

A Comparative Analysis of Nonparametric
Volatility Estimators: An Empirical Evidence
using Option Pricing on S&P 500

Romuald Noel Kenmoe Siyou

22 february 2012

A Dissertation Presented to Department of Quantitative Methods and Business
Science of Milan-Bicocca University in Candidacy for the Degree of Doctor of Phi-
losophy

© Copyright by Romuald Noel Kenmoe Siyou, 2012. All Rights Reserved

Contents

1	Basic notions on stochastic process	7
1.1	Preliminaries	7
1.2	Stopping times and subdivisions	9
1.3	Martingales	10
1.3.1	Semimartingales	12
1.3.2	Stochastic integral	12
1.4	Quadratic variation	15
1.4.1	Stochastic differential equations	17
2	Integrated and Spot Volatility estimation	20
2.1	High frequency data in financial markets	20
2.2	Estimating integrated volatility	21
2.2.1	Estimation with no microstructure noise	22
2.2.2	The effect of microstructure noise	26
2.2.3	Reviews of integrated volatility estimators	30
2.2.3.1	Subsampling	30
2.2.4	Kernel-based estimators	33
2.2.5	Filters	35
2.3	Reviews of spot volatility estimators	36
3	Nonparametric estimation of the diffusion coefficient of local volatility models	42
3.1	Kernel density estimator	42

3.1.1	Definition and construction	42
3.1.2	Measure of discrepancy: mean square error and mean integrated square error	45
3.1.3	Approximate properties of bias and variance	46
3.1.4	The ideal window width and kernel	48
3.2	Local time	49
3.3	The estimation of the diffusion coefficient	52
4	Application of nonparametric estimation	63
4.1	Introduction	63
4.2	Spot volatility estimators	68
4.2.1	Nonparametric estimation	69
4.3	Methodology and results	70
4.4	Data and Empirical Methodology	79
4.5	Results	81
4.6	conclusion	81
5	Option pricing with nonparametric diffusion coefficient	84
5.1	Introduction	84
5.2	Stochastic volatility	87
5.3	Implementation method	93
5.4	Competitive methods	95
5.5	Implementation of numerical method	97
5.5.1	Discretization method	98
5.6	Empirical findings	102
5.6.1	Data description and characteristics	102
5.6.2	Model Comparison Results	108
5.7	Conclusion	116

Abstract

The accurate estimating of the volatility is essential in practically all financial models, viz derivatives pricing, portfolio selection and risk management. This thesis concerns with the problem of estimating the functional form of the diffusion coefficient. The estimation procedure considered here is based on the seminal approach introduced by Florens-Zmirou (1993). The approach is derived from a kernel regression introduced by Nadayara and Watson which is a two step procedures. Firstly we will locally construct the estimator of the spot variation by means of *realized volatility*. The realized volatility has gained much importance due to the increased availability of high-frequency data on practically every financial asset traded in the principal marketplaces. The importance of high frequency data for this purpose traces back from the observation of Merton (1980), who noted that the variance over a fixed interval can be estimated as the sum of squared realizations, provided the data are available at a sufficiently high sampling frequency, see Andersen et al. (2003). After estimating the quadratic variation, the diffusion coefficient can be found out readily using kernel regression.

In this dissertation, two main contributions are made: we review the existing nonparametric estimators in the literature, and then we use others new spot volatility to enlarge these classes of estimators. Secondly, we use the estimated volatility to price European call options written on S&P 500 index.

We organize our dissertation with the following structure.

In Chapter 1 we will focus on presenting basic notions of stochastic processes, semimartingale, stochastic integral and stochastic differential equation which are essential to introduce quadratic variation.

In Chapter 2, we review the main problems concerning high frequency data in finance. Successively we present the different estimators of the quadratic variation, and we also concern with the problems of correcting these estimators when the microstructure are taken into account. Furthermore, from integrated volatility, we derive spot volatilities from which the different nonparametric estimator are constructed.

In Chapter 3 we review the concept of the kernel estimator which will help us to construct our nonparametric volatility estimator and successively we are going to focus on the theory of *local time* with which we will prove the consistency and the asymptotic normality of the estimators.

In Chapter 4 we provide simulations to investigate the finite sample properties and the performance of the competing nonparametric estimators of the diffusion function proposed for the continuous time Itô-diffusion process, when discretely sampled observations over the fix time period are available.

Finally, in Chapter 5 we will use the nonparametric estimators for pricing european call options written on the S&P 500 index. This problem can be tackled using parametric approaches (Black & Schole, 1973; Derman & Kani, 1994; Dupire, 1994), as well as with nonparametric techniques (Stanton, 1997; Jiang, 1998; Ait-Sahalia & Lo, 1998). Following the last stream of the literature, we use the different nonparametric estimators of the diffusion coefficient for pricing options where the main purpose is to compare the competing estimators in their ability in figuring out a “right price”. Assuming that the underlying asset price follows diffusion processes, by imposing suitable conditions on the kernel, we can obtain the volatility function of the underlying asset-return process. The constructed volatility will be the continuous-state analog to the implied binomial tree proposed by Rubinstein (1994) and Derman and Kani (1994), and the implied volatility function in Dupire (1994), and Dumas, Fleming, and Whaley (1995). The prices obtained with the nonparametric are consistent with options pricing theory. Furthermore, the estimated volatility prevailing in our models evidences the main features of the true implied volatility and are in line generally with those in the literature for the six classes in which we have divided our sample.

Acknowledgements

I would like to thank my advisor, Professor Simona Sanfelici. Her trust and support helped me a lot, especially during the final steps of completing my work. I am greatly indebted to her for teaching me fundamental aspects in matlab programming. I express my deepest gratitude to professors Giovanni Zambruno for giving me the opportunity to join this doctoral school and to my tutor Fausto Mignanego for his encouragements and his help.

I owe many warm thanks to all my colleagues, particularly Ilaria Peri whom i passed three nice years.

A special thank goes to my parents for putting all the necessary as far as learning is concerned. Thank you to my brothers Cyrille and Eric for their encouragements and their support.

The end and not the least, i thank my fiancee Gabrielle for all her encouragements and all the discussions we have had during the achievement of this dissertation.

Chapter 1

Basic notions on stochastic process

In this preliminary Chapter, after presenting basic notions on stochastic processes, we are subsequently going to define the quadratic variation and its properties as an important tool in the stochastic process theory (Protter , 2004).

1.1 Preliminaries

We assume as given a complete probability space (Ω, \mathcal{F}, P) . In addition we are given a filtration $(\mathcal{F}_t)_{0 \leq t \leq \infty}$. By filtration we mean a family of σ -algebras $(\mathcal{F}_t)_{0 \leq t \leq \infty}$ that is increasing, i.e., $\mathcal{F}_s \subset \mathcal{F}_t$ if $s \leq t$. For convenience, we will usually write \mathfrak{F} for the filtration $(\mathcal{F}_t)_{0 \leq t \leq \infty}$

Definition 1. A filtered complete probability space $(\Omega, \mathcal{F}, \mathfrak{F}, P)$ is qualified to satisfy the **usual hypotheses** if

- (i) \mathcal{F}_0 contains all the P -null sets of \mathcal{F} ;
- (ii) $\mathcal{F}_t = \bigcap_{u > t} \mathcal{F}_u$, all t , $0 \leq t \leq \infty$; that is, the filtration \mathfrak{F} is right continuous.

We continue by introducing the definition of a process, and some others useful definitions.

Definition 2. A process is a family $X = (X_t)_{t \in \mathbb{R}^d}$ of random variables from Ω to some set E .

In our case, the set E will be usually \mathbb{R} . A process can be thought as a mapping from $\Omega \times \mathbb{R}$ into E .

Definition 3. A trajectory of the process X is defined as the mapping $t \rightarrow X(\omega)$ for a fixed $\omega \in \Omega$.

A stochastic process X is said to be càdlàg if a.s. all its trajectories are right-continuous and admit left-hand limits. It is said to be càg if a.s. all its trajectories are left-continuous.

If a process is càdlàg, it is always possible to define two others processes, X_- and ΔX as follow:

$$X_{0-} = X_0, \quad X_{t-} = \lim_{s \rightarrow t} X_s \quad (1.1.1)$$

$$\Delta X_t = X_t - X_{t-} \quad (1.1.2)$$

If the trajectory is continuous in t , then $X_{t-} = X_t$ and $\Delta X_t = 0$.

Definition 4. A process is said to be adapted to the filtration \mathfrak{F} if $X_t \in \mathcal{F}_t$ (that is, \mathcal{F}_t -measurable) for every $t \in \mathbb{R}_+$

Since the concept of quadratic variation is the cornerstone in studying realized volatility, it is important to identify processes of finite variation for every trajectory. That is

$$Var(X) = \lim_{n \rightarrow \infty} \sum_{1 \leq k \leq n} |X_{tk/n}(w) - X_{t(k-1)/n}(w)| < \infty \quad (1.1.3)$$

We denote by \mathcal{V} the set of all real-valued processes X that are càdlàg, adapted, with $X_0 = 0$ and whose trajectory $X_t(w)$ has a finite variation over each finite interval $[0, t]$.

We then abbreviate by $X \in \mathcal{V}$ the fact that X is an adapted process with finite variation.

Before ending this subsection we give the definition of increasing and predictable processes

Definition 5. A process X is said to be increasing if it is càdlàg, adapted, with $X_0 = 0$ and such that each trajectory is non-decreasing.

Definition 6. A σ -field is said to be predictable, if it is the σ -field on $\Omega \times \mathbb{R}_+$ that is generated by all càg adapted processes. Every process which is measurable with respect to the predictable σ -field is predictable.

1.2 Stopping times and subdivisions

The concept of stopping time is very useful in econometric analysis, since economic data are recorded at discrete points in time.

Definition 7. A random variable $T : \Omega \rightarrow \mathbb{R}_+$ is a stopping time if the event $\{\omega \mid T(\omega) \leq t\} \in \mathcal{F}_t$ for all $t \in \mathbb{R}_+$.

Given a process X and a stopping time T , we define the *stopped process* as $X_t^T = X_{T \wedge t}$. Among other things, stopping times are necessary to introduce the localization procedure.

Definition 8. If \mathcal{C} is a class of processes, we define the localized class \mathcal{C}_{loc} as follows: a process X belong to \mathcal{C}_{loc} if and only if there exists an increasing sequence T_n of stopping time such that $\lim_{n \rightarrow \infty} T_n = \infty$ a.s. and that each stopped process $X^{T_n} \in \mathcal{C}$.

The sequence T_n is called *localizing sequence*. It is clear that $\mathcal{C} \subset \mathcal{C}_{loc}$.

A sequence τ_n of stopping time with $\tau_0 = 0$, $\sup_{n \in \mathbb{N}} \tau_n < \infty$ and $\tau_n < \tau_{n+1}$ on the set $\{\tau_n < \infty\}$ is called an *adapted subdivision*. Among subdivisions, we will consider the Riemann sequence.

Definition 9. A sequence $\tau_{n,m}$, $m \in \mathbb{N}$ of adapted subdivisions is called a *Riemann sequence* if $\lim_{n \rightarrow \infty} \sup_{m \in \mathbb{N}} [\tau_{n,m+1} - \tau_{n,m}] = 0$ for all $t \in \mathbb{R}_+$.

1.3 Martingales

Among processes, a very important class is given by martingales. In this section, we describe some of their basic properties which we shall use throughout this work.

Definition 10. A real-valued, adapted process $X = (X_t)_{t \in \mathbb{R}_+}$ is called a martingale with respect to the filtration \mathfrak{F} if

- (i) $X_t \in L^1(dP)$; that is, $E\{|X_t|\} < \infty$;
- (ii) if $s \leq t$, then $X_s = E[X_t | \mathcal{F}_s]$ a.s.

Definition 11. A martingale X is square-integrable if $\sup_{t \in \mathbb{R}_+} E[X_t^2] < \infty$.

Two special classes of martingales play an important role when dealing with local time analysis, namely: local martingales and locally square-integrable martingales.

Definition 12. A locally square-integrable martingale is a process that belongs to the localized class constructed from the class of square-integrable martingales.

Definition 13. A local martingale is a process that belongs to the localized class of uniformly integrable martingales, that is of martingales X such that the family of random variable is uniformly integrable.

The above definition follows by the fact that, each martingale X which is locally squared-integrable is a local martingale. The class of local martingale can be obtained by localization of the class of martingale also. This allows us to state the following:

Proposition 14. *Each martingale is a local martingale*

Proof. Let X be a martingale, and consider the sequence of stopping times $T_n = n$. Then, for every $t \in \mathbb{R}_+$, we have $X_t^{T_n} = E[X_n | \mathcal{F}_t]$. Since the class of uniformly integrable martingales is stable under stopping, we have that X^{T_n} is uniformly integrable as well. \square

$$\text{Let } X_t^* = \sup_{s \leq t} |X_s| \text{ and } X^* = \sup_s |X_s|.$$

Theorem 15. *Let X be a local martingale such that $E\{X_t^*\} < \infty$ for every $t \geq 0$. Then X is a martingale. If $E\{X^*\} < \infty$, then X is a uniformly integrable martingale.*

Proof. Let $(T_n)_{n \geq 1}$ be a *fundamental sequence* of stopping times for X . Provided $s \leq t$, then $E\{X_{t \wedge T_n} | \mathcal{F}_s\} = X_{s \wedge T_n}$. And applying the Dominated Convergence Theorem, we obtain $E\{X_t | \mathcal{F}_s\} = X_s$. If $E\{X^*\} < \infty$, since each $|X_t| \leq X^*$, it follows that $(X_t)_{t \geq 0}$ is uniformly integrable. Where we call fundamental sequence a sequence of increasing stopping time, T_n , with $\lim_{n \rightarrow \infty} T_n = \infty$ a.s. such that $X_{t \wedge T} 1_{\{T_n > 0\}}$ is a uniform integrable for each n . \square

Definition 16. Two local martingales M, N are said to be orthogonal if their product MN is a local martingale. A local martingale X is called a *purely discontinuous* local martingale if $X_0 = 0$ and if it is orthogonal to all continuous local martingales.

Proposition 17. *Let X be always a local martingale, then*

1. *A local martingale X is orthogonal to itself if and only if X_0 is square integrable and $X = X_0$ up to null sets.*
2. *A purely discontinuous local martingale which is continuous is a.s. equal to 0.*
3. *A local martingale X with $X_0 = 0$ is purely discontinuous if and only if it is orthogonal to all continuous bounded martingales Y with $Y_0 = 0$.*
4. *A local martingale in \mathcal{V} is purely discontinuous.*

Theorem 18. *Any local martingale X admits a unique (up to null sets) decomposition:*

$$X = X_0 + X^c + X^d \tag{1.3.1}$$

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

Theorem 19. *For each pair M, N of locally square-integrable martingales there exists a unique, up to null measure sets, predictable process $\langle M, N \rangle \in \mathcal{V}$ such that $MN - \langle M, N \rangle$ is a local martingale.*

1.3.1 Semimartingales

Let us denote by \mathcal{L} the set of all local martingales M such that $M_0 = 0$.

Definition 20. A semimartingale is a decomposable process X of the form $X = X_0 + M + A$ where X_0 is finite-valued and \mathcal{F}_0 -measurable, $M \in \mathcal{L}$ and $A \in \mathcal{V}$.

If there exists a decomposition such that A is predictable, X is called a *special semimartingale*.

From the definition is clear that if $X \in \mathcal{V}$ then it is a semimartingale. Obviously the decomposition $X = X_0 + M + A$ is not unique, but if X is a special semimartingale then there is a unique decomposition with A predictable (Back, 1991). Given that a semimartingale is the sum of a local martingale and a process of finite variation, we can naturally decompose it in a continuous and discontinuous part in the same fashion of Theorem 18:

Proposition 21. *Let X be a semimartingale. Then there is a unique (up to null sets) continuous local martingale X^c such that $X^{c,0} = 0$ and any decomposition $X = X_0 + M + A$ of type (1.3.1) meets $M^c = X^c$ up to null sets.*

Proof. It is enough to use Theorem 18 and Proposition 17. □

Proposition 22. *Let $F(t)$ be a real-valued function on \mathbb{R}^+ , and define the process $X(t) = F(t)$. Then X is a semimartingale if and only if F is càdlàg, with finite-variation over each finite interval.*

Proof. For the sufficiency, it is enough to use the definition of semimartingales. For the converse, see Jacod and Shiryaev (1987),

1.3.2 Stochastic integral

Let A be an increasing process, fix ω such that, if f is a bounded Borel function on \mathbb{R}_+ , we denote by $\int_0^t f(s)dA_s(\omega)$ the Stieltjes integral of f with respect to $A(\omega)$

Let H be a càdlàg process, when the integrand H has *continuous paths*, the Stieltjes integral $\int_0^t H_s dX_s$ is also known as the *Riemann-Stieltjes* integral (for fixed ω),

Where $X \in \mathcal{V}$. In this case we can define the integral as a limit of an approximated sum.

Theorem 23. Let $X \in \mathcal{V}$ and let H be a jointly measurable process such that a.s. $s \rightarrow H(s, \omega)$ is continuous. Let π_n be a sequence of finite random partitions of $[0, t]$ with $\lim_{n \rightarrow \infty} \text{mesh}(\pi_n) = 0$. Then for $T_k \leq S_k \leq T_{k+1}$,

$$\lim_{n \rightarrow \infty} \sum_{T_k, T_{k+1} \in \pi_n} H_{S_k} (X_{T_{k+1}} - X_{T_k}) = \int_0^t H_s dX_s \text{ a.s.} \quad (1.3.2)$$

This definition stems from the fact that, if $X \in \mathcal{V}$, then its trajectories are the distribution functions of a signed measure. We want now to define the stochastic integral when X is a semimartingale. In this case, the trajectories do not necessarily define a measure; for example, the Wiener process has infinite variation over each finite interval. Now consider a generic process X . The stochastic integral can be naturally defined for processes H such that

$$H_t = H_0 1_{\{0\}} + \sum_{i=1}^n H_i 1_{(T_i, T_{i+1}]}(t)$$

where $0 = T_1 \leq \dots \leq T_{n+1} < \infty$ is a finite sequence of stopping times, $H_i \in \mathcal{F}_{T_i}$ with $|H_i| < \infty$ a.s., $0 \leq i \leq n$. We can then define:

$$J_X(H) = H_0 X_0 + \sum_{i=1}^n H_i (X^{T_{i+1}-T_i}) \quad (1.3.3)$$

The peculiar property of semimartingales is that this definition can be extended to any locally bounded predictable process H if and only if X is a semimartingale. The feasibility of the extension for semimartingales is stated in the following theorem

Theorem 24. Let X be a semimartingale. Then the mapping (1.3.3) has an extension to the space of all locally bounded predictable processes H , with the following properties:

- 1) $G_t = \int_0^t H_s dX_s$ is a càdlàg adapted process
- 2) the mapping $H \rightarrow \int HdX$ is linear
- 3) If a sequence H^n of predictable processes converges pointwise to a limit H , and if $|H^n| < K$, where K is a locally bounded predictable process, then $\int_0^t H_s^n dX_s$ converges to $\int_0^t H_s dX_s$ in measure for all $t \in \mathbb{R}_+$.

