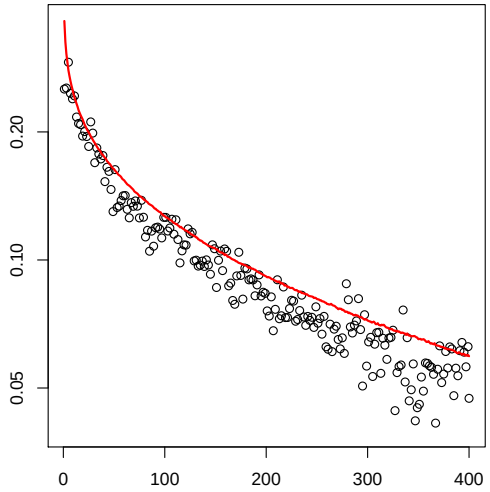
In this section we present an overview of the data fitting of our model with the S&P500 index; in particular, we considered the S&P500 Open Data time series, from 3 Jan 1950 to 7 July 2010. Our analysis on this index has been less precise than the one we did on the DJIA, we just contented ourselves with a rough estimation of the parameters for the simplest version of our model (i.e. $\theta(t) = t^{2D}$). Nonetheless, even with these limitations, we



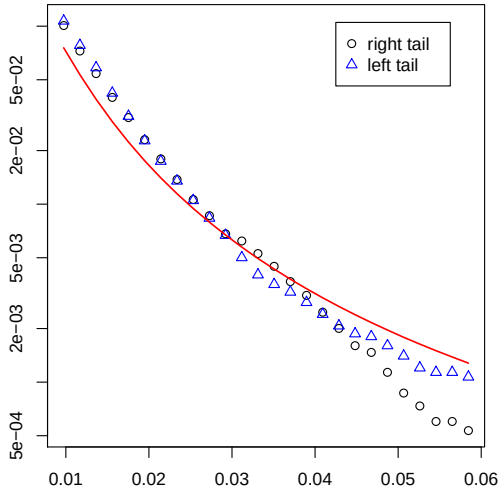
(A) *Density of the empirical distribution of the S&P500 log-returns (1950-2010).*



(B) *Multiscaling in the S&P500 (1950-2010).*

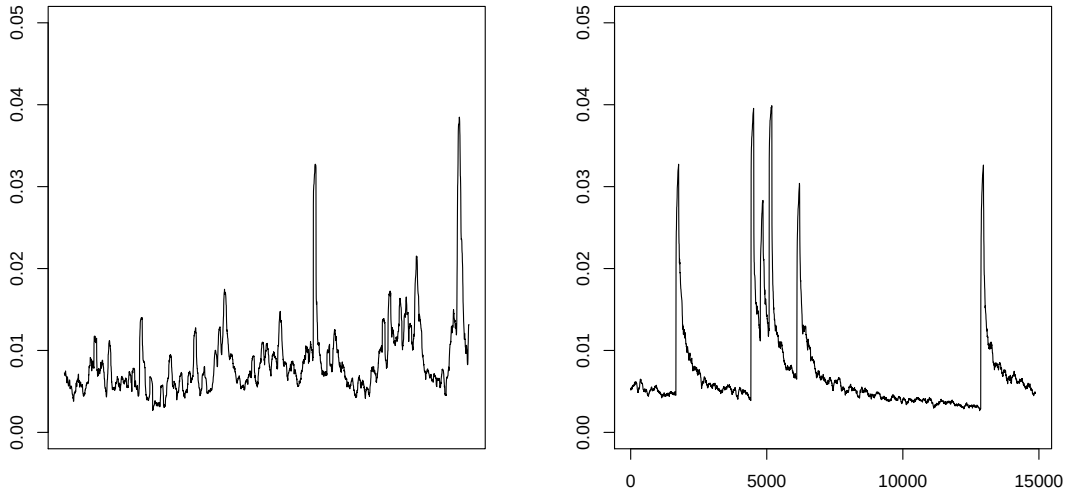


(C) *Volatility autocorrelation in the S&P500 (1950-2010): log plot.*



(D) *Tails of the empirical distribution of the S&P log-returns (1950-2010).*

FIGURE 3.8: *Fitting model for the S&P500 time series (1950-2010).*
(A) The density of the log-return empirical distribution $\hat{p}_1(\cdot)$ in the period 1950-2010 (circles) and the theoretical prediction $p_1(\cdot)$ (line).
(B) The empirical scaling exponent $\widehat{A}(q)$ in the period 1950-2010 (circles) and the theoretical scaling exponent $A(q)$ (line) as a function of q.// (C) Log plot for the empirical 1-day volatility autocorrelation $\widehat{\rho}_1(t)$ in the period 1950-2010 (circles) and the theoretical prediction $\rho(t)$ (line), as functions of t (days).
(D)The integrated right and left tails of the log-return empirical distribution $\hat{p}_1(\cdot)$ in the period 1950-2010 (see legend) and the theoretical prediction $p_1(\cdot)$ (line).
Note. We considered only the simplest version of the model (that is, $\theta(t) = t^{2D}$) for the S&P500 time series.



(A) *Empirical volatility of the S&P (1950-2010).*

(B) *Empirical volatility of a simulation of our model with $C = \infty$.*

FIGURE 3.9: *Local volatility for the S&P and for simulations of the model.*
(A) Empirical volatility for the daily log-returns of the S&P500 time series in the period 1950-2010
(B) Empirical volatility for a simulation of our model with $C = \infty$ over 61 years.
Note: The volatility at day i is evaluated as $\left[\frac{1}{2R+1}\sum_{j=i-R}^{i+R}(x_j - x_{j-1})^2\right]^{1/2}$, with $R = 50$. The parameters of our model are fixed as in (3.6.1).

found a good agreement between the model and the index.

First of all, in Figure 3.7 we can see that the S&P500 index presents the same scaling properties of the DJIA (and, of course, it presents the same stylized facts we mentioned in Section 1.1), i.e. the diffusive scaling of log-returns (see Figure 3.7(A)) and the multiscaling of the moments (see Figure 3.7(B)). Thus, this index is a good candidate for being compatible with our model. After having verified these preliminary conditions, for the estimation of parameters we used the same method we have illustrated in Section 3.2, but only for the version of the model with $\theta(t) = t^{2D}$ (i.e. $C = \infty$), obtaining the following estimates for the parameters:

$$\widehat{D} \simeq 0.10, \quad \widehat{\lambda} \simeq 0.00043, \quad \widehat{\mathbb{E}[\sigma]} \simeq 0.238, \quad \widehat{\mathbb{E}[\sigma^2]} \simeq 0.0567, \quad \frac{\sqrt{\widehat{Var(\sigma)}}}{\widehat{\mathbb{E}[\sigma]}} \simeq 0.015. \quad (3.6.1)$$

Remark 3.11. Note that, as in the DJIA case, σ can be considered almost constant.

In Figure 3.8 we can see the agreement between the empirical and theoretical quantities that we have considered before for the DJIA: the fit is again very good, even if not as impressive as in the Dow Jones case, especially for the bulk of the distribution (Figure 3.8(A)).

For what concerns the volatility (Figure 3.9), we can note that the same problems we found in the DJIA analysis for the model with $C = \infty$, but this was *a priori* totally obvious since we considered only this version for the S&P index.

Chapter 4

Option Pricing

In this chapter we present some results on option pricing with our model. Although no closed formulas is available by now, relevant quantities can be quite efficiently computed via Monte Carlo simulations.