Moreover this extension is unique, up to null measure sets, and in 3) above the convergence is in measure, uniformly on finite intervals: $\sup_{s \leq t} \left| \int_0^t H_s^n dX_s - \int_0^t H_s dX_s \right| \rightarrow 0$.

A complete proof of the above Theorem can be found in Dellacherie and Meyer (1976). It is important to state the following properties, which we state without proof.

Proposition 25. *Let X be a semimartingale and H, K be locally bounded predictable process. Then the following properties hold up to null sets:*

1. The mapping $H \rightarrow \int HdX$ is linear.
2. $\int HdX$ is a semimartingale; if X is a local martingale, then $\int HdX$ is a local martingale
3. If $X \in \mathcal{V}$ then $\int HdX \in \mathcal{V}$.
4. $(\int HdX)_0 = 0$ and $\int HdX = \int Hd(X - X_0)$.
5. $\Delta(\int HdX) = H \Delta X$
6. $\int K d(\int HdX) = \int HK dX$.

The stochastic integral of a predictable process that is left-continuous can be approximated by Riemann sums. Consider a subdivision τ_n . Then the τ -Riemann approximant of the stochastic integral $\int HdX$ is defined as the process $\tau(\int HdX)$ defined by

$$\tau\left(\int HdX\right)_t = \sum_{n \in \mathbb{N}} H_{\tau_n} (X_{\tau_{n+1} \wedge t} - X_{\tau_n \wedge t}) \quad (1.3.4)$$

We then have the following:

Proposition 26. *Let X be a semimartingale, H be a càg adapted process and τ_n -Riemann sequence of adapted subdivisions. Then the τ_n -Riemann approximants converge to $\int HdX$, in measure uniformly on each compact interval*

Proof. Consider $\tau_{n,m}$ and define H_n by

$$H_n = \sum_{m \in \mathbb{N}} H_{\tau_{n,m}} \mathbb{1}_{] \tau_{n,m}, \tau_{n,m+1}]} \quad (1.3.5)$$

Then H_n is predictable, converges pointwise to H , since H is càg. Now consider $K_t = \sup_{s \leq t} |H_s|$. The process K is adapted, càg, locally bounded and $|H^n| < K$. Then the result follow from Theorem 23 and from the property $\tau_n(\int H dX) = \int H_n dX$. \square

1.4 Quadratic variation

The quadratic variation process of two semimartingales, also known as the bracket process, is a simple object that nevertheless plays a fundamental role, especially in mathematical finance. In this subsection we define it and state its most important properties.

Definition 27. The quadratic variation of two semimartingales X and Y is defined by the following process:

$$[X, Y]_t = X_t Y_t - X_0 Y_0 - \int_0^t X_{s-} dY_s - \int_0^t Y_{s-} dX_s \quad (1.4.1)$$

From the definition itself, it is straightforward to verify the following properties:

Proposition 28. *The quadratic variation of two semimartingales X, Y has the following properties:*

1. $[X, Y]_0 = 0$
2. $[X, Y] = [X - X_0, Y - Y_0]$
3. $[X, Y] = \frac{1}{4}([X + Y, X + Y] - [X - Y, X - Y])$ (polarization)

The following analysis is crucial for at least two reasons. First, the name quadratic variation comes after Theorem 29. Second, it is the basis for realized volatility, a

concept which will be illustrated in the following chapters. Indeed, it allows an estimation of quadratic variation.

Theorem 29. *Consider a sequence of possibly random partitions of $[0, T]$. Then for every Riemann sequence $(\tau_n) \equiv \{\tau_{n,m}\}_{m \geq 0}$, $n = 0, 1, 2, \dots$ where $\tau_{m,0} \leq \tau_{m,1} \leq \tau_{m,2} \leq \dots$ satisfy, with probability one, for $n \rightarrow \infty$.*

*$\tau_{n,0} \rightarrow 0$; $\sup_{m \geq 0} \tau_{n,m} \rightarrow T$; $\sup_{m \geq 0} (\tau_{n,m+1} - \tau_{n,m}) \rightarrow 0$ of adapted subdivisions,
Then for $t \in [0, T]$ the process $S_{\tau_n}(X, Y)$ defined by:*

$$S_{\tau_n}(X, Y)_t = \sum_{m \geq 1} (X_{\tau_{n,m+1 \wedge t}} - X_{\tau_{n,m \wedge t}})(Y_{\tau_{n,m+1 \wedge t}} - Y_{\tau_{n,m \wedge t}}) \quad (1.4.2)$$

where $t \wedge \tau = \min(t, \tau)$ converges, for $n \rightarrow \infty$, to the process $[X, Y]_t$, in measure and uniformly on every compact interval.

Proof. By polarization, it suffices to prove the claim for $X = Y$. From equation (1.3.4) we get: $S_{\tau_n}(X, X) = X^2 - X_0^2 - 2\tau_n(\int X_- dX)$ \square

The last term converges to $\tau(\int X_- dX)$, then $S_{\tau_n}(X, X)$ converges to $[X, X]$.

Proposition 30. *Let X and Y be two semimartingales.*

1. $[X, Y] \in \mathcal{V}$
2. $[X, X]$ is increasing.
3. $\Delta[X, Y] = \Delta X \Delta Y$
4. If T is a stopping time, then $[X^T, Y] = [X, Y^T] = [X^T, Y^T]$

The property 3 implies that if X or Y is continuous, then $[X, Y]$ is continuous as well.

Proposition 31. *If X are a special semimartingale and $Y \in \mathcal{V}$ then:*

1. $[X, Y]_t = \int_0^t \Delta X_s dY_s$

2. $X_t Y_t = \int_0^t Y_s dX_s + \int_0^t X_s dY_s$
3. If Y is predictable, then $[X, Y]_t = \int_0^t \Delta Y_s dX_s$
4. If X and Y are continuous then $[X, Y] = 0$

Lemma 32. *Let X be a purely discontinuous square-integrable martingale. Then*

$$[X, X] = \sum_{s \leq t} (\Delta X_s)^2$$

Theorem 33. *If X and Y are semimartingales, and if X^c, Y^c denote their continuous martingale parts, then:*

$$[X, Y]_t = \langle X^c, Y^c \rangle_t + \sum_{s \leq t} \Delta X_s \Delta Y_s \quad (1.4.3)$$

where the process $\langle X^c, Y^c \rangle$ is called the predictable quadratic variation of the pair (X, Y) .

Theorem 34. *(Knight's Theorem) Let M_1, \dots, M_n be orthogonal square-integrable martingales, and consider the time changes:*

$$T_i(t) = \begin{cases} \inf_s [B_i, B_i]_s > t & \text{if this is finite} \\ +\infty & \text{otherwise} \end{cases} \quad (1.4.4)$$

Then the transformed variables:

$$X_i(t) = \begin{cases} B_i(T_i(t)), & \text{if } T_i(t) < \infty \\ B_i(\infty) + W_i(t - [B_i, B_i]_\infty) & \text{otherwise} \end{cases} \quad (1.4.5)$$

where W_1, \dots, W_n is an n -dimensional Brownian motion independent of X_i , are an n -dimensional Brownian motion relative to their generated filtration.

1.4.1 Stochastic differential equations

In this subsection, we are concerned with the following equation:

$$X(t) = \eta + \int_0^t \beta(s, X(s))ds + \int_0^t \sigma(s, X(s))dW(s), \quad (1.4.6)$$

where $W(s)$ is the standard Wiener process, which satisfy the usual hypothesis, and we look for an adapted process $X(t) \in L^2(\Omega)$. The functions σ, β are applications from $[0, T] \times L^2(\Omega) \rightarrow L^2(\Omega)$, while η is an \mathcal{F}_0 measurable process in $L^2(\Omega)$, that is the boundary condition. It is common to write equation (1.4.6) in the shorthand notation:

$$\begin{cases} dX(t) = \beta(t, X(t))dt + \sigma(t, X(t))dW(t) \\ X(0) = \eta \end{cases} \quad (1.4.7)$$

Theorem 35. *Assume the following assumptions hold for the existence and uniqueness of solutions:*

1 β and σ are continuous

2 There exists $M > 0$ such that

$$\|\beta(t, \varsigma)\|^2 + \|\sigma(t, \varsigma)\|^2 \leq M^2(1 + \|\varsigma\|^2) \quad \forall t \in [0, T], \varsigma \in L^2(\Omega)$$

$$\|\beta(t, \varsigma_1) - \beta(t, \varsigma_2)\| + \|\sigma(t, \varsigma_1) - \sigma(t, \varsigma_2)\| \leq M(\|\varsigma_1 - \varsigma_2\|) \quad \forall t \in [0, T], \varsigma_1, \varsigma_2 \in L^2(\Omega)$$

3 $\forall t \in [0, T], \varsigma \in L^2(\Omega)$ such that ς is \mathcal{F}_t -measurable, we have that $\beta(t, \varsigma), \sigma(t, \varsigma) \in L^2(\Omega)$ and are \mathcal{F}_t -measurable

Let $\eta \in L^2(\Omega)$ and \mathcal{F}_0 -measurable. Then there exists a unique (up to null sets) adapted process $X(t) \in L^2(\Omega)$ fulfilling equation (1.4.6).

Corollary 36. *Assume the hypothesis of Theorem 35 hold. Then the (unique) solution process X is a continuous semimartingale, and*

$$[X, X]_t = \int_0^t \sigma^2(s, X(s))ds \quad (1.4.8)$$

The process in (1.4.8) is commonly called integrated volatility or integrated variance in financial econometrics.

Proof. The result comes from the fact that, η is \mathcal{F}_0 -measurable and a finite-valued, $\int_0^t \beta(s, X(s))ds$ is of finite variation, so their quadratic variation is zero and $\int_0^t \sigma(s, X(s))dW(s)$

is a local martingale, since Wiener process is a local martingale and using the proposition 25. The result follows! \square

Lemma 37. *Consider the process Y_t^n defined on $(\Omega, (\mathcal{F}_i)_{i \in \mathbb{N}}, P)$, as*

$$Y_t = \sum_{i=1}^{\lfloor nt/T \rfloor} \Theta_i,$$

where $\lfloor x \rfloor$ is the integer part of x , with Θ_i bounded and adapted to \mathcal{F}_{i-1} , and make the following assumptions:

- (i) $\sum_{i=1}^{\lfloor nt/T \rfloor} \mathbb{E} [\Theta_i \mid \mathcal{F}_{i-1}] \rightarrow 0$ in probability;
- (ii) $\sum_{i=1}^{\lfloor nt/T \rfloor} \mathbb{E} [\Theta_i^2 \mid \mathcal{F}_{i-1}] \rightarrow V_t$ in probability;
- (iii) $\forall \varepsilon > 0, \sum_{i=1}^{\lfloor nt/T \rfloor} \mathbb{E} [\Theta_i^2 I_{\{|\Theta_i| > \varepsilon\}} \mid \mathcal{F}_{i-1}] \rightarrow 0$ in probability; (conditional Lindeberg condition).

Then Y_t converges in distribution to the continuous martingale M_t with quadratic variation $[M, M] = V_t$.

It can be tempting to investigate the link between integrated volatility and realized volatility. Different estimators have been proposed for estimating the integrated volatility. Nowadays, the most popular is realized volatility, The idea behind realized volatility hinges on Theorem 29 ; which will be discussed thoroughly in the next chapter.

Chapter 2

Integrated and Spot Volatility estimation

2.1 High frequency data in financial markets

One of the most important success in the last years consist in studying of the high-frequency data, namely the observations recorded in real-time on the market. Unlike traditional studies, in which the data are measured in equispaced interval, in the case of high-frequency data, any single traded transaction occurred in the market is recorded. The possibility to exploit this huge quantity of information constitute an important advantage because it allows to an accurated estimate of the volatility. These estimations intervene almost in all financial applications including derivatives pricing, portfolio selection, risk management and the liquidity management. Many years ago financial applications were focused on the daily observation, nowadays, many financial companies take advantage of the maximum information by using *tick-by-tick* data. Every *tick* is considered as a logic information unit, as a quotation or a transaction price;

Using daily data, the closed price is considered and this reflects the trend of entire negotiation daytime. This consideration can be misleading, in fact, when intraday movement are not taken into account, the risk of disregarding relevant information

is influent.

Nevertheless, using this type of data open door to several problems as far as their collection and their handling are concerned. In fact, besides the problem of management and put on a large quantity of observations, (million of data can be recorded during one trading day), the observations are not equispaced, Engle and Russell proposed ACD (Autoregressive Conditional Duration) for tackling the latter problem. This regularization consist of modeling the time elapsed between a transaction and the other. One of the major concerns of the high frequency data is the microstructure noise. In the subsection below, we will present several methods of volatility estimation when data are contaminated.

2.2 Estimating integrated volatility

Different estimators for the integrated volatility have been proposed in the literature. Nowadays, the most popular is realized volatility, which will be discussed thoroughly throughout. The idea behind realized volatility hinges on Theorem 29. Suppose that, in a given trading day t , the logarithmic prices are observed tick-by-tick. Consider a grid $\Lambda_t = \{k_0, \dots, k_{n_t}\}$ containing all observation points, and set $p_{t,i}$, $i = 1, \dots, n_t$, to be the i -th price observation during day t , where n_t is the total number of observations at day t .

$r_{t,i} = p_{t,i} - p_{t,i-1}$ is the i -th intraperiod return of day t such that

$$r_t = \sum_{i=0}^{n_t} r_{t,i}$$

Then the realized volatility is given by:

$$RV^{(TOT)} = \sum_{i=1}^{n_t} r_{t,i}^2 = \sum_{i=1}^{n_t} (p_{t,i} - p_{t,i-1})^2 \quad (2.2.1)$$

which is the sum of all available intraday high frequencies squared returns.

In Malliavin and Mancino (2002) it is showed that :

$$\lim_{n \rightarrow \infty} RV^{(TOT)} = \int_{t-1}^t \sigma^2(s, p(s)) ds \quad a.s. \quad (2.2.2)$$

2.2.1 Estimation with no microstructure noise

Suppose that, along day t , the logarithmic prices of a given asset follow a continuous time diffusion process, as follows:

$$dp(t+k) = \mu(t+k)dt + \sigma(t+\tau)dW(t+k), \quad 0 \leq k \leq 1, \quad t = 1, 2, \dots \quad (2.2.3)$$

where $p(t+k)$ is the logarithmic price at time $t+k$, $\mu(t+k)$ is the drift component, $\sigma(t+k)$ is the spot or instantaneous volatility (or standard deviation), and $W(t+k)$ is a standard Brownian motion. In addition, suppose here is a null correlation between $\sigma(t+k)$ and $W(t+k)$ (there are orthogonal), such that there is no leverage effect. Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002) showed that daily returns, defined as $r_t = p(t) - p(t-1)$, are Gaussian conditionally on $\mathcal{F}_t \equiv \mathcal{F}\{\mu(t+k-1), \sigma(t+k-1)\}_{0 \leq k \leq 1}$, the σ -algebra (information set) generated by the sample paths of $\mu(t+k-1)$ and $\sigma(t+k-1)$, $0 \leq k \leq 1$

$$r_t / \mathcal{F}_t = N \left(\int_0^1 \mu(t+k-1)dk, \int_0^1 \sigma^2(t+k-1)dk \right)$$

The term $IV_t = \int_0^1 \sigma^2(t+k-1)dk$ is known as the integrated variance of the period $[t-1, t]$, which is a measure of the day- t ex post volatility. The integrated variance is typically the object of interest as a measure of the true daily volatility.

Under the model (2.2.3), the approximation in (2.2.1) is justified by the theoretical results in stochastic processes which state that

$$plim \sum_{i=1}^{n_t} (p_{t,i} - p_{t,i-1})^2 = \int_0^t \sigma^2(t)dt \quad (2.2.4)$$

as the sampling frequency increases.

In practice, prices are observed at discrete and irregularly spaced intervals. In this sense, there are many ways in which one can sample the data. We always suppose that in a given day t , we partition the interval $[0, 1]$ in n_t subintervals and define the grid of observation times as $\Lambda_t = \{k_0, \dots, k_{n_t}\}$, where $0 = k_0 < k_2 < \dots < k_{n_t} = 1$. The length of the i -th subinterval is given by $\delta_{i,n_t} = k_i - k_{i-1}$. It is assumed that the length of each subinterval shrinks to zero as the number of intraday observations increases. The integrated variance over each of the subintervals is defined as

$$IV_{i,t} = \int_{k_{i-1}}^{k_i} \sigma^2(t + k - 1)dk$$

There are several sampling schemes that can be used, as follows:

1. The most widely used sampling scheme is *calendar time sampling* (CTS), where the intervals are equidistant in calendar time, that is, $\delta_{i,n_t} = \frac{1}{n_t}$ for all i . For example, the prices may be sampled every 5 or 15 minutes. As the intraday data are irregularly spaced, in most cases calendar time sampled data must be constructed artificially (see Andersen and Bollerslev, 1997). Hansen and Lunde (2006b) showed that the previous tick method is a sensible way to sample prices in calendar time. For example, during a five-minute interval, we may observe several prices, in which case the previous tick method takes the first observation as the sampled price.

2. Another sampling alternative is *transaction time sampling* (TrTS), where prices are recorded every m -th transaction.

3. The third sampling scheme is known as *business time sampling* (BTS), where the sampling times are chosen such that $IV_{i,t} = \frac{IV_t}{n_t}$.

4. The last sampling alternative is called *tick time sampling* (TkTS), where prices are recorded at every price change.

An important difference among these distinct sampling schemes is that the observation times in BTS are latent, whereas in CTS, TrTS, and TkTS they are observed.

Several authors have studied the behaviour of realized volatility: Andersen et al. (2003) showed, using a seminal result in semimartingale process theory that, the realized variance using all data available, as defined in equation (2.2.1), is a consistent

estimator of the integrated variance when there is no microstructure noise, such that

$$RV_t^{(TOT)} \xrightarrow{p} IV_t$$

From the results in Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002) derived the asymptotic distribution of the realized variance as

$$n_t^{1/2} \frac{1}{\sqrt{2IQ_t}} (RV_t^{(TOT)} - IV_t) \xrightarrow{\mathcal{L}} N(0, 1), \quad (2.2.5)$$

where the integrated quarticity, IQ_t , is defined as

$$IQ_t = \int_0^1 \sigma^4(t+k-1)dk \quad (2.2.6)$$

Proof. Consider a real number $k = T/n$. We define the equally space sampled returns as:

$$\Delta_j X = X_{jk} - X_{(j-1)k}, \quad j = 1, \dots, n \quad (2.2.7)$$

and we define the realized variance as:

$$RV_k(X)_t = \sum_{j=1}^{[t/k]} (\Delta_j X)^2 \quad (2.2.8)$$

It follows from Ito's lemma that:

$$X_t^2 = [X]_t + 2 \int_0^t X_s dX_s$$

thus

$$r_j^2 = (X_{jk} - X_{(j-1)k})^2 = [X]_{jk} - [X]_{(j-1)k} + 2 \int_0^t (X_s - X_{(j-1)k}) dX_s$$

which implies

$$k^{-1/2} (RV_k(X)_t - [X]_t) = 2k^{-1/2} \sum_{j=1}^{[t/k]} \int_{k(j-1)}^{kj} (X_s - X_{(j-1)k}) dX_s$$

$$2 \int_0^{k[t/k]} (X_s - X_{k[t/k]}) dX_s$$

As $k \rightarrow 0$, the right term converges to (Jacod and Protter,1998);

$$k^{-1/2} \int_0^{k[t/k]} (X_s - X_{k[t/k]}) dX_s \rightarrow \frac{1}{\sqrt{2}} \int_0^t \sigma_s^2 dw_s^t$$

so that

$$k^{-1/2} (RV_k(X)_t - [X]_t) \rightarrow MN(0, 2 \int_0^t \sigma_s^4 ds) \quad (2.2.9)$$

□

Bandi and Russell (2005a) gave an alternative simple proof of the above result.

Furthermore, under the assumption of no microstructure noise, Barndorff-Nielsen and Shephard (2002) showed that the integrated quarticity is consistently estimated by the realized quarticity, which is defined as

$$RQ_t^{(TOT)} = \frac{n_t}{3} \sum_{i=0}^{n_t} r_{t,i}^4 \quad (2.2.10)$$

and

$$n_t^{1/2} \frac{1}{\sqrt{\frac{2}{3} IQ_t}} (RV_t^{(TOT)} - IV_t) \xrightarrow{d} N(0, 1),$$

This asymptotic result allows to construct feasible confidence intervals for the realised volatility estimator.

Barndorff-Nielsen and Shephard (2005b), Meddahi (2002), and Nielsen and Frederiksen (2006) studied the finite sample behavior of the limit theory given in (2.2.4). The main conclusion is that (2.2.4) is poorly sized, but

$$n_t^{1/2} \frac{1}{\sqrt{\frac{2}{3} \frac{RQ_t^{(TOT)}}{(RV_t^{(TOT)})^2}}} [\log(RV_t^{(TOT)}) - \log(IV_t)] \xrightarrow{d} N(0, 1),$$

performs quite well.

2.2.2 The effect of microstructure noise

In this section we discuss the effects of the presence of microstructure noise in the estimation of the integrated variance. Market microstructure noise has many sources, including the discreteness of the price (see Harris, 1990, 1991), and properties of the trading mechanism, as in Black (1976). As in the previous section, we consider the grid of observation times, $\Lambda_t = \{k_0, \dots, k_{n_t}\}$. Following the notation in Zhang et al. (2005), we set $p_{t,i} \equiv p(t + k_i)$. Suppose also that the logarithmic prices are observed with noise, that is,

$$p_{t,i} = p_{t,i}^* + \varepsilon_{t,i}, \quad (2.2.11)$$

where $p_{t,i}^*$ is the latent true, or efficient price process and $\varepsilon_{t,i}$, is the microstructure noise.

It follows that

$$r_{t,i} = r_{t,i}^* + \varepsilon_{t,i} - \varepsilon_{t,i-1} = r_{t,i}^* + \nu_{t,i} \quad (2.2.12)$$

where $r_{t,i}^* = p_{t,i}^* - p_{t,i-1}^*$ is the efficient return. It is clear that $r_{t,i}$, is an autocorrelated process, then $RV_t^{(TOT)}$ cannot be an unbiased estimator of the latent daily volatility, as discussed in section 2.2.1. Furthermore as

$$RV_t^{(TOT)} = \sum_{i=1}^{n_t} (r_{t,i}^*)^2 + 2 \sum_{i=1}^{n_t} r_{t,i}^* \nu_{i,t} + \sum_{i=1}^{n_t} \nu_{t,i}^2,$$

With straightforward calculus, it is possible to show that, conditionally on the efficient returns,

$$E(RV_t^{(TOT)} | r^*) = RV_t^{*(TOT)} + 2n_t E(\varepsilon_{t,i}^2),$$

such that $RV_t^{(TOT)}$ is also a biased estimator of the integrated variance under microstructure effects.

As in Bandi and Russell (2005a), consider the following assumption regarding the noise structure.

Assumption 1 (Noise Structure).

(a) The microstructure noise, $\varepsilon_{t,i}$, has zero mean and is a covariance stationary stochastic process.

(b) The variance of $\nu_{t,i} = \varepsilon_{t,i} - \varepsilon_{t,i-1}$ is $O(1)$.

Under Assumption 1, Bandi and Russell (2005a) showed that

$$RV_t^{(TOT)} \xrightarrow{a.s.} \infty \text{ as } n_t \rightarrow \infty$$

Furthermore, consider the following assumption.

Assumption 2

(a) The microstructure noise, $\varepsilon_{t,i}$, has zero mean and is an independent and identically distributed random variable.

(b) The noise is independent of the price process.

(c) The variance of $\nu_{t,i} = \varepsilon_{t,i} - \varepsilon_{t,i-1}$ is $O(1)$.

Under Assumption 2, it was shown in Zhang et al. (2005) that

$$n_t^{-1/2} [RV_t^{(TOT)} - IV_t - 2n_t E(\varepsilon_{t,i}^2)] \xrightarrow{d} 2[E(\varepsilon_{t,i}^4)]^{1/2} N(0, 1).$$

In practical applications, even sampling at the highest available frequency, the number of intraday observations is finite and the price records are discrete. This introduces a bias due to discretization, such that

$$RV_t^{(TOT)} \stackrel{d}{\approx} IV_t + \underbrace{2n_t E(\varepsilon_{t,i}^2)}_{\text{bias due to noise}} + \underbrace{\left[\underbrace{4n_t E(\varepsilon_{t,i}^4)}_{\text{due to noise}} + \underbrace{\frac{2}{n_t} \int_0^1 \sigma_t^4 dt}_{\text{due to discretization}} \right]}_{\text{total variance}}^{1/2} N(0, 1),$$

where " $\stackrel{d}{\approx}$ " means that, when multiplied by a suitable factor, the convergence is in distribution. Recently, Zhang (2006a) and Aït-Sahalia et al. (2006), considered the case where the noise is not IID, such that Assumption 2 is modified as follows.

Assumption 3 (Dependent Noise Structure).

(a) The microstructure noise, $\varepsilon_{i,t}$, has a zero mean, stationary, and strong mixing

stochastic process, with the mixing coefficients decaying exponentially. In addition, $E[(\varepsilon_{t,i})^{4+\alpha}] < \infty$, for some $\alpha > 0$.

(b) The noise is independent of the price process.

(c) The variance of $\nu_{t,i} = \varepsilon_{t,i} - \varepsilon_{t,i-1}$ is $O(1)$.

Under Assumption 3, Zhang (2006a) and Aït-Sahalia et al. (2006) showed that

$$RV_t^{(TOT)} \stackrel{d}{\approx} IV_t + \underbrace{2n_t E(\varepsilon_{t,i}^2)}_{\text{bias due to noise}} + \underbrace{\left[\underbrace{2n_t \Omega}_{\text{due to noise}} + \underbrace{\frac{2}{n_t} \int_0^1 \sigma_t^4 dt}_{\text{due to discretization}} \right]}_{\text{total variance}}^{1/2} N(0, 1),$$

where

$$\Omega = V[(\varepsilon_{t,1} - \varepsilon_{t,0})^2] + 2 \sum_{i=1}^{\infty} Cov[(\varepsilon_{t,1} - \varepsilon_{t,1})^2, (\varepsilon_{t,i+1} - \varepsilon_{t,i})^2]$$

The most important fact about the last result is that, for large n_t , the realized variance (2.2.1) may have no connection to the true returns. On the contrary, $RV_t^{(TOT)}$ diverges to infinity linearly in n_t . In addition, Bandi and Russell (2005a) and Zhang et al. (2005) showed that, scaled by $(2n_t)^{-1}$, the realized variance estimates the variance of the microstructure noise consistently, such that

$$\frac{1}{2n_t} RV_t^{(TOT)} \xrightarrow{p} E(\varepsilon_{i,t}^2) \tag{2.2.13}$$

As recommended in Andersen et al. (2000a, 2001a, 2003), one possible solution to the microstructure bias is to sample the returns at arbitrarily selected lower frequencies, such as every 5 or 15 minutes, instead of at every tick. This procedure is called *sparse sampling*. However, Zhang et al. (2005) showed that this is not an adequate solution to the problem because most of the information contained in high frequency data is lost. First, define a new grid $\Lambda_t^{(sparse)}$, with $n_t^{(sparse)}$ sparsely

equidistant sampled observation times. Clearly, $\Lambda_t^{(sparse)}$ is a subgrid of Λ_t . Set

$$RV_t^{(sparse)} = \sum_{i=1}^{n_t^{(sparse)}} r_{t,i}^2. \quad (2.2.14)$$

where $r_{t,i}$ is the i -th return over the grid $\Lambda_t^{(sparse)}$.

Based on the results of Rootzen (1980), Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002), Mykland and Zhang (2006), Zhang et al. (2005), Zhang (2006a), and Aït-Sahalia et al. (2006) showed that the bias due to noise is given by $2n_t^{(sparse)}E(\varepsilon_{t,i}^2)$ and that, under Assumptions 2 or 3,

$$RV_t^{(sparse)} \stackrel{d}{\approx} IV_t + \underbrace{2n_t^{(sparse)}E(\varepsilon_{t,i}^2)}_{\text{bias due to noise}} + \underbrace{\left[\underbrace{2n_t^{(sparse)}E(\varepsilon_{t,i}^4)}_{\text{due to noise}} + \underbrace{\frac{2}{n_t^{(sparse)}} \int_0^1 \sigma_t^4 dt}_{\text{due to discretization}} \right]}_{\text{total variance}}^{1/2} N(0, 1),$$

Although the bias is reduced when $n_t^{(sparse)} < n_t$, the variance is increased due to discretization, leading to the well-known bias-variance trade-off.

Bandi and Russell (2005a,b) and Zhang et al. (2005) proposed a method for selecting the optimal sampling frequency based on the minimization of the MSE. When the Assumption 2 holds, defined the mean-square error as:

$$\begin{aligned} &= \mathbb{E}(RV_t^{(TOT)} - RV_t^{*(TOT)}) = MSE(n_t^{(sparse)}) \\ &= 2n_t^{(sparse)}E(\varepsilon_{t,i}^2) + 4n_t^{(sparse)}E(\varepsilon_{t,i}^4) + [8RV_t^{(sparse)}E(\varepsilon_{t,i}^2) - V(\varepsilon_{t,i}^2)] + \\ &\frac{2}{n_t}IQ_t^{(sparse)} \end{aligned}$$

Thus, the optimal sampling frequency may be approximated by

$$n_t^* \approx \left\{ \frac{IQ_t}{4[E(\varepsilon_{t,i}^2)]} \right\}^{1/3} \quad (2.2.15)$$

This optimal sampling are found minimizing the MSE as $\partial MSE / \partial n_t^{sparse} = 0$.

Bandi and Russell (2005a, 2006b) considered equidistant sampling intervals, whereas Zhang et al. (2005) provided a more general formula for irregularly spaced data. However, Bandi and Russell (2005a) also considered optimal sampling with depen-

dent noise, optimal sampling with bias-corrected realized variance estimates, and optimal sampling with prefiltered data.

As discussed previously, $E(\varepsilon_{t,i}^2)$ may be consistently estimated by $\frac{1}{2n_t}RV_t^{(TOT)}$; see Equation (2.2.13). Notwithstanding being a consistent estimation, an important point that must be emphasized is that the integrated quarticity is not known, and hence must be estimated. However, the realized quarticity, as given in Equation (2.2.10), is not consistent in the presence of microstructure noise. Bandi and Russell (2005a, 2006b) adopted the solution of computing (2.2.9) using a sparse set of observations, namely one that is sampled every 15 minutes. The authors showed through simulation that such sparse sampling did not seem to have a harmful effect on the selection of the optimal frequency. Zhang et al. (2005) developed an alternative solution for estimating the integrated quarticity. Even though choosing the sampling frequency on the basis of the finite sample mean-squared-error (MSE) is optimal in the case of realized variance, alternative estimators (discussed below) have been proposed that have the potential, when appropriately implemented, to outperform the classical realized variance estimator

2.2.3 Reviews of integrated volatility estimators

For having a consistent estimator, we need to minimize the bias present in the $RV_t^{(sparse)}$ estimator. We can do this using several ways such as; Subsampling, Kernel-Based Estimators, Filters and the method proposed by Large. For the first two methods we will be quite exhaustive while being very synthetic for the last two.

2.2.3.1 Subsampling

Zhang et al. (2005) proposed a subsampling method in order to estimate the integrated variance consistently in the presence of microstructure noise. The main idea is to explore the fact that, for example, ten-minute returns starting at 9:30 could be measured using the intervals 9:30–9:40, 9:40–9:50, 9:31–9:41, 9:41–9:51, and so on. Formally, suppose that the full grid, $\Lambda_t = \{k_0, \dots, k_t\}$, is partitioned into A nonoverlapping subgrids, $\Lambda_t^{(\alpha)}$ $\alpha = 1, \dots, A$ such that

$\Lambda_t = \bigcup_{\alpha=1}^A \Lambda_t^{(\alpha)}$, where $\Lambda_t^{(\alpha)} \cap \Lambda_t^{(\beta)} = \emptyset$ when $\alpha \neq \beta$

Set $n_t^{(\alpha)}$ as the number of observations in each subgrid, and define the RV for grid α as

$$RV_t^{(\alpha)} = \sum_{i=1}^{n_t^{(\alpha)}} r_{t,i}^2 \quad (2.2.16)$$

The proposal of Zhang et al. (2005) is to use the following estimator for the daily RV:

$$RV_t^{(TOT)} = \frac{1}{A} \sum_{\alpha=1}^{n_t^{(\alpha)}} RV_t^{(\alpha)} - \frac{\bar{n}_t}{n_t} RV_t^{(TOT)}, \quad (2.2.17)$$

where n_t is the number of observations in the full grid, and

$$\bar{n}_t = \frac{1}{A} \sum_{\alpha=1}^A n_t^{(\alpha)} = \frac{n_t - A + 1}{A}$$

The estimator in (2.2.17) is called the Two Time Scales Estimator (TTSE) of the integrated variance. Zhang et al. (2005) showed that, under Assumption 2,

$$n_t^{-1/6} [RV_t^{(TOT)} - IV_t] \xrightarrow{d} + \underbrace{\left[\underbrace{8c^{-2} E[(\varepsilon_{t,i}^2)]^2}_{\text{due to noise}} + \underbrace{\frac{2}{3} IQ_t}_{\text{due to discretization}} \right]^{1/2}}_{\text{total variance}} N(0, 1),$$

where, in the case of equidistant observations,

$$c = \left\{ \frac{1}{12E[(\varepsilon_{t,i}^2)]^2} IQ_t \right\}^{-1/3}$$

In Ait-Sahalia et al. (2006), a small sample refinement to the estimator in (2.2.17) is proposed. The final estimator becomes

$$RV_t^{(ZMA,adj)} = \left(1 - \frac{\bar{n}_t}{n_t} \right) RV_t^{(ZMA)}, \quad (2.2.18)$$

Both of the estimators in (2.2.17) and (2.2.18) are derived under Assumption 2 (IID noise). In order to take into account possibly dependent noise, Zhang (2006a) and Aït-Sahalia et al. (2006) proposed an alternative estimator that is also based on the two time scales idea. All the results are derived under Assumption 3 (non-IID noise).

First, the authors defined the *average lag J realised volatility* $RV_{t,J}^{(AL)}$, which is given by

$$RV_{t,J}^{(AL)} = \frac{1}{J} \sum_{i=0}^{n_t-J} (r_{t,i+J} - r_{t,i})^2 \quad (2.2.19)$$

Then the authors proposed a generalization of the TTSE derived in Zhang et al. (2005), which has the form

$$RV_t^{(ZMA)} = RV_t^{(AL)} - \frac{\bar{n}_t^{(A)}}{\bar{n}_t^J} RV_{t,J}^{(AL)} \quad (2.2.20)$$

where $1 \leq J \leq A \leq n_t$, $A = o(n_t)$, as $n \rightarrow \infty$ $\bar{n}_t^{(A)} = \frac{(n_t-A+1)}{A}$ and $\bar{n}_t^{(J)} = \frac{(n_t-J+1)}{J}$. Note that (2.2.19) becomes the TTSE in Zhang et al. (2005) when $J = 1$ and $A \rightarrow \infty$ as $n_t \rightarrow \infty$. A small sample correction is given by

$$RV_t^{(ZMA,adj)} = \left(1 - \frac{\bar{n}_t^{(A)}}{\bar{n}_t^{(J)}}\right) RV_t^{(ZMA)}, \quad (2.2.21)$$

Zhang (2006a) and Aït-Sahalia et al. (2006) showed that

$$[RV_t^{(ZMA,adj)} \stackrel{d}{\approx} IV_t + \underbrace{\left[\underbrace{c^{-2}\zeta^2}_{\text{due to noise}} + \underbrace{\frac{c}{3} \int_0^t \sigma_t^4 dt}_{\text{due to discretization}} \right]}_{\text{total variance}}]^{1/2} N(0, 1),$$

where c is a constant and

$$\zeta^2 = 16V(\varepsilon_{t,i})^2 + 32 \sum_{i=1}^{\infty} Cov(\varepsilon_{t,0}, \varepsilon_{t,i})^2$$

2.2.4 Kernel-based estimators

Consistently estimating the quadratic variation under the presence of microstructure noise is, in a sense, similar to the well-known autocorrelation corrections that are frequently used in the time-series literature to estimate the long run variances and covariances of stationary stochastic processes (see, for example, Andrews, 1991; Newey and West, 1987). Consequently, it is natural to adapt similar techniques for the present case. For example, Hansen and Lunde (2004, 2006a) considered the following simple kernel-based estimator

$$RV_t^{(HL)} = RV_t^{(TOT)} + 2 \sum_{h=1}^H \frac{n_t}{n_t - h} \hat{\gamma}_h \quad (2.2.22)$$

where

$$\hat{\gamma}_h = \frac{n_t}{n_t - h} \sum_{j=1}^{n_t-h} r_{t,j} r_{j+h} \quad (2.2.23)$$

Zhou (1996) was the first to consider the use of kernel methods to deal with the problem of microstructure noise in high-frequency data. For the case of independent noise, Zhou proposed (2.2.22) with $H = 1$. Hansen and Lunde (2006b) examined the properties of Zhou’s estimator and showed that, although unbiased under Assumption 2, the estimator is not consistent. However, Hansen and Lunde (2006b) advocated that, while inconsistent, Zhou’s kernel method is able to uncover several properties of the microstructure noise, and concluded that the noise:

- (1) Is correlated with the efficient price;
- (2) Is time dependent;
- (3) Is quite small in the DJIA stocks; and

(4) Has properties that change substantially over time. Their results are robust to both CTS and TrTS. Moreover, selecting higher values for H does not solve the consistency problem. However, the estimator in (2.2.22) is unbiased by an upwards scaling of the empirical autocovariances. The h -th autocovariance is scaled by $\frac{n_t}{n_t-h}$ to compensate for the “missing” autocovariance terms.