4.1 The Market

Since the model is an univariate model, we study a market $\mathcal{M}$ composed by only two assets:

(1) the risk-free asset $S^{[0]} \equiv 1$;

(2) the riskly asset $S^{[1]}$. The prices of the riskly asset, that will be denoted by S_t, follows the dynamics

$$\mathrm{d}(\log S_t) = v_t \mathrm{d}B_t + r_t \mathrm{d}t, \qquad \text{i.e.} \qquad S_t = s_0 \exp\left(\int_0^t v_s \mathrm{d}B_s + \int_0^t r_s \, \mathrm{d}s \right), \quad (4.1.1)$$

where r_s is the local trend,

$$B_t = \int_0^{I_t} \frac{1}{I'(I^{-1}(s))} \mathrm{d}W_s, \qquad \text{and} \qquad v_t = \sqrt{I'(t)}, \quad (4.1.2)$$

using the same notations of Property (B), cf. 1.5.

Let us denote by $\mathcal{G}$ the σ-field generated by the two sequences of random variables $(\sigma_k)_{k\in\mathbb{N}_0}$ and $(\tau_k)_{k\in\mathbb{N}_0}$, which coincides with the σ-field generated by the process $I = (I_t)_{t\in[0,\infty)}$. As we have seen in proof of Property (B), the process B defined in (4.1.2) is a Brownian Motion independent of the σ-field $\mathcal{G}$. In particular, the process v defined in (4.1.2) is independent of B.

Remark 4.1. Under the conditional law $P^* := P(\cdot|\mathcal{G})$, equation (4.1.1) becomes particularly simple, because $v = (v_t)_{t\in[0,\infty)}$ is *deterministic*; this will be particularly important for pratically computing the prices, cf. section 4.3.

Following the traditional pricing approach presented in the section "Arbitrage Theory and Option Pricing", cf. the chapter Introduction, the first step is to find an equivalent martingale measure for our market $\mathcal{M}$; in the next section we present a simple choice for it.

4.2 A "Natural" Equivalent Martingale Measure for the Market

In this section we introduce a "natural" equivalent martingale measure for the market $\mathcal{M}$: this measure is "natural" in the sense that, under that, the (discounted) price process is a very sligth modification of our model, and requires to estimate only one new parameter.

Fix $T > 0$: define the law $\widetilde{P}_T$ by the Girsanov formula

$$\frac{\mathrm{d}\widetilde{P}_T}{\mathrm{d}P} = \exp\left(-\int_0^T \frac{v_t}{2} + \frac{r_t}{v_t}\mathrm{d}B_t - \frac{1}{2}\int_0^T \left(\frac{v_t}{2} + \frac{r_t}{v_t}\right)^2 \mathrm{d}t\right), \tag{4.2.1}$$

and analogously define $\frac{\mathrm{d}\widetilde{P}^*_T}{\mathrm{d}P^*}$.

Proposition 4.2. *The process $\widetilde{B} = (\widetilde{B}_t)_{t\in[0,T]}$, defined by*

$$\widetilde{B}_t = B_t + \int_0^t \left(\frac{v_s}{2} + \frac{r_s}{v_s}\right)\mathrm{d}s,$$

is a Brownian Motion under $\widetilde{P}^*_T$ and under $\widetilde{P}_T$.

Proof. The fact that $\widetilde{B}$ is a Brownian Motion under $\widetilde{P}^*_T$ is, for example, a simple application of Girsanov Theorem that uses the Novikov criterion (see [19] pag. 191); then, it is *a fortiori* a Brownian Motion under $\widetilde{P}_T$ too. □

Recalling (4.1.1) and 4.1.2, we can write

$$S_t = s_0 \exp\left(\int_0^t v_t \mathrm{d}\widetilde{B}_s - \frac{1}{2}\int_0^t v_s^2 \mathrm{d}s\right) \;=\; s_0 \exp\left(\int_0^t v_t \mathrm{d}\widetilde{B}_s - \frac{1}{2}I(t)\right), \tag{4.2.2}$$

where the first Itô integral is to be understood under $\widetilde{P}_T$ (or under $\widetilde{P}^*_T$).

By the Itô formula, it then follows immediately that $(S_t)_{t\in[0,T]}$ is a *martingale* under $\widetilde{P}_T$ (and under $\widetilde{P}^*_T$). Therefore, $\widetilde{P}_T$ is a martingale measure for our (discounted) process, and it is a quite convenient choice because under this measure the log-prices $\log S_t$ follow a dynamic that is a simple modification of our model, as we will see in a moment.

Set

$$\widehat{W}_s := \int_0^{I^{-1}(s)} v_t \,\mathrm{d}\widetilde{B}_t = \int_0^{I^{-1}(s)} \sqrt{I'(t)}\,\mathrm{d}\widetilde{B}_t,$$

it is easily checked that, under $\widetilde{P}_T$ and conditionally on $\mathcal{G}$ (that is, under $\widetilde{P}^*_T$), $\widehat{W} = (\widehat{W}_s)_{s\in[0,T]}$ is a Brownian motion (it is a Gaussian process with the right covariance). It follows that $\widehat{W}$ is, under $\widetilde{P}_T$, a Brownian motion independent of $\mathcal{G}$, and hence independent of I. Note that by construction, we can write

$$S_s = s_0 \exp\left(\widehat{W}_{I(s)} - \frac{1}{2}I(s)\right), \tag{4.2.3}$$

and this show that, under the martingale measure $\widetilde{P}_T$,

$$\log S_s = \log s_0 + \widehat{W}_{I(s)} - \frac{1}{2}I(s) \stackrel{d}{=} \log s_0 + X_s - \frac{1}{2}I(s), \tag{4.2.4}$$

where $X_s = W_{I(s)}$ follows our model dynamics; moreover, its parameters are the ones estimated under the "real world" probability. In fact, under $\widetilde{P}_T$ the process $\log S$ is equal in distribution to the process X_s under P_T except that for the "drift term" $\frac{1}{2}I(s)$; it has, in particular, the same driving Poisson point process of X_s, only the driving Brownian Motion changes in such a way to get rid of the drift. Then, we don't need to re-estimate the parameters of the process under $\widetilde{P}$, we can use the ones estimated on the hystorical prices time series.

Remark 4.3. It can be directly checked that the process in (4.2.3) is a martingale, just from the fact that $\widehat{W}$ is a Brownian motion independent of the process I.

Remark 4.4. Clearly the market $\mathcal{M}$ is incomplete, since we have three sources of randomness and only one riskly asset; in particular, there exist many equivalent martingale measures. The one we chosed, anyway, besides allowing for an easy pricing theory, is quite reasonable; in fact, the Poisson point process model the flow of "shocks" in the market, and we can think that they remain the same both in the"real" and in the "risk-neutral" world. Moreover, the distribution of Σ too remains the same under P and $\widetilde{P}$.

4.3 Option Pricing

In Section 4.2 we have introduced a simple equivalent martingale measure for the market $\mathcal{M}$, and now we use it to price an option according to the Theorem 0.36.