The upward scaling has the drawback that it increases the variance of the estimator. For this reason, Hansen and Lunde (2005b) consider the Bartlett kernel and define the estimator

$$RV_t^{(HL, Bartlett)} = RV_t^{(TOT)} + 2 \sum_{h=1}^H \left(1 - \frac{h}{H+1}\right) \hat{\gamma}_h \quad (2.2.24)$$

where H is determined as $H = \left\lceil \left(\frac{4n_t}{100}\right)^{2/9} \right\rceil$ and $\hat{\gamma}_h$ is defined as in (2.1.23). However, the estimator (2.1.24) is also inconsistent.

Recently, Barndorff-Nielsen et al. (2006a) proposed the flat-top kernel-based estimator

$$RV_t^{(BHLS)} = RV_t^{(TOT)} + \sum_{h=1}^H k\left(\frac{h-1}{H}\right) (\hat{\gamma}_h + \hat{\gamma}_{-h}) \quad (2.2.25)$$

where $k(x)$ for $x \in [0, 1]$ is a nonstochastic weight function such that $k(0) = 1$ and $k(1) = 0$. The authors made several contributions to the literature by:

- (1) proving that the statement that all kernel based RV estimators were inconsistent is wrong and proposed several consistent kernel-based estimators;
- (2) designing a kernel that has a smaller variance than the multiscale estimator;
- (3) proposing an estimator for data with endogenously spaced observations, such as that in databases on transactions;
- (4) considering the case where the microstructure noise is endogenous.

Barndorff-Nielsen et al. (2006a) showed that, if $H = cn_t^{2/3}$, then the resulting estimator is asymptotically mixed Gaussian, converging at rate $H = n_t^{1/6}$. The constant, c , can be optimally chosen as a function of the kernel $k(x)$. For example, the value of c that minimizes the variance of the estimator is given by

$$c = \left\{ \frac{2[k'(0)^2 + k'(1)^2]}{\int_0^1 k(x)^2 dx} \right\}^{1/3} \frac{[E(\varepsilon_{t,i}^2)]^{2/3}}{IQ_t^3}$$

Barndorff-Nielsen et al. (2006a) also compared three different kernels:

- (1) Bartlett where $k(x) = 1 - x$;
- (2) the 2nd order where $k(x) = 1 - 2x - x^2$;

(3) Epanechnikov where $k(x) = 1 - x^2$.

Their findings are summarized as follows: the Bartlett kernel has the same asymptotic distribution as the TTSE of Zhang et al. (2005) and is more efficient than the Epanechnikov alternative, but is less efficient than the 2nd order kernel. Moreover, if $k'(0) = 0$ and $k'(1) = 0$, then setting $H = cn_t^{1/2}$, the asymptotic distribution of the estimator is mixed normal with convergence rate equal to $n_t^{1/4}$. Under microstructure effects, Barndorff-Nielsen et al. (2006a) discussed the choice of the constant c in a simplified framework where the variance of the efficient price is held constant. In their article, the authors compared eight different kernels satisfying $k'(0) = 0$ and $k'(1) = 0$. The cubic kernel, where $k(x) = 1 - 3x^2 + 2x^3$, has the same asymptotic distribution as the multiscale estimator of Aït-Sahalia et al. (2006) and Zhang (2006a). The Tukey–Hanning kernel, where $k(x) = \frac{[1 - \cos\pi(1-x)^2]}{2}$, seems to be the best option in terms of efficiency.

Barndorff-Nielsen et al. (2006a) also showed that the findings above are robust to endogenous and/or dependent noise, and endogenously spaced observations, as in tick data databases. They also provided Monte Carlo evidence in favor of their estimators in finite samples.

2.2.5 Filters

In the early days of modelling RV, another common alternative to attenuate the effects of the microstructure noise was to pre-filter the intraday returns. For example, in Bollen and Inder (2002), an autoregressive (AR) filter was used, while a moving average (MA) filter was considered in Ebens (1999), Maheu and McCurdy (2002), and Andersen et al. (2001a). More recently, Hansen et al. (2008), showed that the MA(1) structure considered in Ebens (1999) and Andersen et al. (2001a) is well specified when the market microstructure noise is IID. Moreover, when correcting the estimator by a scaling factor, it becomes a consistent estimator of the integrated variance (see Hansen et al., 2008 for further details).

Recently, Large (2006) proposed an interesting estimator of quadratic variation which controls for microstructure effects when the best quotes change by jumping

the minimum price tick. The estimator compares the number of alternations, where quotes jump back to their previous price, with the number of other jumps. If the alternations are uncorrelated, the estimator is consistent in a limit theory where jumps are very frequent and small.

2.3 Reviews of spot volatility estimators

In this section we analyze the technical aspects of implementation of several spot volatility estimators. First of all, we define the notion of spot volatility. From the formula (2.2.3), with the assumption that $\{\sigma_t\}$ has continuous sample path, the spot volatility can be obtained by taking the derivative of the integrated volatility with respect to time path wisely

$$\sigma_t^2 = \lim_{h \rightarrow 0} \frac{[p, p]_{t+h} - [p, p]_t}{h} = \lim_{h \rightarrow 0} \frac{1}{h} \int_t^{t+h} \sigma_s^2 ds \tag{2.3.1}$$

Our interest in this part is to estimate σ_t for some fixed point using the discretely observed high frequency data over a defined time interval.

Always defining $p(t)$ as in (2.2.3), we assume that,

- a) $E[\varepsilon_{i,t}] = 0$, $\sup_t E[\varepsilon_{i,t}^2] = C_\varepsilon^2 < \infty$ (C_ε = unknown positive constant)
- b) For any arbitrary point set $\{t_k \in [0, T], 1 \leq k \leq n\}$ of arbitrary size n , the random variable $\{\varepsilon_{i,t}\}$ are independent.

let us briefly recall the theory of all spot volatility estimators employed in the comparison. Given a discrete realization of the process X_t , namely given $N' + 1$ observations $X(t_0), X(t_1), \dots, X(N)$ in the interval $[0, T]$, we want to estimate the spot volatility $\sigma^2(t'_k)$ at time t'_k . We denoted by $\{t_k, 0 \leq k \leq N\}$ the finite set of points in $[0, T]$ at which the noisy price process is observed. For simplicity in calculus and notations, we suppose that the points are equally spaced and we set $t_k = k\Delta$, $\Delta = T/N$. And we set

$$t'_k = t_{kD}, 1 \leq k \leq N'$$

where N', D are natural numbers satisfying the relation $N'D = N$. Where for each observation at time t'_k , we divide this interval into D steps, and $N + 1$ represents the amount of daily data (without infill).

The first model we are going to deal with is the normalized squared returns. This estimator is obtained by setting

$$\tilde{\sigma}_i^2(t'_k) = \frac{(X(t'_k) - X(t'_{k-1}))^2}{\Delta'} \quad \text{where } \Delta' = \frac{T}{N}$$

Unfortunately, this estimator is very noisy, because it uses low frequency data.

A more robust estimator of the Realized Volatility type has been analyzed by several authors, among others Comte and Renault (1998). It is given by

$$\tilde{\sigma}_{CR}^2(t'_k) = \frac{1}{m} \sum_{j=1}^m \frac{(X(t'_{k-m/2+j}) - X(t'_{k-m/2+j-1}))^2}{\Delta'} \quad \text{where } \Delta' = \frac{T}{N} \quad (2.3.2)$$

Comte and Renault (1998), Mykland and Zhang (2008), for instance study its asymptotic properties. That estimator is an average of m squared returns around t'_k . The larger m , the smoother the estimates. In the presence of microstructure effects due to tick-by-tick quotes, data must be sampled at a lower frequency in order to filter out the high frequency noise component.

A general class of nonparametric spot volatility filters is based on a kernel weighted measure of the integrated volatility, see for example D.Kristensen (2006), Fan and Jiang (2008), Mattiussi and Renò (2010) and can be seen as a continuous-time weighted moving average of the instantaneous volatility. Given a kernel $K : \mathbb{R} \rightarrow \mathbb{R}$ normalized to $\int_{\mathbb{R}} K(z) dz = 1$ and a bandwidth $h > 1$, define $K_h(z) = \frac{1}{h} K(\frac{z}{h})$. Then, based on standard results for kernel estimators, a natural spot volatility estimator is

$$\hat{\sigma}^2(t) = \lim_{h \rightarrow 0} \sum_{i=0}^N K_h(t_{i-1} - t) [X(t_i) - X(t_{i-1})]^2.$$

This is Nadaraya-Watson type kernel estimator and it is simply the limit for shrink-

ing bandwidth sequences of a kernel weighted average of the squared increments of data or, in other words, a kernel regression estimator in the time domain. By an appropriate choice of the kernel, it can potentially deal with the presence of market microstructure effects. Consistency and both pointwise and global asymptotic distributions are established in Fan and Wang (2008) and D.Kristensen (2006). This broad class of estimators includes as a special case the standard realized volatility type estimators and the rolling window estimators proposed by Andreou and Ghysels (2002) and Foster and Nelson (1996). In Andreou and Ghysels (2002) the k-day spot volatility estimator is defined as a one-sided moving k-day average of one-day spot volatility given by a rescaled sum of squared intraday returns. This estimator looks very close to the real-time scheme which will be resumed in the next paragraph. In particular, in our analysis we consider the following schemes from the class of kernel estimators: the *one-sided exponential filter*

$$\tilde{\sigma}_E^2(t'_k) = (1 - \lambda) \sum_{j=1}^i \lambda^j [X(t'_{k-j+1}) - X(t'_{k-j})]^2 \quad (2.3.3)$$

where λ is a smoothing parameter, which can be set equal to 0.94 for daily data according to J.P. Morgan standard: the *one-sided rolling daily window* volatility, defined as

$$\tilde{\sigma}_W^2(t'_k) = \sum_{j=1}^{n_L} \omega_j [X(t'_{k-j+1}) - X(t'_{k-j})]^2 \quad (2.3.4)$$

where n_L is the lag length of the rolling window and $w_j = \exp(-\alpha j)$. The parameter α can be optimized to minimize the asymptotic measurement error variance as explained in Foster and Nelson (1996). These two volatility measures are then normalized by the sampling interval in order to obtain a measure of volatility per unit time.

The so called real-time estimator of the spot volatility reconstructs a volatility which can work in a real-time manner, namely an estimator which can estimate the spot volatility at each time almost immediately on receiving the observed data around that time. Given the observed log-price process $X(t)$, a first regularization is

performed to obtain a smoothed process $\bar{X}(t)$ which is then used to recover the spot volatility. The scheme proposed by Ogawa and Sanfelici (2010) is the following:

As above, we denoted by $\{t_k, 0 \leq k \leq N\}$ the finite set of points in $[0, T]$ at which the noisy price process is observed. For simplicity in calculus and notations we suppose that the points are equally spaced and we set $t_k = k\Delta$, $\Delta = T/N$. Let $\{t'_k, 1 \leq k \leq N'\} \subset \{t_k, 1 \leq k \leq N\}$, be the subset of estimation points such that

$$t'_k = t_{kD}, \quad 1 \leq k \leq N'$$

where N', D are natural numbers satisfying the relation $N'D = N$. The remaining data $X(t_i)$ for $t_i \notin \{t'_k, 1 \leq k \leq N'\}$ will be used to construct the regularized data $\bar{X}(t)$, which are defined in the following way

$$\bar{X}(t'_k) = \frac{1}{M} \sum_{i=1}^M X(t_{kD-i+1}), \quad 1 \leq k \leq N' \quad (2.3.5)$$

where M is an integer not necessarily equal to D . The regularization of other processes $p(t)$, ε_i , i at $t = t'_k = t_{kD}$ is defined in the same way. Thus we have again the relation

$$\bar{X}(t'_k) = \bar{p}(t'_k) + \bar{\varepsilon}(t'_k).$$

It can be showed that the family $\{\bar{\varepsilon}(t'_k)\}$ is almost orthogonal in the sense that

$$E[\bar{\varepsilon}(t'_k)\bar{\varepsilon}(t'_l)] = 0 \text{ for } |k - l| > \rho := \frac{M}{D}$$

Now taking the increment over $[t'_{(k-1)}, t'_k]$ of each member in the relation above, we have

$$\Delta'_k \bar{X} = \Delta'_k \bar{p} + \Delta'_k \bar{\varepsilon}$$

where $\Delta'_k \bar{Y} = \bar{Y}(t'_k) - \bar{Y}(t'_{(k-1)}) = \bar{Y}(t_{kD}) - \bar{Y}(t_{(k-1)D})$, ($Y = X, p, \varepsilon$).

Given all that assumptions, the estimator is defined in the following way

For each t'_k ($1 \leq k \leq N'$) and set

$$\tilde{\sigma}_{OS}^2(t'_k) = \frac{G(\rho)^{-1}}{L} \sum_{i=1}^L \frac{(\Delta'_{k+l} \bar{X})^2}{\Delta'}, \quad (2.3.6)$$

where, $\Delta' = T/N' = D\Delta$ and $G(\rho) = \frac{3MD - D^2 + 1}{3M^2} = \frac{(3\rho - 1)D^2 + 1}{3\rho^2 D^2}$. And the normalizing constant is $G(\rho) = \frac{1}{DM^2} \sum_{i,j=1}^D \{M - |i - j|\}$,

In Ogawa and Sanfelici (2010), the proof of an explicit L^1 bound for the estimate over which, suitable choice of M and L yields the consistency of the estimator (2.3.3) is proposed.

Consider, in a filtered probability space, the process (2.2.3), where it is assumed that $E[\int_0^{2\pi} (\mu(t))^2] < \infty$, $E[\int_0^{2\pi} (\sigma(t))^4] < \infty$ so that a single solution of the stochastic differential equation (2.2.3) exists in the interval $[0, 2\pi]$, and W_t a standard Brownian motion. Define the Fourier coefficients of dr and σ^2 as follows:

$$\begin{aligned} a_0(dr) &= \frac{1}{2\pi} \int_0^{2\pi} dr, & a_0(\sigma^2) &= \frac{1}{2\pi} \int_0^{2\pi} \sigma^2(t) dt \\ a_k(dr) &= \frac{1}{\pi} \int_0^{2\pi} \cos(kt) dr, & a_k(\sigma^2) &= \frac{1}{\pi} \int_0^{2\pi} \cos(kt) \sigma^2(t) dt \\ a_k(dr) &= \frac{1}{\pi} \int_0^{2\pi} \sin(kt) dr, & a_k(\sigma^2) &= \frac{1}{\pi} \int_0^{2\pi} \sin(kt) \sigma^2(t) dt \end{aligned}$$

in Malliavin and Mancino (2002) it is shown that, given an integer $n_0 > 0$, we have

$$a_0(\sigma^2) = \lim_{N \rightarrow \infty} \frac{\pi}{N + 1 + n_0} \sum_{k=n_0}^N \frac{1}{2} (a_k^2(dr) + b_k^2(dr)) \quad (2.3.7)$$

$$a_q(\sigma^2) = \lim_{N \rightarrow \infty} \frac{\pi}{N + 1 + n_0} \sum_{k=n_0}^N \frac{1}{2} (a_k(dr) a_{k+q}(dr) + b_k(dr) b_{k+q}(dr)) \quad (2.3.8)$$

$$b_q(\sigma^2) = \lim_{N \rightarrow \infty} \frac{\pi}{N + 1 + n_0} \sum_{k=n_0}^N \frac{1}{2} (a_k(dr)a_{k+q}(dr) + b_k(dr)b_{k+q}(dr)) \quad (2.3.9)$$

almost surely. Then, the spot volatility can be estimated using its Fourier coefficient as

$$\tilde{\sigma}(t) = \lim_{M \rightarrow \infty} \sum_{k=0}^M [a_k(\sigma^2)\cos(kt) + b_k(\sigma^2)\sin(kt)] \quad (2.3.10)$$

When actually implementing the estimator, it is suggest to add a linear trend on the observed process X_t such that we get $X(2\pi) = X(0)$, which does not have an influence on the volatility estimate. Moeover, we set $n_0 = 1$ and we filter progressively high modes by using the following approximation of the volatility

$$\tilde{\sigma}_F(t) = \lim_{M \rightarrow \infty} \sum_{k=0}^M \varphi_F(\delta k) [a_k(\sigma^2)\cos(kt) + b_k(\sigma^2)\sin(kt)] \quad (2.3.11)$$

where $\varphi(x) = \sin^2(x)/x^2$ is the Fejer kernel and $\varphi(0) = 1$. δ is an a parameter adapted to the scale which is expected to give an appropriate resolution of the volatility and to filter out high frequency noise modes. When choosing δ , it is convenient to have $\delta \geq \frac{1}{M}$, the larger δ , the smoother and less detailed the estimated volatility. After selecting a suitable value of the cutting frequency N_F , it is convenient to use the maximum M that can be computed, namely $M = N_F - n_0$, and successively tune δ to filter high frequencies. Finally, we remark that according to well-known diffraction phenomena of Fourier series near the boundary of the time window this estimator performs badly at the boundaries.

Chapter 3

Nonparametric estimation of the diffusion coefficient of local volatility models

In this chapter we review the concepts of kernel estimator and *local time*. We introduce the general framework for deriving nonparametric estimation of the diffusion coefficient. Using spot volatility (2.3.2, 2.3.3, 2.3.4 and 2.3.6), we construct nonparametric studied estimators and prove that they are consistent and asymptotically normally distributed.

3.1 Kernel density estimator

3.1.1 Definition and construction

We suppose that X_1, X_2, \dots, X_n are independent identically distributed and drawn from a (unknown) density function $f(x)$, of which we want to have an estimate.

We remember that the probability density f , of a random variable X is the

derivative of the cumulative distribution function F (if it exists). Then we can write

$$\begin{aligned} f(x) &= \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} \\ &= \lim_{h \rightarrow 0} \frac{P\{x-h < X \leq x+h\}}{2h} \end{aligned}$$

an estimator of $f(x)$ is then

$$\begin{aligned} \hat{f}(x) &= \frac{1}{2h} \frac{\#\{i : x-h < X_i \leq x+h\}}{n} \\ &= \frac{1}{2hn} \sum_{i=1}^n I_{\{x-h < X_i \leq x+h\}} \\ &= \frac{1}{2hn} \sum_{i=1}^n I_{\{-1 < \frac{x-X_i}{h} \leq 1\}} \end{aligned}$$

where I_A is the indicator function, taking the value 1 if A is true and 0 elsewhere. The fraction $\frac{1}{2nh}$ in front of the sum normalizes $\hat{f}(x)$ making its integral equal to one. We can prove that, if $h \rightarrow 0$, fast enough, $\hat{f} \rightarrow f$ as $n \rightarrow \infty$.

We can express the estimator more transparently, defining the weight function w by

$$w(x) = \begin{cases} \frac{1}{2} & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.1.1)$$

then it is easy to see that the naive estimator can be written as

$$\hat{f}(x) = \frac{1}{2n} \sum_{i=1}^n \frac{1}{h} w\left(\frac{x-X_i}{h}\right)$$

The estimator constructed above presents many flaws (has jumps, non derivability at all points etc...), We can generalize the naive estimator to resolve some difficulties stated above.

We can replace the weight function w by a general *kernel function* K which

satisfies the condition

$$\int_{-\infty}^{\infty} K(x)dx = 1 \quad (3.1.2)$$

Usually, but not always, we can suppose that K is a symmetric probability density function, the normal density for instance or the weight function w used in the construction of the first estimator. Similar to the definition of the naive estimator, the *kernel estimator* with kernel K is defined by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (3.1.3)$$

where h is the *window width*, also called the *smoothing parameter* or *bandwidth* by some authors.

In each observation X_i , we place a “bump”, the kernel estimator resulting is merely the sum of the “bumps”.

The kernel K determines the form of the “bumps”, and the window h determines the width of the “bumps” as precised above.

We can easily see that the kernel estimator

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

has the following properties

- If K is a probability density, the estimator \hat{f} is a probability density as well.
- \hat{f} has the same properties of continuity and differentiability as K .

1. If K is continue, \hat{f} will be also a continuous function
2. If K is differentiable, \hat{f} will be also a differentiable function
3. If K can take negative values, then \hat{f} should take negative values as well

3.1.2 Measure of discrepancy: mean square error and mean integrated square error

Considering the kernel estimator

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K \left(\frac{x - X_i}{h} \right) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i)$$

where we introduce the notation

$$K_h(\cdot) = \frac{1}{h} K \left(\frac{\cdot}{h} \right)$$

for a transformed version of K .

To calculate the bias of the kernel estimator, we first observe that $\mathbb{E}\{\hat{f}(x)\} = \mathbb{E}\{K_h(x - X)\} = \int K_h(x - y)f(y)dy$.

We recall that the convolution between two functions is defined as

$$(f * g)(x) = \int_{\mathbb{R}} f(x - y)g(y)dy.$$

Then, we have

$$\mathbb{E}\{\hat{f}(x)\} - f(x) = \underbrace{(K_h * f)(x)}_{\text{smooth version of } f} - f(x)$$

we calculate the variance as follow

$$Var\{\hat{f}(x)\} = \mathbb{E}\{\hat{f}^2(x)\} - [\mathbb{E}\{\hat{f}(x)\}]^2 = \frac{1}{n} \{(K_h^2 * f)(x) - (K_h * f)^2(x)\}$$

When considering estimation at single point, a natural measure of the discrepancy is the *mean square error* (MSE), defined by

$$MSE_x(\hat{f}) = \mathbb{E}\{\hat{f}(x) - f(x)\}^2 \tag{3.1.4}$$

by standard properties of mean and variance, we can express it as

$$MSE_x(\hat{f}) = \{\mathbb{E}\hat{f}(x) - f(x)\}^2 + Var\hat{f}(x) = Var\hat{f}(x) + [Bias(\hat{f}(x))]^2 \quad (3.1.5)$$

The most general way of placing a measure on the global accuracy of \hat{f} as an estimator of f is the *mean integrated square error* (MISE) defined as

$$MISE(\hat{f}) = \int MSE(\hat{f}). \quad (3.1.6)$$

since

$$\int (K_h^2 * f)(x) dx = \int \frac{1}{h^2} \left\{ \int K^2 \left(\frac{x-y}{h} \right) f(y) dy \right\} dx = \frac{1}{h} \int \int K^2(u) \{f(x-uh) dx\} du, \quad \text{with } u = \frac{x-y}{h}$$

By substitution in (3.1.5) we find that

$$MISE(\hat{f}) = \frac{1}{nh} \int K^2(u) du + \left(1 - \frac{1}{n}\right) \int (K_h * f)^2(x) dx - 2 \int (K_h * f)(x) f(x) dx + \int f^2(x) dx.$$

3.1.3 Approximate properties of bias and variance

We suppose that the kernel K is a symmetric function satisfying the following properties

$$K \geq 0 \quad \int K(t) dt = 1, \quad \int tK(t) dt = 0, \quad \text{and } 0 < \int t^2 K(t) dt < \infty \quad (3.1.7)$$

and that the unknown density f has continuous derivatives of all orders required.

An asymptotic approximation of the expectation of the estimator $\hat{f}(x)$ is given by

$$\begin{aligned} \mathbb{E}\{\hat{f}(x)\} &= \int K_h(x-y) f(y) dy \\ &= \int K(u) f(x-uh) dy, \quad \text{with } u = \frac{x-y}{h} \quad du = -\frac{1}{h} dy \\ &= \int K(u) [f(x) - f'(x)uh + \frac{1}{2} f''(x)u^2 h^2 + \dots] du \quad \text{by} \end{aligned}$$

by Taylor expansion

$$= f(x) \int K(u)du - f'(x)h \int K(u)udu + \frac{1}{2}f''(x)h^2 \int K(u)u^2du + o(h^2)$$

the bias can be obtained as

$$\mathbb{E}\{\widehat{f}(x)\} - f(x) = \frac{1}{2}f''h^2 \int K(u)u^2du + o(h^2) \quad (3.1.8)$$

Given that

$$\text{Var}\{\widehat{f}(x)\} = \frac{1}{n}\{\mathbb{E}K_h^2(x - X) - [\mathbb{E}K_h(x - X)]^2\}$$

$$\begin{aligned} \mathbb{E}K_h^2(x - X) &= \frac{1}{h^2} \int K^2\left(\frac{x-y}{h}\right) f(y)dy \\ &= \frac{1}{h} \int K^2(u)f(x - uh)du, \quad \text{with } u = \frac{x-y}{h} \\ &= \frac{1}{h} \int K^2(u)[f(x) - f'(x)hu + \dots]du, \quad \text{by Taylor expansion} \end{aligned}$$

we finally find that $\text{Var}\{\widehat{f}(x)\} = \frac{1}{nh}f(x) \int K^2(u)du + o(\frac{1}{nh})$

To sum up, we have therefore established that

$$\text{Bias}\{\widehat{f}(x)\} = \frac{1}{2}f''(x)\beta h^2 + o(h^2) \quad \text{where } \beta = \int K(u)u^2du$$

$$\text{Var}\{\widehat{f}(x)\} = \frac{1}{nh}f(x)R(K) + o(\frac{1}{nh}) \quad \text{where } R(K) = \int K^2(u)du \quad (3.1.9)$$

If $h = h_n \rightarrow 0$ as $n \rightarrow \infty$, then $\text{Bias}\{\widehat{f}(x)\} \rightarrow 0$ as $n \rightarrow \infty$

If $h = h_n \rightarrow 0$ as $n \rightarrow \infty$, then $\text{Var}\{\widehat{f}(x)\} \rightarrow 0$ as $n \rightarrow \infty$

We see that as in many branches of statistics, there is a trade-off between the bias and the variance terms; the bias can be reduced at the expense of increasing the variance, and vice versa by adjusting the amount of smoothing.

3.1.4 The ideal window width and kernel

From (3.2.8) we have the following asymptotic approximation for MSE and MISE

$$\begin{aligned} MSE\{\widehat{f}_n(x)\} &= \frac{1}{4}f''(x)\beta^2h^4 + \frac{1}{nh}f(x)R(K) + o(h^4 + \frac{1}{nh}) \\ MISE\{\widehat{f}_n(x)\} &= \frac{1}{4}\beta^2h^4 \int \{f''(x)\}^2 dx + \frac{1}{nh}R(K) + o(h^4 + \frac{1}{nh}) \end{aligned}$$

under appropriate integrability conditions of f and his derivative

We denote the asymptotic approximation of the MSE by

$$AMSE\{\widehat{f}_n(x)\} = \frac{1}{4}f''(x)\beta^2h^4 + \frac{1}{nh}f(x)R(K) \quad (3.1.10)$$

and the asymptotic approximation of the MISE by

$$AMISE\{\widehat{f}_n(x)\} = \frac{1}{4}\beta^2h^4 \int \{f''(x)\}^2 dx + \frac{1}{nh}R(K) \quad (3.1.11)$$

The ideal value of h is obtained by minimizing the approximate mean integrated square error, and finally we obtain

$$h_{opt} = \beta^{-2/5} \left\{ \int K(t)dt \right\}^{1/5} \left\{ \int f''(x)^2 dx \right\}^{-1/5} n^{-1/5} \quad (3.1.12)$$

Substituting the value of h_{opt} from (3.1.11) inside the value of the approximate formula (3.1.10) for the mean integrated square error shows that, if h is chosen optimally, then the approximate value of the mean integrated square error will be

$$\frac{5}{4}C(K) \left\{ \int f''(x)^2 dx \right\}^{1/5} n^{-4/5} \quad (3.1.13)$$

where the constant $C(K)$ is given by

$$C(K) = \beta^{2/5} \left\{ \int K(t)^2 dt \right\}^{4/5} \quad (3.1.14)$$

The constant $C(K)$ depends on the choice of the kernel and on the second derivative of the target density function. $R(f'') = \int f''(x)^2 dx$ denotes the curvature. Thus,

densities with small curvature will ask for a large bandwidth (more smoothing), conversely, densities with large curvature will ask for a small bandwidth. The criterion (3.1.11) is infeasible. However, several techniques exist to estimate $R(f'')$. The simplest is based on the assumption of a gaussian kernel K and that f is normal density with variance σ^2 .

$$\int f''(x)^2 dx = \int \phi''(x)^2 dx = \frac{3}{8} \pi^{-1/2} \sigma^{-5} \quad (3.1.15)$$

where σ is the standard deviation of a family of variables containing f and ϕ is a standard normal density. Then the window width obtained by substitution gives the value

$$h_{opt} = 1,06 \sigma n^{-1/5} \quad (3.1.16)$$

A quick way of estimating the smoothing parameter, therefore would be to estimate σ from the data and then to substitute the estimate into (3.2.15). Either the usual sample standard deviation or a more robust estimator of σ could be used.

3.2 Local time

In what follows we introduce some preliminary results regarding the *local* or *sojourn time* of a continuous semimartingale (SMG). These results will be useful in the development of our analysis (Revuz and Yor (1998) is a standard reference). The concept of continuous semimartingale was presented thoroughly in the first chapter. In this section we will deal with the stochastic differential equation of the form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \quad (3.2.1)$$

with the initial condition $X_0 = \bar{X}$, where $\{W_t : t \geq 0\}$ is a standard Brownian motion defined on the filtered probability space $(\Omega, \mathfrak{F}^W, (\mathcal{F}_t^W)_{t \in \mathbb{R}^+}, P)$

Stochastic differential equations like (3.2.1) are known to have solutions that are SMGs since $\bar{X} + \int_0^t \mu(X_s)ds$ is a continuous adapted process of finite variation and

$\int_0^t \sigma(X_s)dB_s$ is a continuous local martingale. Hence, our theory comes within the ambit of SMG analysis.

The local time of a continuous SMG M_t is defined as follows:

Lemma 38. *(The Tanaka Formula) For any real number a , there exists a non-decreasing continuous process $L_M(\cdot, a)$ called the local time of M_t at a , such that*

$$|M_t - a| = |M_0 - a| + \int_0^t \text{sgn}(M_s - a)dM_s + L_M(t, a), \quad (3.2.2)$$

$$(M_t - a)^+ = (M_0 - a)^+ + \int_0^t 1\{M_s > a\}dM_s + \frac{1}{2}L_M(t, a), \quad (3.2.3)$$

$$(M_t - a)^- = (M_0 - a)^- + \int_0^t 1\{M_s \leq a\}dM_s + \frac{1}{2}L_M(t, a), \quad (3.2.4)$$

In particular, $|M_t - a|$, $(M_t - a)^-$, $(M_t - a)^+$ are SMGs.

Lemma 39. *(Continuity of SMG Local Time) For any continuous SMG M_t , there exists a version of the local time such that the map $(t, a) \mapsto L_M(t, a)$ is a.s. continuous in t and càdlàg in a .*

Lemma 40. *(The Occupation Time Formula) Let M_t be a continuous SMG with quadratic variation process $[M_t]$ and let L^a be the local time at a . Then,*

$$\int_0^t f(M_s, s)d[M]_s = \int_{-\infty}^{\infty} da \int_0^t f(a, s)dL_M(s, a) \quad (3.2.5)$$

for every positive Borel measurable function f . If f is homogeneous, then the expression simplifies to

$$\int_0^t f(M_s, s)d[M]_s = \int_{-\infty}^{\infty} f(a)L_M(t, a)da \quad (3.2.6)$$

Lemma 41. *If M_t is a continuous SMG then, almost surely*

$$L_M(t, a) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_0^t 1_{[a, a+\varepsilon]}(M_s) d[M]_s \quad (3.2.7)$$

Lemma 42. *If M_t is a continuous local martingale then, almost surely*

$$L_M(t, a) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^t 1_{]a-\varepsilon, a+\varepsilon]}(M_s) d[M]_s \quad (3.2.8)$$

The process $L_M(t, a)$ is called the local time of M_t at the point a over the time interval $[0, t]$.

It is measured in units of the quadratic variation process and gives the amount of time that the process spends in the vicinity of a . The “chronological local time” (terminology from Phillips and Park (1998)) is a standardized version of the conventional local time that is defined in terms of pure time units. It can be easily derived in the Brownian motion case. From (3.2.8), the local time of a standard Brownian motion W_t is

$$L_W(t, a) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^t 1_{(|W_s - a| < \varepsilon)} ds \quad a.s. \quad \forall a, t. \quad (3.2.9)$$

Now, consider the Brownian motion $B_t = \sigma W_t$ with local variance σ^2 . We can write, as in Phillips and Park (1998),

$$L_B(t, a) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^t 1_{(|B_s - a| < \varepsilon)} \sigma^2 ds = \sigma L_W(t, \frac{a}{\sigma}) \quad a.s. \quad \forall a, t. \quad (3.2.10)$$

Since the quadratic variation of Brownian motion is deterministic, the chronological local time can be obtained as a scaled version of the conventional sojourn time as

$$\bar{L}_B(t, a) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^t 1_{(|B_s - a| < \varepsilon)} ds = \sigma^{-2} L_B(t, a) \quad a.s. \quad \forall a, t. \quad (3.2.11)$$

Equation (3.2.11) clarifies the sense in which $\bar{L}_B(t, a)$ measures the amount of time (out of t) that the process spends in the neighborhood of a generic spatial point a . It turns out that a similar expression can be defined for more general processes

such as those driven by stochastic differential equations like (3.2.1). In this case, the measure $d[X]_s$ is random and equal to $\sigma^2(X_s)ds$. Hence, given the limit operation, a natural way to define the chronological local time is by

$$\bar{L}_X(t, a) = \frac{1}{\sigma^2(a)} \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_0^t 1_{[a, a+\varepsilon]}(X_s) \sigma^2(X_s) ds = \frac{1}{\sigma^2(a)} L_X(t, a) \text{ a.s. } \forall a, t. \quad (3.2.12)$$

This is the notion of local time that we will use extensively in what follows. It appears in other recent work on the nonparametric treatment of diffusion processes as Florens-Zmirou (1993), where it is sometimes referred to simply as “local time”.

We can estimate the local time of a diffusion via the following approximation:

$$L_M^n(t, a) = \frac{T}{nh_n} \sum_{i=1}^{\lfloor \frac{nt}{T} \rfloor} K\left(\frac{X_i - a}{h_n}\right) \quad (3.2.13)$$

where $[x]$ is the integer part of x . We have indeed:

Proposition 43. (Florens-Zmirou, 1993) *If as $n \rightarrow \infty$ we have $nh_n^4 \rightarrow 0$, then $L_M^n(t, a) \rightarrow L_M(t, a)$ in the \mathcal{L}^2 sense. The convergence is almost sure if $\frac{\log n}{nh_n^2} \rightarrow 0$.*

Proof. This can be reach following the Proposition 3 in Florens-Zmirou (1993). \square

3.3 The estimation of the diffusion coefficient

In this section we will concentrate on the models of the kind (3.2.1).

By constructing the nonparametric estimator of the diffusion process, we assume that, the process X_t , solution of the equation (3.2.1) is observed at $\{t_0 = 0 < t_1 < \dots < t_n = T\}$ in the time interval $[0, T]$, with $T > 0$. Subsequently, we let $\{X_t = X_{\Delta_n}, X_{2\Delta_n}, \dots, X_{n\Delta_n}\}$ be n equispaced observations at discrete points $\{t_0 = \Delta_n, t_2 = 2\Delta_n, \dots, t_n = n\Delta_n\}$, where $\Delta_n = \frac{T}{n}$. It is possible to derive the same construction per nonequispaced data. Given the above conditions, the theory of the estimation of coefficient diffusion is built on the Nadayara-Watson estimator of the

kind:

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^n K\left(\frac{x-\hat{X}_i}{h}\right)\tilde{\sigma}_i^2}{\sum_{i=1}^n K\left(\frac{x-\hat{X}_i}{h}\right)} \quad (3.3.1)$$

where $\tilde{\sigma}_i$ is a consistent estimate of the spot volatility at time t_i and h is a bandwidth parameter.

The following conditions will be used in the study of (3.2.1)

Assumption 4

1. $X_0 \in \mathcal{L}^2(\Omega)$ is independent of W_t , $t \in [0, T]$ and measurable with respect to \mathcal{F}_0
2. $\mu(x)$ and $\sigma(x)$ are defined on a compact interval I . $\mu(x)$ is once continuously differentiable, while $\sigma(x)$ is twice continuously differentiable.
3. A constant K exists such that $0 < \sigma(x) \leq K$ and $|\mu(x)| \leq K$.
4. (Feller condition for non-explosion). Given:

$$S(\alpha) = \int_0^\alpha e^{\int_0^y -\frac{2\mu(x)}{\sigma^2(x)} dx} dy \quad (3.3.2)$$

$$V(\alpha) = \int_0^\alpha S'(\alpha) \int_0^y \frac{2}{S'(x)\sigma^2(x)} dx \quad (3.3.3)$$

then $V(\alpha)$ diverges at the boundaries of I .

The first fully nonparametric estimator of the diffusion coefficient was proposed by Florens-Zmirou (1993). Her estimator is obtained by setting $\tilde{\sigma}_i^2 = \frac{n}{T}(\hat{X}_{i+1} - \hat{X}_i)^2$ in (3.3.1) thus implies

$$\hat{\sigma}_{FZ}^2(x) = \frac{n \sum_{i=1}^{n-1} K\left(\frac{x-\hat{X}_i}{h}\right)(X_{i+1} - X_i)^2}{T \sum_{i=1}^{n-1} K\left(\frac{x-\hat{X}_i}{h}\right)} \quad (3.3.4)$$

Florens-Zmirou (1993) uses naive kernel estimator; however the extension to a generic kernel was proposed by Jiang and Knight (1997)

The diffusion coefficient proposed in Florens-Zmirou (1993); Santon(1997); Jiang and Knight (1997) is constructed on the following quantity:

$$V^n(a) = \frac{1}{Th_n} \sum_{i=0}^{n-1} K \left(\frac{X_i - a}{h_n} \right) (X_{i+1} - X_i)^2 \quad (3.3.5)$$

Proposition 44. (Florens-Zmirou, 1993) *If $nh_n^4 \rightarrow 0$ as $n \rightarrow \infty$, then $V^n(x)$ converges to $\sigma^2(x)L_T(x)$ in the \mathcal{L}^2 sense.*

Thus dividing $V^n(x)$ by $L_T^n(x)$, we get a consistent estimator of $\sigma^2(x)$ given in (3.3.4). Results are also found about the asymptotic distribution if, additionally, $nh_n^3 \rightarrow 0$.