In order to do this, we need to know how this process behaves under the measure $\widetilde{P}_T$, that is the same that to know how W, Σ and $\mathcal{T}$ transform under this measure. First of all, $\mathcal{T}$ and Σ remain the same under P and $\widetilde{P}_T$, as we have seen ahead, while $\widetilde{W}$ is still a Brownian Motion independent of I, that just gets rid of the trend; then we are done, because all the parameters remain the same both under P and $\widetilde{P}_T$ and we can use the estimation obtained by the hystorical prices time series. Using the parameters estimated in Chapter 3 and (4.2.4), we theoretically have all the ingredients to compute the price of any option. In fact, given a T-claim X, its price at time t is

$$\Pi_X(t) = \mathbb{E}^{\widetilde{P}_T}\left[X|\mathcal{F}_t\right]; \tag{4.3.1}$$

in particular, if the option X is a simple claim (that is, if it depends only on the value of S at time T), observing that the triple $(S, \Sigma, \mathcal{T})$ is Markov, we obtain

$$\Pi_X(t) = \mathbb{E}^{\widetilde{P}_T}\left[X|(S_t, \sigma_{i(t)}, \tau_{i(t)})\right]. \tag{4.3.2}$$

Since a simple claim can be written as $X = f(S_T))$, recalling (4.2.3) we can rewrite (4.3.2) more explicitly as:

$$\Pi_X(t) = \mathbb{E}\left[f\left(S_t \exp\left(\widetilde{W}_{I(T-t)} - \frac{1}{2}I(T-t)\right)\right)\Big|S_t = \hat{S}_t, \sigma_{i(t)} = \hat{\sigma}_{i(t)}, \tau_{i(t)} = \hat{\tau}_{i(t)}\right], \tag{4.3.3}$$

where $\widetilde{W}$ is a Standard Brownian Motion and I is the subordinator process with the parameters estimated on the time series.

The expectation in (4.3.3), that doesn't admit a close formula even for simple claims such as the Call or the Put options, can be anyway easily evaluated via Montecarlo Simulations of its trajectories, at least if we know the two values $\hat{S}_t, \hat{\tau}_{i(t)}$. Since the prices $\hat{S}_t$ are clearly available, all we need is an estimation of $\tau_{i(t)}$ and $\sigma_{i(t)}$. Since $\mathcal{T}$ and Σ have the same distribution under P and $\widetilde{P}$, we can estimate this data both from the time series of the prices of the index, both from the time series of an option on this index. For simplicity reasons, we consider the simplest version of our model, that is $\theta(t) = t^{2D}$, and we set $\hat{\sigma}_{i(t)} = \hat{\sigma}$, where $\hat{\sigma}$ is the one estimated in Chapter 3. Then, in the following section we illustrate two possible methods for the estimation of $\hat{\tau}_{i(t)}$, the first using the time series of the index and the second using the time series of the Call options on the index.

4.4 Estimation of $\tau_{\mathbf{i(t)}}$

In this section we analyze two methods for the estimation of $\hat{\tau}_{i(t)}$, with a pratical application to the DJIA and S&P500 time series. It must be stressed that this method is based in heuristics and empirical evidences; no formal convergence proof is given.

4.4.1 Estimation on the index data

This method completely relies on just the hystorical prices time series of the index for the estimation of $\tau_{i(t)}$, and then require only data that are often really easy to obtain. As first thing, we will explain a preliminary version of the method that contains the foundamental idea, then we will refine this method to obtain the one we really used for the applications.

Consider a time series $\widehat{S} = \left(\widehat{S}_{t_j}\right)_{j\in\{1,\dots,N\}}$ of equispaced points taken by a trajectory $S_t(\omega)$ of our model, then fix a time $t_{\bar{j}}$; we want to estimate $\tau_{i(t_{\bar{j}})}(\omega)$, that is the last point of $\mathcal{T}(\omega)$ before $t_{\bar{j}}$. Consider the function

$$g_{t_{\bar{j}}}(n) = \frac{1}{n}\sum_{k=1}^{n}\left(\widehat{S}_{t_{\bar{j}-k}} - \widehat{S}_{t_{\bar{j}-k-1}}\right)^2, \tag{4.4.1}$$

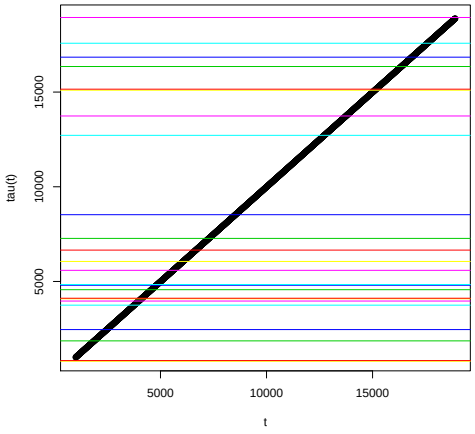
and define $\hat{n}$ as the first point where $g_{t_{\bar{j}}}$ assume a relative maximum; we set $\hat{\tau}_{i(t_{\bar{j}})}(\omega) = t_{\bar{j}-\hat{n}}$.

Why such a choice for $\tau_{i(t_{\bar{j}})}(\omega)$; the reason is the following. Assume for simplicity that $\tau_{i(t_{\bar{j}})}(\omega) \in t_{\bar{j}-1}, \dots, t_1$, then each term $\left(\widehat{S}_{t_{\bar{k}-1}} - \widehat{S}_{t_{\bar{j}-k-1}}\right)^2$ is, conditionally to $\mathcal{G}$, just a χ^2 variable multiplied for a coefficient A_k, with

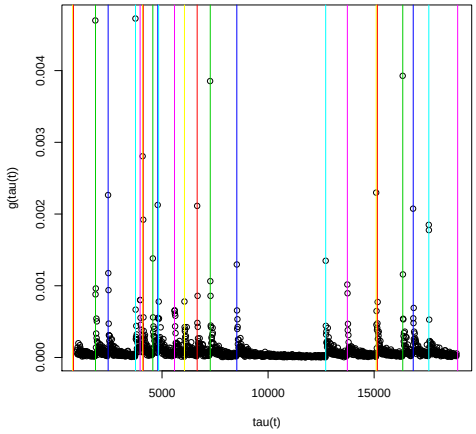
$$A_k = \sigma^2\left[\theta\left(t_{\bar{j}-k} - \tau_{i(t_{\bar{j}-k})}\right) - \theta\left(t_{\bar{j}-1-k} - \tau_{i(t_{\bar{j}-1-k})}\right)\right].$$

Note that, until $t_{\bar{j}-1-k} \geq \tau_{i(\tau_{\bar{j}})}$, the A_k is increasing; in fact, we can write

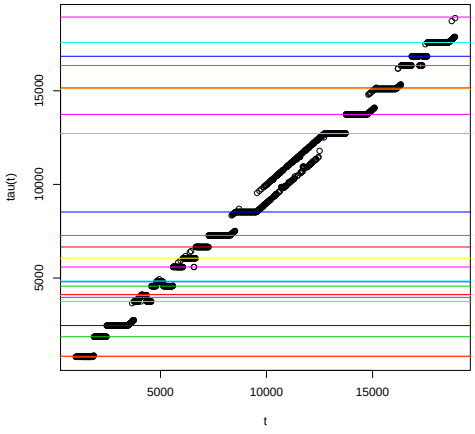
$$A_k = \sigma^2\int_{t_{\bar{j}-k-1}}^{t_{\bar{j}-k}} \theta'(s)\mathrm{d}s$$



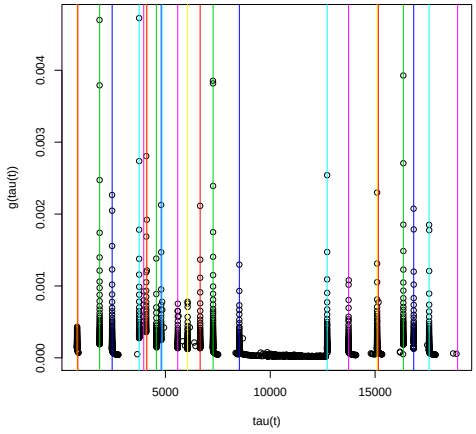
(A) $\hat{\tau}_{i(t)}$ *with the non refined method.*



(B) $\hat{\tau}_{i(t)}, g(\hat{\tau}_{i(t)})$ *with the non refined method.*



(C) $\hat{\tau}_{i(t)}$ *with the non refined method.*



(D) $\hat{\tau}_{i(t)}, g(\hat{\tau}_{i(t)})$ *with the non refined method.*

FIGURE 4.1: *Estimation of* $\tau_{(i(t))}$ *on simulated data.*
(A): Estimation with the non-refined method of $\tau_{i(t_j)}$ (y-axes) at each time t_j (x-axes) for a simulated time series; the line are in corrispondence of the points $\tau_k \in \mathcal{T}$ of the simulated time series.
(B): Estimation with the non-refined method of $\tau_{i(t_j)}$ (x-axes) at each time t_j, with the relative value $g(\hat{\tau}_{i(t_j)})$ (y-axes) for a simulated time series; the line are in corrispondence of the points $\tau_k \in \mathcal{T}$ of the simulated time series.
(C): Estimation with the refined method of $\tau_{i(t_j)}$ (y-axes) at each time t_j (x-axes) for a simulated time series; the line are in corrispondence of the points $\tau_k \in \mathcal{T}$ of the simulated time series.
(D): Estimation with the refined method of $\tau_{i(t_j)}$ (x-axes) at each time t_j, with the relative value $g(\hat{\tau}_{i(t_j)})$ (y-axes) for a simulated time series; the line are in corrispondence of the points $\tau_k \in \mathcal{T}$ of the simulated time series.
The simulated time series used is the same for all the four graphics, and is just a realization of our model with $\theta = t^{2D}$ and parameters fixed as in (3.2.4).

and, since θ' is decreasing and the t_i are equispaced, it follows $A_k \geq A_{k-1}$.

Proposition 4.5. *With the notations above, the function*

$$\widetilde{g}(n) := \frac{1}{n}\sum_{i=1}^{n} A_i \tag{4.4.2}$$

is increasing, until n is such that $t_{\bar{j}-1-n} \geq \tau_{i(\tau_{\bar{j}})}$.

Proof. As we have noted above, until $t_{\bar{j}-1-k} \geq \tau_{i(\tau_{\bar{j}})}$ we have that $A_k \geq A_{k-1}$. Then, we proceed by induction:

for the case $n = 1$,

$$A_1 = \frac{1}{2}A_1 + \frac{1}{2}A_1 \leq \frac{1}{2}A_1 + \frac{1}{2}A_2.$$

For the inductive step, suppose that the property holds until n, then

$$\frac{1}{n+1}\sum_{i=1}^{n+1} A_i = \frac{n}{n+1}\cdot\frac{1}{n}\sum_{i=1}^{n} A_i + \frac{A_{n+1}}{n+1} \geq \frac{n}{n+1}\cdot\frac{1}{n-1}\sum_{i=1}^{n-1} A_i + \frac{A_{n+1}}{n+1} \geq$$

$$\frac{n}{n-1}\cdot\frac{1}{n+1}\sum_{i=1}^{n-1} A_i + \frac{n}{n-1}\cdot\frac{A_{n+1}}{n+1} = \frac{n}{n^2-1}\cdot\sum_{i=1}^{n} A_i,$$

that is greater or equal to $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} A_i$ because $\dfrac{n}{n^2-1} \geq \dfrac{1}{n}$. □

We want to investigate now what happens when n is such that the term $t_{\bar{j}-1-n}$ becomes smaller than $\tau_{i(\tau_{\bar{j}})}$; first of all note that, since θ has a singularity in 0, A_k increases faster and faster as $t_{\bar{j}-1-k}$ approaches $\tau_{i(t_{\bar{j}})}$, and the same does $\widetilde{g}(n)$ as $t_{\bar{j}-1-n}$ approaches $\tau_{i(t_{\bar{j}})}$. Instead, when $t_{\bar{j}-1-k}$ overcomes $\tau_{i(t_{\bar{j}})}$ (that is, when $t_{\bar{j}-k} = \tau_{i(t_{\bar{j}})}$), it is very likely that the term A_k becomes much smaller than the term A_{k-1}, because in this case

$$A_k = \theta\left(\lambda\left(t_{\bar{j}-k-1} - \tau_{i(t_{\bar{j}-k-1})}\right)\right),$$

with $E\left[t_{\bar{j}-k-1} - \tau_{i(t_{\bar{j}-k-1})}\right] \approx \frac{1}{\lambda}$. Then, if A_k is sufficiently small, which occurs in the (likely) case the two consecutive shocks are not too close, the inequality $\widetilde{g}(n) \leq \widetilde{g}(n-1)$ holds for n such that $t_{\bar{j}-n} = \tau_{i(t_{\bar{j}})}$; in other words, in this case it holds $t_{\bar{j}-n} = \tau_{i(t_{\bar{j}})}$, where n is the first point of (relative) maximum for $\widetilde{g}$.

Since $\dfrac{1}{n}g_{t_{\bar{j}}}(n) = \displaystyle\sum_{k=1}^{n} A_i\,\chi_i^2$, where the χ_i^2 are n independent χ^2 random variables (and then have a relatively small standard deviation), one could hope that $g(n)$ behaves like $\widetilde{g}(n)$, so it is reasonable to set $\tau_{i(t_{\bar{j}})}(\omega) = t_{\bar{j}-\hat{n}}$, with $\hat{n}$ the first point where $g_{t_{\bar{j}}}$ assume a relative maximum.

Unfortunately, when we tested the algorithm on simulated data, we found that the estimation for the points of $\mathcal{T}$ was wrong, as we can see in Figure 4.1(A). The problem is

that, at least when we consider $t_{\bar{j}-k}$ far from the points of $\mathcal{T}$, $\theta(t_{\bar{j}-k} - t_{\bar{j}-k-1})$ increases slowly (this is due to the fact that θ increases slowly far from 0), so the fluctuations given by the χ^2 term destroy the increasing property of $\widetilde{g}$.

Anyway, if we plot (as in Figure 4.1(B)) both $\hat{\tau}_{i(t_j)}$ and $g(\hat{\tau}_{i(t_j)})$ on the same plane as functions of t_j, we can note that, when t_j is near to $i(t_j)$, two things happen:

- the estimated $\hat{\tau}_{i(t_j)}$ effectively collapses to the same point $\tau_{i(t_j)}$;
- the values $g(\hat{\tau}_{i(t_j)})$ remarkably increases as t_j approach to $i(t_j)$.