In our analysis, relying on the result of the previous Chapter, we want to substitute the quantity (3.3.5) with the following:

$$W^n(x) = \frac{1}{Tnh_n} \sum_{i=0}^{n-1} K \left(\frac{X_i - x}{h_n} \right) \tilde{\sigma}^2(t_i) \quad (3.3.6)$$

where $\tilde{\sigma}^2(t_i)$ is one of the spot volatility (2.3.2, 2.3.3, 2.3.4 and 2.3.6) computed on the observed trajectory of X_t .

We then define the estimator:

$$S^n(x) = \frac{\sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) \tilde{\sigma}^2(t_i)}{T \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right)} \quad (3.3.7)$$

We now prove that $S^n(x)$ is a consistent estimator of $\sigma^2(x)$.

Theorem 45. *If $nh_n^4 \rightarrow 0$ as $n \rightarrow \infty$, then $S_n(x)$ computed with $\tilde{\sigma}_{CR}^2, \tilde{\sigma}_E^2, \tilde{\sigma}_W^2$ and $\tilde{\sigma}_{OS}^2$ are consistent estimators of $\sigma^2(x)$ in the \mathcal{L}^2 sense.*

Proof. We first suppose that $\mu(x) = 0$ in (3.2.1). For every $w \in \Omega$, we consider the solution $X_t(w)$ of the previous equation and define the process $Z_t(w)$ which can be written by

$$Z_t(w) = \sigma(X_t) dW_t' \quad (3.3.8)$$

where W_t' is a standard Brownian motion in an auxiliary probability space $(\Omega', (\mathcal{F}'_t)_{0 \leq t \leq T}, P')$ and $\sigma(X_t(w))$ is the realization of $\sigma(X_t)$. We suppose that the solution of (3.2.1)

exists and is unique. Then is plausible to construct random variables in Ω by taking expectations in Ω' . We denote by E' the expected value in Ω' and by \mathbb{E} the expected value in Ω . We can write the following relation

$$\begin{aligned}
 \frac{1}{\Delta t} E' [Z_{i+1} - Z_i]^2 &= \frac{1}{\Delta t} \int_{t_i}^{t_{i+1}} \sigma^2(X_t(\omega) dt)^2 \text{ and we set } Z_i = Z_{t_i}(w) \\
 &= \frac{1}{\Delta t} \lim_{m_i \rightarrow \infty} \sum_{j=0}^{m_i} \left(Z_{i+\frac{m_i}{2}+j} - Z_{i+\frac{m_i}{2}+j-1} \right)^2 \\
 &= \frac{1}{\Delta t} \sum_{j=0}^{m_i} \left(Z_{i+\frac{m_i}{2}+j} - Z_{i+\frac{m_i}{2}+j-1} \right)^2 + \frac{1}{\Delta t} \sum_{j=m_i}^{\infty} \left(Z_{i+\frac{m_i}{2}+j} - Z_{i+\frac{m_i}{2}+j-1} \right)^2 \\
 &= \tilde{\sigma}_{m_i}^2(X_{t_i}(w)) + R_{m_i}(i, \Delta t)
 \end{aligned} \tag{3.3.9}$$

where the first part of the sum is the estimator of the spot volatility and the other is the remainder which tends to 0 as $m_i \rightarrow \infty$ and $\Delta t \rightarrow 0$ such that $m_i \Delta t \rightarrow 0$

Now it is possible to prove that, almost surely

$$\mathbb{E}[E' [(Z_t(w) - Z_s(w))^2] / \mathcal{F}_s] = \mathbb{E}[(X_t - X_s)^2 / \mathcal{F}_s], \tag{3.3.10}$$

where X is the solution of the previous equation and both equalities yield $\int_s^t \mathbb{E}[\sigma^2(X_u) / \mathcal{F}_s] du$ and for Cauchy-Schwartz inequality, we get almost surely:

$$\mathbb{E}[E'^2 [(Z_t(w) - Z_s(w))^2] / \mathcal{F}_s] \leq \mathbb{E}[(X_t - X_s)^4 / \mathcal{F}_s], \tag{3.3.11}$$

Redefining (3.3.6) as:

$$U_t^n(x) = \frac{T}{nh} \sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right) \sigma_{CR}^2(t_i)$$

where σ_{CR}^2 is computed as in Comte and Renault on the observed trajectory of X_t .

Now let us denote the $\mathcal{L}^2(\Omega)$ norm of X by $\|X\|^2 = \mathbb{E}[X^2]$. We then use almost sure identity of the previous relation and get:

$$\begin{aligned}
 \|U_t^n(x) - \sigma^2(x)L_t^n(x)\| &= \left\| \frac{T}{h_n} \sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right) \left\{ \frac{1}{\Delta t} E' [(Z_{i+1}^w - Z_i^w)^2] - R_{m_i}(i, \Delta t) \right\} - \sigma^2(x)L_t^n(x) \right\| \\
 &\leq \left\| \frac{T}{nh_n} \sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right) \left(\frac{1}{\Delta t} E' [(Z_{i+1}^w - Z_i^w)^2] - \sigma^2(x)L_t^n(x) \right) \right\| + \left\| \frac{T}{nh_n} \sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right) R_{m_i}(i, \Delta t) \right\| \\
 &\leq \left\| \frac{T}{h_n} \sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right) (X_{i+1} - X_i)^2 - \sigma^2(x)L_t^n(x) \right\| + \left\| \frac{T}{nh_n} \sum_{i=0}^{n-1} K\left(\frac{X_i - x}{h_n}\right) R_{m_i}(i, \Delta t) \right\|
 \end{aligned}$$

the first term tends to 0 as $nh_n^4 \rightarrow 0$ and $n \rightarrow \infty$ and the second term tends to 0 as $nh_n \rightarrow \infty$ and $\frac{m_i}{n} \rightarrow 0$.

If $\mu(t) \neq 0$, then we can write :

$\mathbb{E}[E'[(Z_t - Z_s)^2]] = \mathbb{E}[(X_t - X_s)^2] - \mathbb{E}\left[\left(\int_s^t \mu(X_u)du\right)^2\right]$ and the second term on the right hand side vanish as $s \rightarrow t$ i.e $n \rightarrow \infty$

We can derive the same results for the one side Exponential estimator and the Rolling Estimator

Noting that; $\tilde{\sigma}_E^2(t'_k) = (1 - \lambda) \sum_{j=1}^i \lambda^j [X(t'_{k-j+1}) - X(t'_{k-j})]^2 \leq \sigma_{CR}^2$, because $0 \leq \lambda \leq 1$ and $\tilde{\sigma}_W^2(t'_k) = \sum_{j=1}^{nL} \omega_j [X(t'_{k-j+1}) - X(t'_{k-j})]^2 \leq \sigma_{CR}^2$ and $\omega_j = \exp(-\alpha j)$ the results follow immediately.

We can use the same decomposition introduced above to prove the consistency of the estimator of Ogawa and Sanfelici .

$$\frac{1}{\Delta t} E' [Z_{i+1} - Z_i]^2 = \frac{1}{\Delta t} \int_{t_i}^{t_{i+1}} \sigma^2(X_t(\omega)) dt = \lim_{L \rightarrow \infty} \frac{1}{\Delta t} \sum_{l=1}^L (\Delta'_{k+l} \bar{X})^2 = RV_{OS} + R_{m_i}(i, \Delta t)$$

where $\Delta t = \frac{T}{n}$ the step between two observations

and noting that $\lim_{\Delta t \rightarrow 0} R_{m_i}(i, \Delta t) = 0$ almost surely in Ω . Following the previous demonstration step by step the result follows.

Lemma 46. (Florens-Zmirou) *If $nh^3 \rightarrow \infty$, then*

$$\sum_{i=1}^{[nt]} \mathbb{E} \left[\frac{n}{h_n} \left(K \left(\frac{X_{i-1} - x}{h_n} \right) [X_i - X_{i-1}]^2 - \sigma^2(x)/n \right)^2 \mid \mathcal{F}_{i-1} \right] \rightarrow \sigma^4(x) L_t(x)$$

where the above convergence is in probability

The proof is found in Florens-Zmirou (1993) lemma 2(b)

Lemma 47. *Let $g(x) : \mathbb{R} \rightarrow \mathbb{R}$ be a continuously differentiable bounded function, with bounded first derivative. Let $nh_n^3 \rightarrow 0$ when $n \rightarrow \infty$. Consider :*

$$G_t(x) = \frac{1}{\sqrt{nh_n}} \sum_{i=1}^{[nt/T]} K \left(\frac{X_i - x}{h_n} \right) [g(X_i) - g(x)] \quad (3.3.12)$$

then, as $n \rightarrow \infty$, $G_t(x) \rightarrow 0$ in \mathcal{L}^1 sense, and then in probability.

Proof: We have the known relationship

Proof.

$$\mathbb{E}[|G_t(x)|] \leq \mathbb{E} \left[\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{\lfloor nt/T \rfloor} K \left(\frac{X_i - x}{h_n} \right) |g(X_i) - g(x)| \right] \quad (3.3.13)$$

dividing the sum in terms such that $|X_i - x| \leq n^{-\frac{1}{3}}$ and their complementary $|X_i - x| > n^{-\frac{1}{3}}$. Then:

$$\begin{aligned} & \mathbb{E}[|G_t(x)|] \leq \\ & \mathbb{E} \left[\frac{1}{\sqrt{nh_n}} \sup_{|X_i - x| \leq n^{-\frac{1}{3}}} |g(X_i) - g(x)| \sum_{|X_i - x| \leq n^{-\frac{1}{3}}} K \left(\frac{X_i - x}{h_n} \right) + \frac{1}{\sqrt{nh_n}} \sum_{|X_i - x| > n^{-\frac{1}{3}}} K \left(\frac{X_i - x}{h_n} \right) |g(X_i) - g(x)| \right] \\ & \leq \mathbb{E} \left[\sqrt{nh_n} \sup_{|X_i - x| \leq n^{-\frac{1}{3}}} |g(X_i) - g(x)| \sum_{|X_i - x| \leq n^{-\frac{1}{3}}} \frac{1}{nh_n} K \left(\frac{X_i - x}{h_n} \right) \right] + \\ & + \mathbb{E} \left[\frac{1}{\sqrt{nh_n}} \sup_{|X_i - x| > n^{-\frac{1}{3}}} |g(X_i) - g(x)| \sum_{|X_i - x| > n^{-\frac{1}{3}}} K \left(\frac{X_i - x}{h_n} \right) \right] \quad (3.3.14) \end{aligned}$$

Now, by Taylor's rule, and given the boundedness of the first derivative of $g(\cdot)$, we get that $\sup_{|X_i - x| \leq n^{-\frac{1}{3}}} |g(X_i) - g(x)| = o(n^{-\frac{1}{3}})$. Then, using Proposition 43, we have

that the first term goes as $(nh_n^3)^{\frac{1}{6}}$, then it goes to zero as $n \rightarrow \infty$. The second term goes to zero given the boundedness of g , the fact that the argument of the kernel is greater than $\frac{1}{n^{\frac{1}{3}}h_n} = \frac{1}{(nh_n^{\frac{1}{3}})^3}$, and the fact that the kernel goes to zero faster than inverse polynomials when its argument goes to infinity, as in this case. \square

Theorem 48. *If $nh_n^3 \rightarrow 0$ then*

$$\sqrt{nh_n} \left(\frac{S_n(x)}{\sigma^2(x)} - 1 \right) \xrightarrow{\mathcal{L}} \frac{1}{\sqrt{L_T(x)}} \mathcal{N}(0, 1), \quad (3.3.15)$$

where the above convergence is in law, and $\mathcal{N}(0, 1)$ is a standard normal variable.

Where $S_n(x)$ refers to all nonparametric estimators

Proof of theorem : Consider the discrete filtration $\mathcal{F}_i = \mathcal{F}_{t_i}$, $i = 0, \dots, n$. Define the following term

$$\Theta_i = \sqrt{\frac{T}{nh_n}} K \left(\frac{X_{i-1} - x}{h_n} \right) [\tilde{\sigma}_{CR}^2(X_{i-1}) - \sigma^2(x)] \quad (3.3.16)$$

since $\tilde{\sigma}^2(\cdot)$ is bounded, Θ_i is bounded, and it is adapted to \mathcal{F}_{i-1} . We want now to verify the conditions of Lemma 37. In the limit $n \rightarrow \infty$, we have the following:

- 1 $\sum_{i=1}^{[nt/T]} \mathbb{E}[\Theta_i | \mathcal{F}_{i-1}] = \sum_{i=1}^{[nt/T]} \Theta_i$ which yields the following

$$\sqrt{\frac{T}{nh_n}} K \left(\frac{X_{i-1} - x}{h_n} \right) [\tilde{\sigma}^2(X_{i-1}) - \sigma^2(x)] + \sqrt{\frac{T}{nh_n}} K \left(\frac{X_{i-1} - x}{h_n} \right) [\tilde{\sigma}_{CR}^2(X_{i-1}) - \sigma^2(x)]$$

The first term tends to zero in probability, given Lemma 47, The second term tends to zero, given Proposition 43 and because the approximation of the σ_{CR}^2 which is bounded.

- 2 We have to prove that $\sum_{i=1}^{[nt/T]} \mathbb{E}[\Theta_i^2 | \mathcal{F}_{i-1}] \rightarrow \sigma^4(x)L_t((x))$.

Using lemma 46, this is equivalent to prove that:

$$H_t := \sum_{i=1}^{[nt/T]} \mathbb{E} \left[\frac{T}{nh_n} \left\{ K \left(\frac{X_i - x}{h_n} \right) [\tilde{\sigma}_{CR}^2(x) - \frac{n}{T}(X_i - X_{i-1})^2] \right\}^2 \mid \mathcal{F}_{i-1} \right] \rightarrow 0$$

We can use the derivation (3.3.9) and use the same reasoning the proof of the consistency, together with inequality (3.3.11), to get

$$\begin{aligned} H_t &= \sum_{i=1}^{[nt/T]} \mathbb{E} \left[\frac{T}{nh_n} \left\{ K \left(\frac{X_i - x}{h_n} \right) \left[\frac{n}{T} E'[(Z_i^w - Z_{i-1}^w)^2] - R_{m_i}(i, \Delta t) - \sigma^2(x) \right] \right\}^2 \mid \mathcal{F}_{i-1} \right] \\ &\leq \sum_{i=1}^{[nt/T]} \mathbb{E} \left[\frac{T}{h_n} \left\{ K \left(\frac{X_i - x}{h_n} \right) [(X_i - X_{i-1})^2] - R(i, \Delta t) - \sigma^2(x) \right\}^2 \mid \mathcal{F}_{i-1} \right] \end{aligned}$$

From this inequality, Proposition 43, and the remainder term of equation (3.3.9), we get the result.

- 3 We have to prove conditional Lindeberg condition. We have:

$$\mathbb{E} [|\mathbb{E} [\Theta_i^2 I_{\{|\Theta_i| > \varepsilon\}} | \mathcal{F}_{i-1}]|] = \sum_{|\Theta_i| > \varepsilon} \Theta_i^2 \quad (3.3.17)$$

Now, the sum (3.3.17) is bounded by $\sigma^4(x)L_t(x)$; moreover, we can rewrite $|\Theta_i| > \varepsilon$ as:

$$K \left(\frac{X_i - x}{h_n} \right) |\sigma^2(X_i) - \sigma^2(x)| > \varepsilon \sqrt{nh_n/T} \quad (3.3.18)$$

The left-hand side of equation (3.3.18) is bounded, thus as $n \rightarrow \infty$ we have $\varepsilon \sqrt{nh_n} \rightarrow \infty$ and the sum (3.3.17) vanishes in probability.

Thus, we fulfill the assumptions of lemma 37, then if we define $Y_t^n(x)$ as:

$$Y_t^n(x) = \sum_{i=1}^{[nt/T]} \Theta_i(x) \quad (3.3.19)$$

then we have that $Y_t^n(x)$ converges in law to the continuous martingale M_t with quadratic variation $[M, M]_t = \sigma^4(x)L_t(x)$. We then set $M_t = B(\sigma^4(x)L_t(x))$, where $B(t)$ is a standard Brownian motion. Now consider:

$$Z_t^n(x) = \sum_{i=1}^{[nt/T]} (W_{t_i} - W_{t_{i-1}}) \quad (3.3.20)$$

where W_t is the standard Brownian motion in (3.2.1). It is clear that Z_t^n converges in law to the standard Brownian motion W_t . Moreover we have:

$$\sum_{i=1}^{[nt/T]} \mathbb{E} [\Theta_i(W_{t_i} - W_{t_{i-1}}) | \mathcal{F}_{i-1}] = 0 \quad (3.3.21)$$

By equation (3.3.21), we get that M_t and W_t are orthogonal. We can also write $B(t) = M_{T(t)}$, where $T(t) = \inf_s (\frac{s}{\sigma^4(x)L_s(x)})$. Then, by Knight's Theorem (Knight, 1970), we get that $B(t)$ and W_t are independent Brownian motions. Then $B(t)$ and $L_t(x)$ are independent, since the filtration generated by X_t is included in the filtration generated by W_t . We then have that $Y_t^n(x) \rightarrow \sqrt{L_t(x)}\sigma^2(x)\mathcal{N}(0, 1)$ where $\mathcal{N}(0, 1)$ is

a standard normal random variable independent of $L_t(x)$. Since $L_t^n(x)$ converges in probability to $L_t(x)$, we have the desired result.

The asymptotic properties for other estimators are obtained in the same way, with the appropriate modifications

Mancini and Renò (2006) introduce the jump diffusion model. Following the classical framework, they stand always in a filtered probability space $(\Omega, (\mathcal{F}_t)_{t \in [0, T]}, \mathcal{F}, P)$, which satisfies the usual conditions, where W is a standard Brownian motion and J is a pure jump levy process or a doubly stochastic Poisson process with finite activity. Assuming that $(X_t)_{t \in [0, T]}$ is a real process such that $X_0 \in \mathbb{R}$ and

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dJ_t, \quad t \in]0, T] \quad (3.3.22)$$

and when J is a double stochastic process, either the non-parametric estimator function of the diffusion function $\sigma(x)$, or the $\mu(x)$ and $\lambda(x)$ counterpart can be obtained. So that in our setting, the volatility can be estimated as:

$$\hat{\sigma}_n^2(X) := \frac{\sum_{i=1}^n K\left(\frac{X-X_i}{h}\right) (\Delta_i X)^2 I_{\{(\Delta_i X)^2 \leq \vartheta(\delta)\}}}{\sum_{i=1}^n K\left(\frac{X-X_i}{h}\right) \delta} \quad (3.3.23)$$

where δ is the step between two consecutive observations and $\delta \rightarrow 0$ and $\vartheta(\delta)$ is a threshold function so that $\frac{\delta \ln \frac{1}{\delta}}{\vartheta(\delta)}$ then to zero. Furthermore, the estimated function satisfies

$$\sqrt{nh}(\hat{\sigma}_n^2(X) - \sigma^2(X)) \xrightarrow{st} MN\left(0, 2 \frac{\sigma^6(X) L_T^{**}(X)}{(L_T^*)(X)}\right) \quad (3.3.24)$$

where the above convergence is stable in law, $MN(0, U^2)$ is a random variable having a mixed normal law with the characteristic function $\phi(u) = E[e^{-\frac{U^2 u^2}{2}}]$ and $L_T^{**}(X) = L_T(X) \int_{\mathbb{R}^+} K^2(u) du + L_T(X^-) \int_{\mathbb{R}^-} K^2(u) du$.