These two features, that are connected to the fact that I_t has a singularity on each point of $\mathcal{T}$, suggest a more refined method for the estimation of $\tau_{i(t_j)}$. With the same notations as above, we set $\hat{\tau}_{i(\bar{t}_j)} = t_{\bar{j}-\hat{n}}$, but this time $\hat{n}$ is the point where $g_{t_{\bar{j}}}(n)\big|_{n\in\{1,\dots,N\}}$ assume the *absolute* maximum value. In words: instead of taking $\hat{n}$ as the first relative maximum point of $g_{t_{\bar{j}}}$, we set it as the point that maximize $g_{t_{\bar{j}}}$ over an appropriate subset $\{1,\dots,N\}$ of $\mathbb{N}$, that corresponds to the interval $[t_{\bar{j}-N}, t_{\bar{j}}]$ (see Remark 4.6). From what we have seen above, if we assumwe $[t_{\bar{j}-N}, t_{\bar{j}}] \cap \mathcal{T} = \left\{\tau_{i(t_{\bar{j}})}\right\}$, it is quite likely that the estimated point $\hat{\tau}_{i(\bar{t}_j)}$ coincides with $\tau_{i(\bar{t}_j)}$ (this is no more true if $|[t_{\bar{j}-N}, t_{\bar{j}}] \cap \mathcal{T}| \neq 1$, see Remark 4.7). In Figure 4.1(C) and Figure 4.1(D), we can see that effectively this method outperforms the previous one on the same simulated data, and this is the method we will consider by now.

If we consider just the estimation $\hat{\tau}_{i(t_{\bar{j}})}$ for the single time $t_{\bar{j}}$, again this is not very reliable; anyway, this completely changes if we apply this method to a *window of values* $\hat{\tau}_{i(t_{\bar{j}})}, \hat{\tau}_{i(t_{\bar{j}-1})}, \dots, \hat{\tau}_{i(t_{\bar{j}-M})}$. In fact, even if each single value $\hat{\tau}_{i(t_{\bar{j}-k})}$ have a considerable possibility of being wrong, most of them will be right and the estimated points will gather in corripondence of the points of $\mathcal{T}$; so, we will identify the points of $\mathcal{T}$ with clustering of points $\hat{\tau}_{i(t_{\bar{j}-k})}$, as in Figure 4.1(D). Note that all the points $\tau_k \in \tau$ are perfectly catched in this way.

Remark 4.6. This last method requires to determine for each $t_{\bar{j}}$ an appropriate interval $[t_{\bar{j}-N}, t_{\bar{j}}]$, that is equivalent to determine the subset $\{1,\dots,N\}$ on which we maximize the function $g_{t_{\bar{j}}}$. Since we would like to have $\left|[t_{\bar{j}-N}, t_{\bar{j}}] \cap \mathcal{T}\right| = 1$, we want to maximize $P(\left|[t_{\bar{j}-N}, t_{\bar{j}}] \cap \mathcal{T}\right| = 1)$, that has the solution $t_{\bar{j}} - t_{\bar{j}-N} = \frac{1}{\lambda}$.

Remark 4.7. What happens if despite our choice for the interval $[t_{\bar{j}-N}, t_{\bar{j}}]$, two (or more) points of $\mathcal{T}$ (name it τ_k and τ_{k+1}) fall on it?

There are two main possibilities, depending (partially) on the relation between σ_k and σ_{k-1}:

- if $\sigma_{k-1} >> \sigma_k$, the point τ_{k+1} will be "hidden" by the point τ_k; the points

  $$\hat{\tau}_{i(t_j)}, \hat{\tau}_{i(t_{j-1})}, \dots, \hat{\tau}_{i(t_{j-M})}$$

  will gather around a value between τ_k and τ_{k+1} (possibly around t_k itself) even if the estimation is based on points "after" τ_{k+1}. In other terms, if

  $$i(t_{\bar{j}}) = i(t_{\bar{j}-1}) = \dots = i(t_{\bar{j}-M}) = k+1,$$

  then the estimated points $\hat{\tau}_{i(t_{\bar{j}})}, \hat{\tau}_{i(t_{\bar{j}-1})}, \dots, \hat{\tau}_{i(t_{\bar{j}-M})}$ will gather around a time t such that $\tau_k \leq t < \tau_{k+1}$.

- if σ_{k-1} is bigger than σ_k, but comparable with it, then initially the point τ_{k+1} can be "hidden" by τ_k as in the previous case, but when we based our estimation on points "after" τ_{k+1} but not too far from it, τ_{k+1} will become visible; this is the situation in Figure 4.2(A).

After the tests on simulated data, we applied the method to empircal time series, namely to the DJIA and to the S&P 500 hystorical prices. As we can see in Figure 4.2, the clustering phenomenon for the estimated points $\hat{\tau}_{i(t_j)}$ is present in the empirical time series. The estimated $\hat{\tau}_{i(t_j)}$ are then consistent with the model, and, moreover, they are consistent with the estimated λ. In fact we have that, for the DJIA, the estimated λ is $\hat{\lambda} \approx \frac{1}{1000}$, and then the mean number of "jump" points in 75 years is about 19; with this method we have 17 estimated "jump" points, that is a perfectly consistent number.

Besides, the function $\widetilde{\tau}(t_k) : t_k \longrightarrow \hat{\tau}_{i(t_k)}$ is almost piece-wise constant, cf. Figure 4.2(D).

Remark 4.8. The concistency between the number of points estimated with this method and the mean number of points implied from the estimated λ parameter is more striking if one thinks that the two have been found using very different quantities: the function g for the direct estimation of the points τ_k, the volatility autocorrelation decay for the estimation of λ.

Finally, we want stress two major advantages of this method:

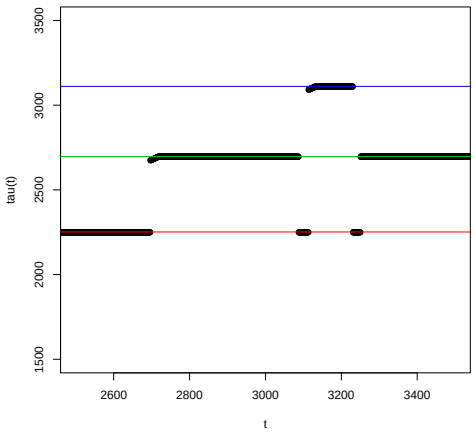
- the algorithm (quite easy to implement) is really fast: the estimation of $\tau_{t_{\bar{j}}}$ takes far less than a second;
- all we need for the estimation is the time series of hystorical prices of the index, no other data are required.

Remark 4.9. Beside the applications to the option pricing, this method has its own interest; in fact, it allows to determine the big economic crises in history directly from the hystorical prices of the index. For the Dow Jones, we found that the crises (that is, the estimated points of $\mathcal{T}$) are the following: 1 October 1939, 10 May 1940, 31 August 1946, 23 June 1950, 23 September 1955, 26 May 1962, 23 September 1970, 5 October 1974, 17 August 1982, 6 May 1982, 16 October 1987, 13 October 1989, 25 October 1997, 29 August 1998, 8 September 2001, 19 July 2002. Some data corresponds perfectly with important hystorical events (as 10 May 1940, that coincides with the falling of France to the german army, or as the 23 June 1950, the Korean War); others corresponds to periods of general depression. Finally, some points seems to be linked to recovery of the market, such as the ones on 1982.

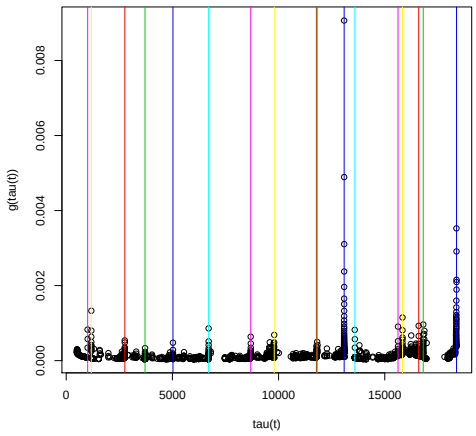
4.4.2 Estimation on the option data

The method we now describe is somehow analougous similar to what is done when one calculate the implied volatility of options prices.

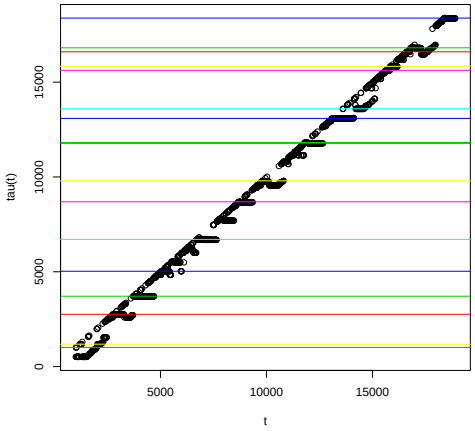
Consider a Call option X_C with fixed time maturity $T > t$, and fixed strike price K; since the function $\Pi_{X_C}(t)$ is strictly monotone with respect to $\tau_{i(t)}$ once that S_t is fixed (the price decreases as $\tau_{i(t)}$ becomes closer to t), we can define the inverse function $\widetilde{\tau} : \mathbb{R} \longrightarrow \mathbb{R}^+$



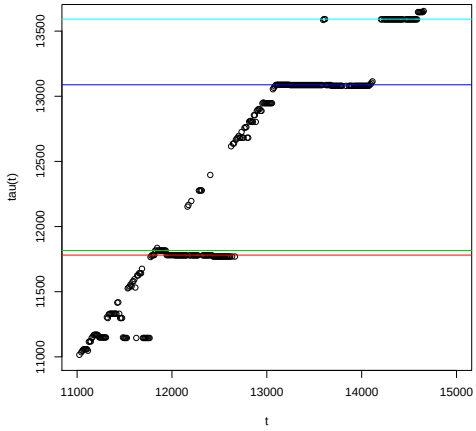
(A) $\hat{\tau}_{i(t)}$ *when* τ_k *is near to* τ_{k+1}.



(B) $\hat{\tau}_{i(t)}, g(\hat{\tau}_{i(t)})$ *for the DJIA.*



(C) $\hat{\tau}_{i(t)}$ *for the DJIA.*



(D) *a "zoom" on* $\hat{\tau}_{i(t)}$.

FIGURE 4.2: *Estimation of* $\tau_{(i(t))}$ *on simulated data.*
(A): Estimation (with the refined method) of $\tau_{i(t_j)}$ (y-axes) at each time t_j (x-axes) for a simulated time series when there are near points of $\mathcal{T}$ with different σ_i in increasing order. The τ_i we see are $\tau_3 = 2252, \tau_4 = 2697, \tau_5 = 3111$, while the σ_i are, respectively, $\sigma_3 = 0.137, \sigma_4 = 0.130, \sigma_5 = 0.093$. The line are in corrispondence of the points $\tau_j \in \mathcal{T}$ of the simulated time series.
(B): Estimation (with the refined method) of $\tau_{i(t_j)}$ (x-axes) at each time t_j, with the relative value $g(\hat{\tau}_{i(t_j)})$ (y-axes) for the DJIA time series; the line are in corrispondence of the points that we decided to consider really points of $\mathcal{T}$
(C): Estimation (with the refined method) of $\tau_{i(t_j)}$ (y-axes) at each time t_j (x-axes) for the DJIA time series; the line are in corrispondence of the points that we decided to consider really points of $\mathcal{T}$.
(D): A "zoom" of (C), considering just the years 1979-1995.

that associates to $\Pi_{X_C}(t)$ the unique $\widetilde{\tau}_{(i(t))}(\Pi_{X_C}(t))$ such that

$$\mathbb{E}\left[\left(S_t \exp\left(\widetilde{W}_{I(T-t)} - \frac{1}{2}I(T-t)\right) - K\right)^+ \Big| S_t = \hat{S}_t, \tau_{i(t)} = \widetilde{\tau}_{(i(t))}(\Pi_{X_C}(t))\right] = \Pi_{X_C}(t).$$

We name $\widetilde{\tau}_{(i(t))}(\Pi_{X_C}(t))$ as the *implied jump point.*

As for the implied volatility, varying T and K we obtain a surface for $\widetilde{\tau}$; given a data set of option prices, we set the estimated "jump time' '$\hat{\tau}_{i(t)}$ as the best fit in the minimal least-squares sense.

This method presents two main difficulties:

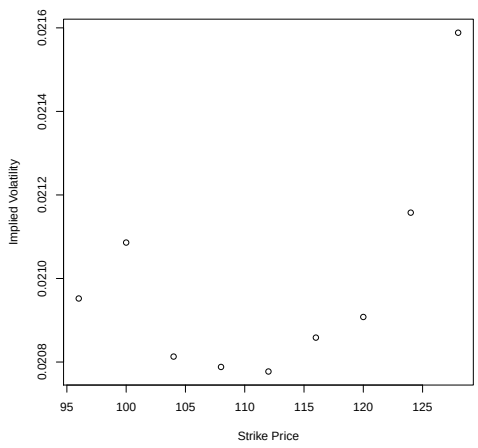
- the prices of options on an index are far less many and far less available than the hystorical prices of the index itself: in particular, we can't use this method to estimate the big economic crises as with the method above.
- since we don't have a close formula for $\Pi_{X_C}(t)$, the least squares minimization problem is solved completely by numerical means, making this method computatyonally quite hard (but we must remark that we implemented this method in a really straightforward way, so probably it can be greatly speeded up).

4.4.3 A comparison between the two methods

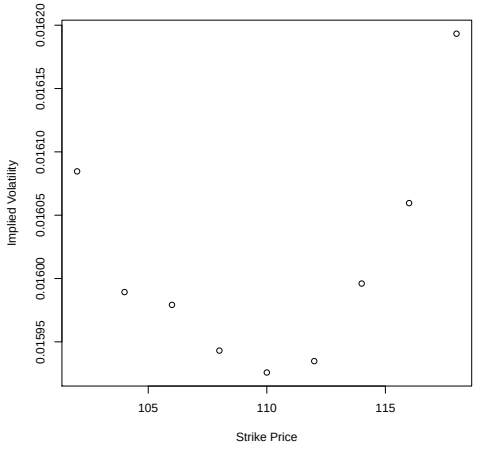
To compare the two methods for the estimation of $\hat{\tau}_{i(t_{\bar{j}})}$, we did as follows. We considered the DJIA index, daily prices, and

- we first estimated $\hat{\tau}_{i(t_{\bar{j}})}$ with the first method for $t_{\bar{j}}$ = 22 September 2010, 23 September 2010, ..., 22 October 2010 (using all the hystorical prices time series for the estimation),
- then we estimated $\hat{\tau}_{i(t_{\bar{j}})}$ with the second method using the prices of the Call options bought from the 22 September 2010 to the 22 October 2010 (again, daily prices), with expire data 20 November 2010 and strike prices $K = 105, 106, \ldots, 118$.