In particular, they proved the consistency and the asymptotic normality of the diffusion coefficient in the presence of jumps with infinite activity and finite variation, such as the gamma process. The interested reader can find more details in the

original papers.

We can find other versions of kernel estimator in the literature. Except the Nadayara-Watson case to which we shall refer for in all this thesis, two other types of kernel smoothers deserve to be considered. The first introduced by Priestley and Chao (1972), is defined by

$$\hat{\sigma}_h^{PC}(x) = \frac{1}{h} \sum_{i=1}^n (X_i - X_{i-1}) \tilde{\sigma}_i K\left(\frac{x - X_i}{h}\right),$$

In the same spirit, Gasser-Müller (1979) proposed the following estimator

$$\hat{\sigma}_h^{PC}(x) = \frac{1}{h} \sum_{i=1}^n \tilde{\sigma}_i \int_{s_{i-1}}^{s_i} K\left(\frac{x - u}{h}\right) du,$$

where $s_0 = 0$, $s_i = (X_i - X_{i+1})/2$, $i = 1, \dots, n-1$, and $s_n = 1$. The Gasser-Müller estimator may also be written as

$$\frac{1}{h} \int_0^1 \sigma_n(u) K\left(\frac{x - u}{h}\right) du \tag{3.3.25}$$

where $\sigma_n(\cdot)$ is a piecewise constant function

$$\sigma_n(u) = \sum_{i=1}^n \sigma_i I_{[s_{i-1}, s_i]}(u)$$

In other words, The Gasser-Müller estimate is the convolution of $\sigma_n(\cdot)$ with $K(\cdot/h)/h$.

The previous estimator present some important differences, Chu and Marron (1991) refer to Nadaraya-Watson estimator as being an “evaluation type” and the Gasser-Müller one as a “convolution type”. Nevertheless, when the design points are approximately evenly spaced, the difference between the evaluation and the convolution estimators is unimportant. However, as Chu and and Marron (1991) argued, “ when the design point are not equally spaced, or when they are i.i.d random vari-

able, there are very substantial and important differences in these estimators. There are, nonetheless, certain basic principles that both estimator types obey. In all this thesis, we will not encounter such a problem because our simulated data and market data is chosen in order to be equally spaced. Therefore the use of the Nadaraya-Watson estimator is justified and we expect to obtain consistent results. There are wide classes of nonparametric regressions making use of ideas from orthogonal series. These include orthogonal polynomials and wavelets. The interested reader can find an exhaustive review in Hart (1997).

Chapter 4

Application of nonparametric estimation

In this chapter, we will use the Nadayara-Watson estimator to compare several spot volatilities performance in estimating the diffusion coefficient. The principal purpose is to construct our model with selective estimators of spot volatility proposed by several authors, which are based on the observed trajectories. Many nonparametric estimators have been proposed. However, microstructure noise is not taken into account. We test the new proposed estimators by means of Monte Carlo simulations of the univariate model of the short rate pioneered by Chan et al., and compare it with the previous analog estimators available in the literature, namely the estimators proposed by Florens-Zmirou and R. Renò. The results strengthen the findings available in the literature on the functional form of the diffusion coefficient.

4.1 Introduction

We are concerned with the problem of confronting several spot volatilities and analyse their capability as nonparametric estimator of diffusion coefficient. Since the early nineties, many authors have been questioned about the best way of estimating diffusion coefficient, namely volatility. The main motivation being the fact that it is

the substratum in practically every financial application.

In most continuous-time finance literature, the model specified as the underlying process of the state of asset prices, exchange rates, or spot interest rates is a time-homogeneous Ito diffusion process represented by the following stochastic differential equation (SDE):

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \quad (4.1.1)$$

with initial condition $X_0 = \eta$, where W_t is a standard real Brownian motion and the real function $\mu(x)$ and $\sigma(x)$ are such that a single solution X_t of the stochastic differential equation (4.1.1) exists. Our specific problem is to estimate the diffusion term $\sigma(x)$ when we observe a discrete realization of the process X_t , namely $n + 1$ observations $\hat{X}_0, \dots, \hat{X}_n$ at times $t_0=0 < t_1 < \dots < t_n = T$ in the interval $[0, T]$. We will consider throughout this chapter that the state variable refers to the spot interest rate so that, we suppose $X_t = r_t$. We need then to estimate $\sigma(r)$ where r_t satisfies the above setting.

The theory is constructed on Nadaraya-Watson estimators type which are given by the following formula :

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^n K\left(\frac{x-\hat{X}_i}{h}\right) \tilde{\sigma}_i^2}{\sum_{i=1}^n K\left(\frac{x-\hat{X}_i}{h}\right)} \quad (4.1.2)$$

where $\tilde{\sigma}_i$ is a consistent estimate of the volatility at time t_i and h is a smoothing parameter.

In many studies, the authors are concerned with estimating the parameters of equation (4.1.1), especially the diffusion coefficient. Returns and Volatilities are directly related to asset allocation, risk management, option pricing and proprietary trading. To achieve these objectives, the stochastic dynamics of the underlying state variables have to be specified correctly. For instance, option pricing theory allows us to value stock, index options or any value of a general asset and hedge against the risk of option writers once the model for the dynamic of underlying state variables is available. See the book of mathematical finance by Bjork (2009), Willmott (1998)

among others. Although many of the stochastic models in use are simple and convenient ones to facilitate mathematical derivations and statistical inference, they are not derived from any economics theory and hence cannot be expected to fit time series financial data. Thus, while the pricing theory gives relatively beautiful formulas when the underlying dynamics is correctly specified, it offers no or little guidance in choosing or validating the model. There is always the risk that misspecification due to parametric approach leads to erroneous valuation and hedging strategies. Furthermore there do not always exist the closed form solutions for the state variable and the derivative pricing specified by the function $\sigma(r) = \sigma(r; \vec{\theta})$ with $\vec{\theta}$ being a vector of real parameters. Hence, there are genuine needs for flexible stochastic modeling. Nonparametric approaches offer a complete and aesthetic treatment for tackling the above problems.

To avoid misspecification, recent researches have used nonparametric estimation techniques in order to avoid having to specify the functional form for the diffusion coefficient. They are useful for examining the extent to which the dynamic of a financial asset vary over the time. This approach can be applied in any branch of quantitative finance. For instance, it can be used to test some commonly accepted economic theories such as capital asset pricing model and shed a light on the question such as if the geometric Brownian motion really fits stocks index, whether the one factors interest rate models fit yields of the bond, and if the interest rate dynamics evolve with the time. Furthermore, when empirical data are available, one can fit directly the observed contingent claim prices with their associated characteristics and check if the estimated prices are consistent with the theoretical ones.

Ait-Sahalia (1996a) proposed a semiparametric model, in the sense that he assumed that the drift follows a linear equation of the form

$$\mu(r) = \nu[\theta - r] \tag{4.1.3}$$

which can be estimated using ordinary least squares

Using the Kolmogorov forward equation, Ait-Sahalia constructed his estimator on the fact that, provided the spot rate has the dynamic of the equation (4.1.1), the

volatility can be estimated by the following:

$$\sigma(r) = \frac{2 \int_{-\infty}^r \mu(y)\pi(y)dy}{\pi(r)} \quad (4.1.4)$$

where $\pi(r)$ is the unconditional distribution of r . Equation (4.1.4) is resolvable substituting $\mu(\cdot)$ by (4.1.3). This estimator is not fully nonparametric, because the drift needs to be specified. After specifying the drift, we need to identify the unconditional distribution $\pi(r)$. Nevertheless, following Scott (1992), Silverman (1986), we can accomplish this task with nonparametric techniques and replace $\pi(r)$ in the (4.1.4). Suppose our observations are equally spaced, and denote them by the discrete realization of the process r_t namely, \hat{r}_i $i = 0, 1, \dots, n$. Then the nonparametric estimator of the density is given by:

$$\hat{\pi}(r) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{r - \hat{r}_i}{h}\right) \quad (4.1.5)$$

where $K(\cdot)$ is the kernel function and h a bandwidth parameter which depends on n . One popular way of estimating densities through formula (4.1.5) is the histogram, where the kernel function is an indicator function of a compact real interval centered around zero, that is called *naive method*. Well-known serious drawback of the naive method is that it is by definition not a continuous function but has jumps at the endpoints of the window and zero derivatives everywhere else.

In his paper, Ait-Sahalia showed that estimation results contradict all the findings related to classical univariate models such as Vasicek and CIR. Ait-Sahalia (1996a) showed that under certain regularity conditions the estimator (4.1.4) is pointwise consistent and asymptotically normal. However, this estimation procedure has to rely on the parametric specifications of the drift term and works only for the strictly stationary diffusion processes.

The first example of a fully nonparametric estimator of the diffusion coefficient was proposed by Florens-Zmirou (1993). Her estimator is fundamentally different from that used in Ait-Sahalia (1996a), in the sense that she does not impose any

restriction on the functional form of the drift. Her estimator is obtained by setting $\tilde{\sigma}_i^2 = \frac{n}{T}(\hat{r}_{i+1} - \hat{r}_i)^2$ in (4.1.2), thus getting:

$$\hat{\sigma}^2(r) = \frac{T \sum_{i=1}^{n-1} K\left(\frac{r-\hat{r}_i}{h}\right) (\hat{r}_{i+1} - \hat{r}_i)^2}{n \sum_{i=1}^{n-1} K\left(\frac{r-\hat{r}_i}{h}\right)} \quad (4.1.6)$$

The variance estimator (4.1.6) can be seen as the conditional variance of r , and it seems to have nice properties because the same kernel is in the numerator and in the denominator; this feature cancels out the biases in finite samples coming from nonparametric estimation of the density. The estimator (4.1.6) has been used by Jiang and Knight (1997) on Canadian interest rates, and by Stanton (1997) on U.S. interest rates.

An analogous estimator to that of Florens-Zmirou (1993) has been proposed by Bandi and Phillips (2003), and implemented in Bandi (2002). The corresponding spot volatility estimator is given by:

$$\tilde{\sigma}_i = \frac{n}{T} \frac{1}{m_i} \sum_{j=0}^{m_i} [\hat{r}_{t_{i,j+1}} - \hat{r}_{t_{i,j}}]^2 \quad (4.1.7)$$

where $t_{i,j}$ is the subset of indexes such that

$$t_{i,0} = \inf\{k \geq 0 : |\hat{r}_k - \hat{r}_i| \leq \varepsilon_s\},$$

and

$$t_{i,j+1} = \inf\{k \geq t_{i,j} + \Delta t : |\hat{r}_k - \hat{r}_i| \leq \varepsilon_s\},$$

m_i is the number of time that $|\hat{r}_k - \hat{r}_i| \leq \varepsilon_s$, ε_s is the parameter to be selected and Δt is the time step between adjacent observations. While the Florens-Zmirou estimator weights the observation r_t with the quadratic variation at time t , the Bandi and Phillips estimator weights the observation r_t with the average quadratic variation of price at all observations that differ from r_t of the parameter less than ε_s . It is important to remark that Bandi and Phillips do not require the process (4.1.1) to be stationary, but only the weaker condition to be recurrent. This condition can be

important theoretically, since Bandi (2002), showed that there is no strong support to the assumption of stationarity of interest rate data.

In the framework of Nadayara-Watson, Renò (2008) proposed two new estimators. The first estimator is based on results of the Fourier coefficients of volatility implemented in Malliavin and Mancino (2002), and the second is based on the realized volatility proposed by T.Andersen and T.Bolloserv (1998).

These estimators were shown to be consistent and asymptotically normal.

This chapter is organized as follows. Section 4.2 describes the volatility models that we consider, while Section 4.3 describes methodology and results. In Section 4.4, we discuss our empirical findings and determine the best volatility model, while Section 4.5 present some results based on S&P 500 data and Section 4.6 concludes.

4.2 Spot volatility estimators

We are going to compare the performance of different spot volatilities models in terms of their ability to estimate the volatility paths of returns, defined as in (4.1.1). The basic spot volatility model is defined in the following way:

$$\hat{\sigma}^2(t'_k) = \frac{(X(t'_k) - X(t'_{k-1}))^2}{\Delta'} \quad \text{where } \Delta' = \frac{T}{N}$$

Nevertheless due to the simplicity of this estimator, it is shown to be very noisy. Therefore, academicians look for estimators which can reconstruct accurately the path of the volatility.

The literature on modeling and forecasting ex-post financial market volatility extends the original idea of Merton (1980), who observed that the variance over a fixed interval could be estimated arbitrarily, although accurately, as the sum of squared realizations, provided the data are available at a sufficiently high sampling frequency. Unfortunately, the crude utilisation of this realized variance do not take microstructure noise into account. Zhang et al.(2005) introduce a consistent model that allows for the microstructure contaminations.

Several different volatility models are developed in the literature by assuming

Comte and Renault (1998)
$\tilde{\sigma}_R^2(t'_k) = \frac{1}{m} \sum_{j=1}^m \frac{(X(t'_{k-m/2+j}) - X(t'_{k-m/2+j-1}))^2}{\Delta'} \quad \text{where } \Delta' = \frac{T}{N}$
Ogawa and Sanfelici(2010)
$\tilde{\sigma}^2(t'_k) = \frac{G(\rho)^{-1}}{L} \sum_{i=1}^L \frac{(\Delta'_{k+i} \bar{X})^2}{\Delta'}$
Foster and Nelson (1996)
$\tilde{\sigma}_E^2(t'_k) = (1 - \lambda) \sum_{j=1}^i \omega_j [X(t'_{k-j+1}) - X(t'_{k-j})]^2 ..$
Andreou and Ghysels(2002)
$\tilde{\sigma}_W^2(t'_k) = \sum_{j=1}^{n_L} \omega_j [X(t'_{k-j+1}) - X(t'_{k-j})]^2.$
Malliavin and Mancino(2002)
$\tilde{\sigma}(t) = \lim_{M \rightarrow \infty} \sum_{k=0}^M [a_k(\sigma^2) \cos(kt) + b_k(\sigma^2) \sin(kt)].$

Table 4.1: Specification of the Spot Volatilities

different structures for σ_t^2 . We list the specification of all the spot volatility models used in this chapter in table 4.1.

4.2.1 Nonparametric estimation

Let's reconsider the following SDE:

$$\begin{cases} dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \\ X_0 = \eta \end{cases} \quad (4.2.1)$$

defined over the interval $[0, T]$, in the filtered probability space $(\Omega, (\mathcal{F}_t)_{0 \leq t \leq T}, P)$ satisfying the usual conditions.

The nonparametric estimator for the diffusion function $\sigma^2(X_t)$ of a general diffusion process is based on observing X_t at $\{t_1, t_2, \dots, t_n\}$ in the time interval $[0, T]$. We will discuss only the equispaced data case. Subsequently, we let $\{X_t = X_{\Delta_n}, X_{2\Delta_n}, \dots, X_{n\Delta_n}\}$ be n equispaced observations at discrete points $\{t_1 = \Delta_n, t_2 = 2\Delta_n, \dots, t_n = n\Delta_n\}$, where $\Delta_n = \frac{T}{n}$.

We assume $K(\cdot) \in L^2(\mathbb{R})$ to be a bounded kernel, that is, $\int_{-\infty}^{\infty} K(x)dx =$

1, continuously differentiable and with bounded first and positive derivative, with $\lim_{x \rightarrow \infty} K(x) = \lim_{x \rightarrow -\infty} K(x) = 0$. We opt to use the gaussian kernel, and therefore we set

$$K(s) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \quad (4.2.2)$$

We also define a sequence of bandwidths h_n such that:

$(h_n)_{n \in \mathbb{N}}$ is a real sequence such that, as $n \rightarrow \infty$, we have $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$.

A very common value for h_n used in applications (Scott, 1992), (Silverman, 1986) is the following

$$h_n = h_s \hat{\sigma} n^{-\frac{1}{5}} \quad (4.2.3)$$

where h_s is a real constant to be tuned, and $\hat{\sigma}$ is the standard deviation of analysed sample, $\hat{\sigma}^2 = \text{Var}[\hat{r}_i]$. Following the same idea of chapter 3, we construct the nonparametric estimator as

$$S_n(x) = \frac{\sum_{i=1}^n K\left(\frac{X_{i\Delta_n} - x}{h}\right) \tilde{\sigma}_i^2}{\sum_{i=1}^n K\left(\frac{X_{i\Delta_n} - x}{h}\right)} \quad (4.2.4)$$

where $\tilde{\sigma}_i$ is a consistent estimator of the spot volatility, and the first factor at the numerator and the denominator can be use to have a discrete approximation of the kernel function.

4.3 Methodology and results

For analysing the spot volatility, we simulate discrete data from the model proposed by Chan et al. (1992) which includes the CIR (1985) and Vasicek (1977) models as a particular case, in which we add microstructure contaminations. By Monte Carlo experiments, we evaluate the performance of the competing estimated volatilities in measuring the value of the true volatility path. The instantaneous variation of the

interest rate is given by the following formula

$$dr_t = \mu(\alpha - r_t)dt + \sigma r_t^\gamma dW_t \quad (4.3.1)$$

where W_t is a standard Brownian motion. The model has affine drift, with stochastic volatility, with the parameter γ measuring the degree of dependence of the variance on the interest rate level. We assume further that the logarithmic noises are normal and independent from r in order to avoid the leverage effect.

We simulate the model (4.3.1) through a simple Euler first order discretization scheme to get the trajectories of the short rate r_t and from that we deduce the value of the spot volatility as $\sigma_t^2 = \hat{\sigma}^2 r_t^{2\hat{\gamma}}$, which is also known as CEV model. When taking into account microstructure noise, observed returns must be sampled at a lower frequency so as to have accurate realized volatility type estimates. In our analysis, we face two types of errors, the first inserted voluntarily and the second which is attributed to discretization. So that one part of the bias is caused by this error and the other part is related to discretization. As point out by Jiang (1998), direct estimation of Chan model is not possible from discretely observed data because their transition density or exact moment conditions is out of reach. However, several authors have estimated the parameters of the model (4.3.1), and in this work we will use the ones estimated by Jiang (1998) using indirect inference. Parameter estimates are $\hat{\alpha} = 0.079(0.044)$, $\hat{\mu} = 0.093(0.100)$, $\hat{\gamma} = 1.474(0.008)$, $\hat{\sigma} = 0.794(0.019)$, where standard errors are in brackets. When simulating the Chan model, the choice of spot rate interest is crucial, so that, we simulate the model with parameters resembling actual interest rates distribution. We use as starting value 3% which is almost equal to the smallest value in the 3-months T-bill analyzed by Jiang (1998). In the same vein Mancini and Renò (2006) argued that 3-months T-bill is an appropriate choice since 7-day Eurodollar produces many jumps. The true and observed data are generated over a daily trading period of $T = 24$ hours as in the foreign exchange markets, since it has no geographical location and no “business hour” limitation. Andersen and Bollerslev (1998) and other authors computed the ex-post daily foreign exchange volatility by aggregating $D = 288$ squared five-minute returns. In our

simulations we will also compute the spot volatility with $D = 144$, corresponding to ten-minutes returns, and $D = 720$ corresponding to two-minutes returns. In the end, we introduce a microstructure error with zero mean and the variance to the noise to signal rate equal to 2. In order to assess the performance of different procedures for estimating the volatility, we calculate the value of the following three loss functions. Firstly, we determine the integrated squared error $\int_0^T (\hat{\sigma}_t^2 - \sigma_t^2)^2 dt$ between the generated and the reconstructed variance trajectories and we calculate the average and the standard deviation of this measure over a given number of days. Broadly speaking, the estimation approach is better when the average (or standard deviation) is smaller. Successively, we evaluate the statistic relative error defined as $\pi_t = \frac{\hat{\sigma}_t^2 - \sigma_t^2}{\sigma_t^2}$ and the relative bias is afterwards defined by: $\bar{\pi} = \frac{1}{N} \sum_{t=1}^N \pi_t$ and σ_t^2 is the simulated volatility paths. Finally, we calculate The Root Mean Square Error defined as $RMSE = \left(\frac{1}{N} \sum_{t=1}^N \pi_t^2 \right)^{1/2}$.