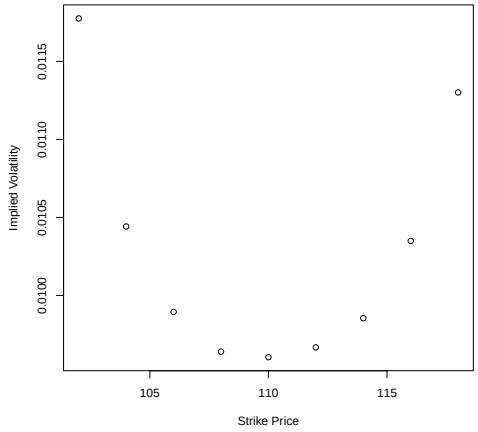
The resulting estimation for $\hat{\tau}_{i(t_{\bar{j}})}$ with the two methods is very different: in fact, $t_{\bar{j}} - \hat{\tau}_{i(t_{\bar{j}})}$ is about 500 days according to the first method, while it is about 260 days according to the second one. To find out what estimation was more useful for the option pricing, we then computed the prices for Call options given by our model, with $t_{\bar{j}} - \hat{\tau}_{i(t_{\bar{j}})} = 260$ and $t_{\bar{j}} - \hat{\tau}_{i(t_{\bar{j}})} = 500$ respectively, and we compared them with the market prices. We consider as test benchmark the Call options bought from 22 September 2010 to 22 October 2010 (daily prices), with strike prices $K = 105, 106, \ldots, 118$ and expire data respectively 20 November 2010 (first test benchmark, they are the same prices on wich we estimated $\hat{\tau}_{i(t_{\bar{j}})}$ with the second method) and 18 December 2010. In both cases, the prices obtained using as $\hat{\tau}_{i(t_{\bar{j}})}$ the one estimated with the second method result closer to the market prices than the ones obtained using the first method to estimate $\hat{\tau}_{i(t_{\bar{j}})}$. While this was *a priori* straightforward for what concern the first test benchmark, this was not obvious for the second benchmark. Anyway, more analysis should be done to obtain more reliable results.



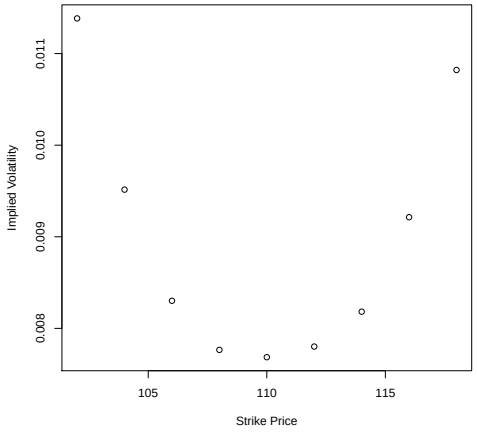
(A) *Implied volatility for* $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 20$.



(B) *Implied volatility for* $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 50$.



(C) *Implied volatility for* $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 250$.



(D) *Implied volatility for* $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 500$.

FIGURE 4.3: *Volatility smile for simulated prices.*
(A): Implied volatility for the prices given by our model to a Call option with time to maturity 10 days, for the strike prices $K = 96, 100, 104, \ldots, 128$. The value of the index at time $t_{\bar{j}}$ is $S_{t_{\bar{j}}} = 110$, while $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 20$.
(B): Implied volatility for the prices given by our model to a Call option with time to maturity 10 days, for the strike prices $K = 102, 104, 106, \ldots, 118$. The value of the index at time $t_{\bar{j}}$ is $S_{t_{\bar{j}}} = 110$, while $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 50$.
(C): Implied volatility for the prices given by our model to a Call option with time to maturity 10 days, for the strike prices $K = 102, 104, 106, \ldots, 118$. The value of the index at time $t_{\bar{j}}$ is $S_{t_{\bar{j}}} = 110$, while $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 250$.
(D): Implied volatility for the prices given by our model to a Call option with time to maturity 10 days, for the strike prices $K = 102, 104, 106, \ldots, 118$. The value of the index at time $t_{\bar{j}}$ is $S_{t_{\bar{j}}} = 110$, while $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 500$.
We considered the version of our model with $\theta = t^{2D}$ and parameters fixed as in (3.2.4).

4.5 Implied Volatility Surface

In this section we analyze the implied volatility surface for the prices given by our model, where $\theta(t) = t^{2D}$ and with parameters fixed as in (3.2.4); we considered the implied volatility surface fixing different values for $t_{\bar{j}} - \tau_{i(t_{\bar{j}})}$.

For all the choices of $t_{\bar{j}} - \tau_{i(t_{\bar{j}})}$ the prices predicted by the model imply the *volatility smile* feature, at least in some cases, and that this feature is more evident for bigger values of $t_{\bar{j}} - \tau_{i(t_{\bar{j}})}$, cf. Figure 4.3(A), Figure 4.3(B), Figure 4.3(C) and Figure 4.3(D).

For what concern the *volatility term structure* (we analyzed the "at the money" case), we can note the following behaviour:

- for $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 20$ and $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 50$ (and, more generally, for relatively small values of $t_{\bar{j}} - \tau_{i(t_{\bar{j}})}$), the implied volatility is increasing with the time to maturity, see Figure 4.4(A);
- for $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 200$ and for $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 250$ (and, more generally, for intermediate values of $t_{\bar{j}} - \tau_{i(t_{\bar{j}})}$), the implied volatility has a very ireggular behaviour, see Figure 4.4(B) and Figure 4.4(C);
- for $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 500$ and $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 1000$ (and, more generally, for relatively big values of $t_{\bar{j}} - \tau_{i(t_{\bar{j}})}$), the implied volatility is decreasing with the time to maturity, see Figure 4.4(D).

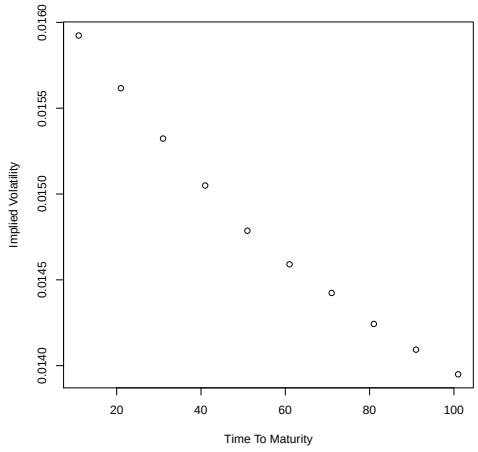
We think that this behaviour can be due to the following facts:

- when we are near to a point of $\mathcal{T}$, the volatility is inusually big because of the recent crisis, but it is very likely that it will fast decreases; this is the situation for $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 20$ and $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 50$.
- When we are far from a point of $\mathcal{T}$, the actual volatility is quite small, but the possibility that a crisis (that is, a point of $\mathcal{T}$) will take place between now and the date of maturity of the option, makes the implied volatility to grow bigger as the time to maturity increases; this is the situation for $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 500$ and $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 1000$.
- Finally, when we are nor far nor near to a crisis, the two effects described above mix togheter, with very irregular results; this is the situation for $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 200$ and $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 250$.

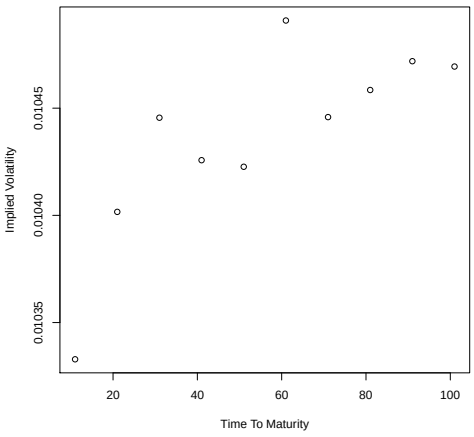
Anyway, more research has to be done about it.

Acknowledgements

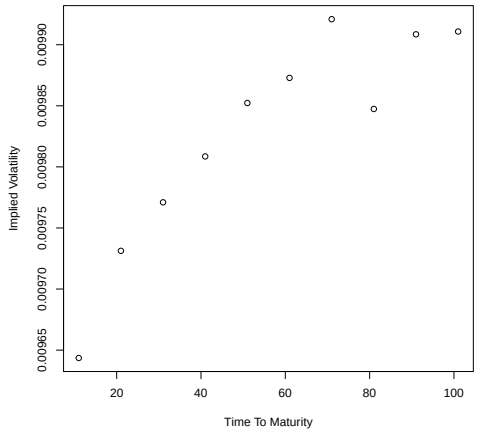
We thank Fulvio Baldovin, Massimiliano Caporin, Wolfgang Runggaldier and Attilio Stella for fruitful discussions.



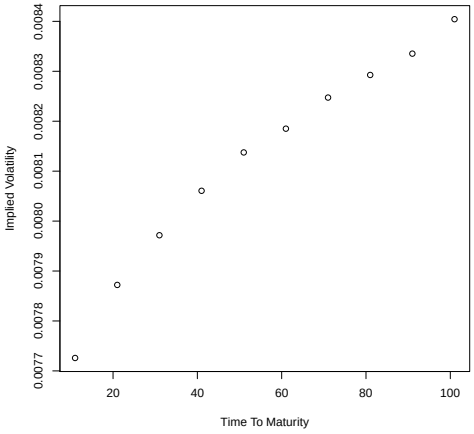
(A) *Volatility term structure,* $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 50$.



(B) *Volatility term structure,* $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 200$.



(C) *Volatility term structure,* $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 250$.



(D) *Volatility term structure,* $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 500$.

FIGURE 4.4: *Volatility term structure for simulated prices.*
(A): Implied volatility for the prices given by our model to a Call option "at the money" (that is, $S_{t_{\bar{j}}} = K = 110$) with time to maturity $1, 11, 21, \ldots, 101$ days. In this picture $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 50$.
(B): Implied volatility for the prices given by our model to a Call option "at the money" (that is, $S_{t_{\bar{j}}} = K = 110$) with time to maturity $1, 11, 21, \ldots, 101$ days. In this picture $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 200$.
(C): Implied volatility for the prices given by our model to a Call option "at the money" (that is, $S_{t_{\bar{j}}} = K = 110$) with time to maturity $1, 11, 21, \ldots, 101$ days. In this picture $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 250$.
(D): Implied volatility for the prices given by our model to a Call option "at the money" (that is, $S_{t_{\bar{j}}} = K = 110$) with time to maturity $1, 11, 21, \ldots, 101$ days. In this picture $t_{\bar{j}} - \tau_{i(t_{\bar{j}})} = 500$.
We considered the version of our model with $\theta = t^{2D}$ and parameters fixed as in (3.2.4).

Bibliography

[1] L. Accardi, Y. G. Lu, *A continuous version of de Finetti's theorem*, Ann. Probab. **21** (1993), 1478–1493.

[2] Y. AÏt-Sahalia, J. Jacod, *Testing for jumps in discretely observed processes*, Ann. Statistics **37** (2009), 184-222.

[3] R. T. Baillie, *Long memory processes and fractional integration in econometrics*, J. Econometrics **73** (1996), 5–59.

[4] L.Heston, *A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*, Rev. Fin. Studies **6** (1993), 327–343.

[5] D.B. Madan, E. Seneta, *The variance gamma (V.G.) model for share market returns*, J. Business **63** (1990), 511–524.

[6] O. E. Barndorff-Nielsen, N. Shephard, *Non Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics*, J. R. Statist. Soc. B **63** (2001), 167–241.

[7] R. Engle, *Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation*, Econometrica **55** (1982), 987–1008.

[8] P. Billingsley, *Probability and Measure*, Third Edition, John Wiley and Sons (1995).

[9] F. Baldovin, A. Stella, *Scaling and efficiency determine the irreversible evolution of a market*, PNAS **104**, n. 50 (2007), 19741–19744.

[10] T. Bollerslev, *Generalized Autoregressive Conditional Heteroskedasticity*, J. Econometrics **31** (1986), 307–327.

[11] T. Bollerslev, H.O. Mikkelsen, *Modeling and pricing long memory in stock market volatility*, J. Econometrics **31** (1996), 151–184.

[12] T. Bollerslev, U. Kretschmer, C. Pigorsch, G. Tauchen, *A discrete-time model for daily S & P500 returns and realized variations: Jumps and leverage effects*, J. Econometrics **150** (2009), 151–166.

[13] T. Bollerslev, V. Todorov, *Jump Tails, Extreme Dependencies, and the Distribution of Stock Returns*, CREATES Research Paper 2010-64 (2010).

[14] T. Di Matteo, T. Aste, M. M. Dacorogna, *Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development*, J. Banking Finance **29** (2005), 827–851.

[15] R. F. Engle, *Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation*, Econometrica **50** (1982), 987–1008.

[16] D. A. Freedman, *Invariants Under Mixing Which Generalize de Finetti's Theorem: Continuous Time Parameter*, Ann. Math. Statist. **34** (1963), 1194–1216.

[17] J. C. Hull, *Options, Futures and Other Derivatives*, Pearson/Prentice Hall (2009).

[18] J. Jacod, V. Todorov, *Testing for common arrivals of jumps for discretely observed multidimensional processes*, Ann. Statistics **37** (2009), 1792-1838.

[19] I. Karatzas, S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer (1988).

[20] P. E. Protter, *Stochastic Integration and Differential Equations*, Springer (2004).

[21] C. Kluppelberg, A. Lindner, R. A. Maller, *A continuous time GARCH process driven by a Lévy process: stationarity and second order behaviour*, J. Appl. Probab. **41** (2004) 601–622.

[22] C. Kluppelberg, A. Lindner, R. A. Maller, *Continuous time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models*, in *From Stochastic Calculus to Mathematical Finance*, Yu. Kabanov, R. Lipster and J. Stoyanov (Eds.), Springer (2007).

[23] Wolfram Research, Inc., *Mathematica*, Version 7.0, Champaign, IL (2008).

[24] B. Øksendal, Stochastic Differential Equations, Springer-Verlag (2003).

[25] R Development Core Team (2009), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. URL: `http://www.R-project.org`.

[26] A. L. Stella, F. Baldovin, *Role of scaling in the statistical modeling of finance*, Pramana **71** (2008), 341–352.

[27] L. Weiss, *The Stochastic Convergence of a Function of Sample Successive Differences*, Ann. Math. Statist. **26** (1955), 532–536.

[28] F. Delbaen, W. Schachermayer, *The Mathematics of Arbitrage*, Springer (2006).

[29] T. Bjoörk, *Arbitrage Theory in Continuous Time*, Oxford University Press (2009).