Once the values of the loss functions are calculated, it is possible to order the models according to their losses. The model with the minimum loss is the best model.

Figure 4.3.1 shows the true and the estimated variance paths for $N_d = 50$ days of transaction and sampling frequency of 10 minutes. Each figure contains the simulated path of the true volatility on the bottom and the estimated volatility obtained with different estimators on the top for any graph. We do not show the results related to 5 minutes and 2 minutes here because the trends are almost the same.

For computing $\hat{\sigma}_{CR}^2$, we extract the daily observation times in such a way that the difference is 7.2 minutes, $D = 200$, where D represents the number of step we use to divide intervals for any observation $t_i = i$. At the same time for the real-time scheme estimator we use the whole data set. In the estimating of the Ogawa and Sanfelici estimator the parameters L and M are considered but, M is the smoothest parameter needed to filter the noise in the presence of microstructure effects. For the one-sided exponential and the one sided rolling window volatility, two smoothing parameters have been added $\lambda = 0.94$ and $\alpha = 0.665$. The Fourier parameters are chosen arbitrarily as $N_F = 50$ and $\delta = 2$ in order to filter out high frequency noise modes. Moreover, this estimator has been computed by reconstructing the Fourier expansion

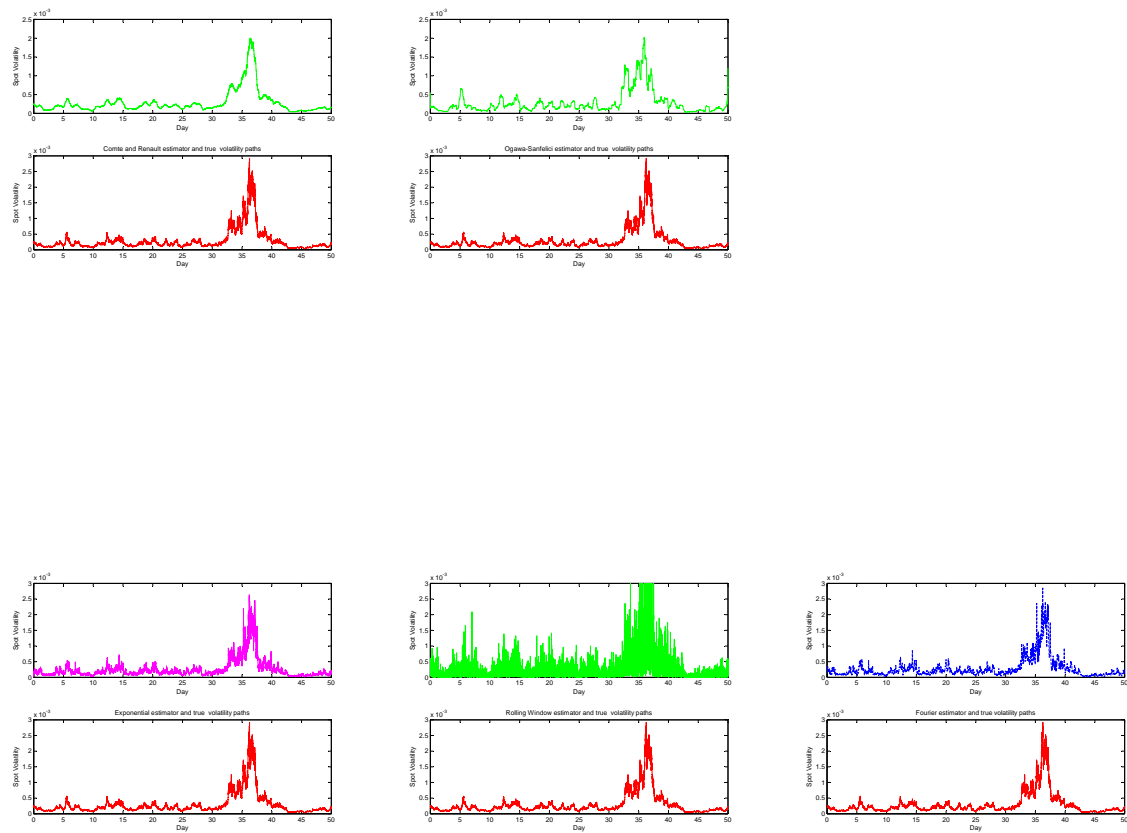


Figure 4.3.1: True and Estimated variance paths: Top Left Comte and Renault estimator; Top Right Ogawa and Sanfelici Estimator; Bottom Left Exponential Estimator; Bottom middle rolling Estimator; Bottom Right Fourier Estimator.

on each day separately, because otherwise it would provide a too smooth trajectory but nevertheless capable of capturing the peaks of the volatility path. The best volatility estimates are provided by the σ_F^2 , $\tilde{\sigma}_{CR}^2$ and the $\tilde{\sigma}_{OS}^2$ in term of the Integrated Squared Error, when the frequency of observation is high, it is easy to notice how any volatility estimator react with respect to microstructure contaminations. The Bias and the RMSE statistic equal in absolute value to around $1.3874e - 04$ for all the estimators. Table 4.2 and Table 4.3 report the results of different loss functions that we consider, as well as the mean and the standard deviation of every spot volatility estimator.

In order to compare the ability of reconstructed estimators in estimating the diffusion coefficient, we simulate 100 trajectories containing 500 days each and substitute the corresponding $\tilde{\sigma}_i$ with each spot volatility estimators. For this purpose, we still use the Chan model, successively we compute the spot volatility path, of the simulated trajectories by using the asymptotic properties of the estimator (4.2.4), we reconstruct the corresponding nonparametric estimator for any spot volatility estimator.

When comparing the estimators in small sample, we have to check not only the unbiasedness and the precision of the estimators, but the reliability of the asymptotic confidence intervals, since those are the ones actually used to draw inference. The variance of $\hat{\sigma}^2(x)$ can be consistently estimated by $\hat{V}[\hat{\sigma}^2(x)] = \hat{\sigma}^4(x) / \sum_{i=1}^n K((X_i - x)/h_n)$ (Jiang 1998).

Figure 4.3.2 shows the average measurements on Monte carlo simulations, estimated diffusion coefficient for 500 days of transaction for the different estimators, together with the corresponding 5% and 95% confidence intervals, obtained with the estimator (4.2.4).

We suppose to have intraday data, where the number of observations is $N=50000$, the number of observations intraday is $D=100$, and the number of days take into account is $ND=500$. We will furthermore select $h_s = 1.06$ and the variance is backed out from the considered data set.

We want to confront the proposed estimators to the popular estimators of Florens-

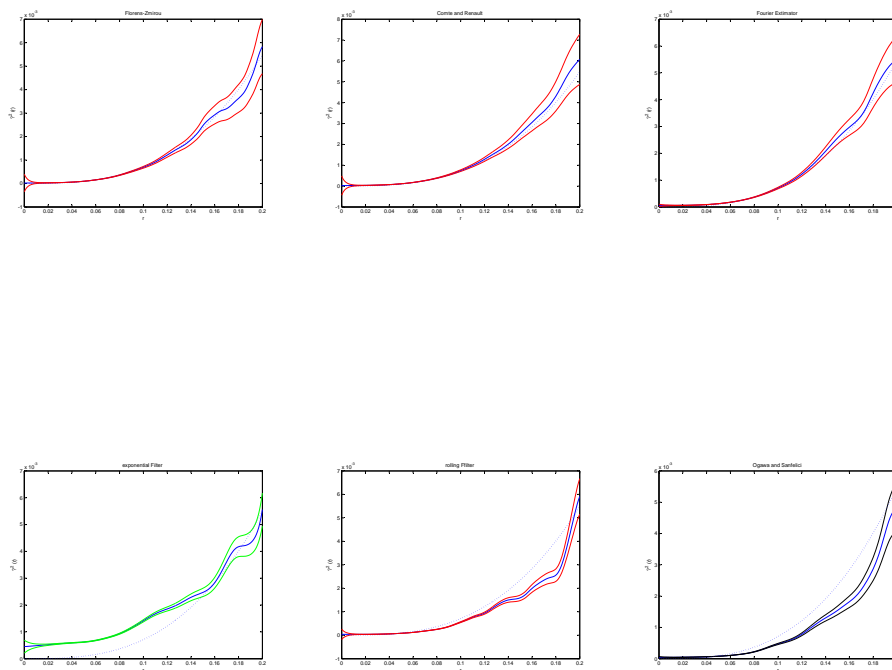


Figure 4.3.2: The generated Monte Carlo squared diffusion coefficient according to the CEV model together with the average estimate, using 100 replications, with six methods: the classical estimator (4.2.4), top; Florens-Zmirou, Comte and Renault and Fourier estimator and the Exponential, Rolling and Sanfelici and Ogawa estimator, bottom. The outer lines are 5% and 95% confidence intervals, as computed with the simulations.

Zmirou, Fourier and Comte and Renault since the latter has been shown to be more precise because it has narrower confidence intervals Renò (2008). We notice that, the Florens-Zmirou, the Fourier, Exponential filter and Rolling filter estimators use low frequency data, while the Comte and Renault and Ogawa and Sanfelici estimator use the whole dataset. The Comte and Renault displays the narrowest confidence bands. Notwithstanding, the Comte and Renault, Florens-Zmirou, Ogawa and Sanfelici estimators being consistent, the Florens-Zmirou seems to be the less precise one. We notice that, the Exponential filter and Rolling estimators are biased, Exponential filter provides a very striking feature. This can be attributed to the smoothing parameter which slows the value of the spot volatility in long run time.

Figure 4.3.3 shows the simulated competing nonparametric estimators with microstructure noise contaminations. When introducing microstructure noise contaminations, the order of precision seem to be the same, however the Florens-Zmirou estimator remain always the less precise. It is worth noting that, the bandwidth have to be increased when we introduce microstructure noise. For this purpose we consider $h=4$ which is the considered value when analyse the real data

In both cases, the result found by Jiang (1998) are confirmed in the sense that the volatility function is neither constant nor a linear function of the state variable as demonstrated by the trend of all estimators. The nonparametric estimator captures very well the functional form of the volatility evidenced in the Chan model, as estimated by Jiang (1998) and Renò (2008), because the function is nonlinear and overall increasing of the short rate. We can also remark that, low interest rates are associated with low volatility and high interest rates are associated with high volatility, suggesting that low interest rates are more like absorbing states and interest rates are more likely to stay low than high. This is confirmed futhermore by the fact that, the 95% confidence band is narrower in the begining but tends to get wider dramatically toward the end for the lack of enough observations around the high levels of interest rate.

When simulating the spot volatility and the nonparametric estimation, we don't take into account the model pioneered by T.Andersen and T.Bolloso (1998) and used in Renò (2008) since theoretical results and Monte Carlo simulations show that

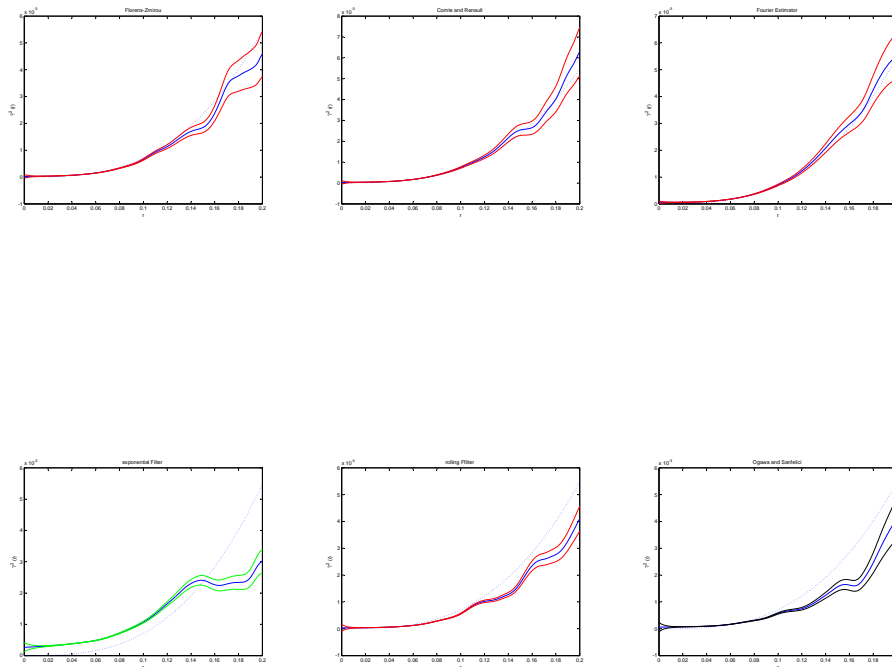


Figure 4.3.3: The generated Monte Carlo squared diffusion coefficient with microstructure noise according to the CEV model together with the average estimate, using 100 replications, with six methods: the classical estimator (4.2.4), top; Florens-Zmirou, Comte and Renault and Fourier estimator and the Exponential, Rolling and Sanfelici and Ogawa estimators, bottom. The outer lines are 5% and 95% confidence intervals, as computed with the simulations.

Method	Int.Sq.Err	Relative Bias	RMSE	Mean	Std Dev
$\tilde{\sigma}_R^2$	$2.1652e - 05$	$-1.3874e - 04$	$1.3874e - 04$	$1.0828e - 03$	$1.7336e - 03$
$\tilde{\sigma}_E^2$	$3.2723e - 05$	$-1.3875e - 04$	$1.3875e - 04$	$1.175e - 03$	$1.8928e - 03$
$\tilde{\sigma}_W^2$	$2.2723e - 05$	$-1.3878e - 04$	$1.3878e - 04$	$7.6741e - 03$	$2.0531e - 03$
$\tilde{\sigma}_{OS}^2$	$9.4157e - 05$	$-1.3879e - 04$	$1.3879e - 04$	$1.0738e - 03$	$1.5493e - 03$
$\tilde{\sigma}_F^2$	$10.304e - 05$	$-1.3177e - 04$	$1.3177e - 04$	$1.8906e - 02$	$1.9487e - 02$

Table 4.2: Integrated Squared Error, Rel.Sta.Err, RMSE, Mean and standard deviation of the Spot volatility estimator with

Method	Int.Sq.Err	Relative Bias	RMSE	Mean	Std Dev
$\tilde{\sigma}_R^2$	$1.9374e - 05$	$-6.9372e - 05$	$6.9372e - 05$	$1,10454e - 03$	$2.1207e - 03$
$\tilde{\sigma}_E^2$	$2.8713e - 05$	$-6.9376e - 05$	$6.9376e - 05$	$9.8229e - 04$	$2.259e - 03$
$\tilde{\sigma}_W^2$	$1.67375e - 05$	$-6.9393e - 05$	$6.9376e - 05$	$7.4101e - 04$	$2.2502e - 03$
$\tilde{\sigma}_{OS}^2$	$1.3057e - 05$	$-6.9385e - 05$	$6.6323e - 05$	$1.81234e - 03$	$1.7354e - 03$
$\tilde{\sigma}_F^2$	$7.72471e - 05$	$-6.5873e - 05$	$6.5873e - 05$	$5.1434e - 02$	$1.2031e - 02$

Table 4.3: Integrated Squared Error, Rel.Sta.Err, RMSE, Mean and standard deviation of the Spot volatility estimator with D=288

it is indistinguishable from the model of Comte and Renault (1998). Furthermore, it is worth noting that, Bandi and Phillips (2003) estimator nests the previous models and the Florens-Zmirou (1993) estimator in estimating the diffusion coefficient; see Renò, Roma and Schaefer (2006) for the latter case. This depends on the choice of the parameter ε_s in Bandi and Phillips (2003) estimator formula. Which can be tantamount to latter when ε_s is small and contains the T.Andersen and T.Bolloserv estimator when ε_s get bigger.

Method	Int.Sq.Err	Relative Bias	RMSE	Mean	Std Dev
$\tilde{\sigma}_R^2$	$1.9374e - 07$	$-4.4672e - 06$	$4.4672e - 06$	$1,10454e - 03$	$2.1207e - 03$
$\tilde{\sigma}_E^2$	$2.8713e - 06$	$-4.5676e - 06$	$4.5676e - 06$	$9.8229e - 04$	$2.259e - 03$
$\tilde{\sigma}_W^2$	$1.67375e - 06$	$-4.7893e - 06$	$4.7896e - 06$	$7.4101e - 04$	$2.2502e - 03$
$\tilde{\sigma}_{OS}^2$	$1.3057e - 06$	$-4.3985e - 06$	$4.3985e - 06$	$1.81234e - 03$	$1.7354e - 03$
$\tilde{\sigma}_F^2$	$7.72471e - 07$	$-3.5783e - 06$	$3.5787e - 06$	$5.1434e - 02$	$1.11e - 02$

Table 4.4: Integrated Squared Error, Rel.Sta.Err, RMSE, Mean and standard deviation of the Spot volatility estimator with D=720

N	Min	Max	Mean	Std. Dev.	Skew	Kurt	Autocorr
11611296	-5.6316	3.6570	1.1956e-005	0.0212	-2.4358	1.4893e+003	-8.8959e+003

Table 4.5: Summary Statistics of tick-by-tick returns on the S&P 500 of the January 2, 1990 to december 29, 2006.

4.4 Data and Empirical Methodology

Our empirical analysis is based on the data from the NYSE Transaction and Quote (TAQ) database which records all trades and quotations for the securities listed on the NYSE, AMEX, NASDAQ, CME, and the regional exchanges. The sample consists of the tick-by-tick data of S&P 500 index futures recorded at Chicago Mercantile Exchange. The data spans from January 2, 1990 to December 29, 2006, a period of 4,274 trading days (11,611,297 observations). We preferred futures data over the index data. First, the S&P 500 index is calculated based on the last transaction price of each of the (500) stocks comprising the index, and not every stock trades every minute. This results in an infrequent trading problem, meaning the index lags actual developments, especially at the opening of trading since it takes some time before each of the 500 stocks begins trading. For the measuring and forecasting of stock market volatility it is of crucial importance to correctly measure the trade-by-trade return.

Since the consecutive prices are nonstationary, we find appropriate to study changes in price. Followings other reseachers, we study the compound return (difference in the logarithmic value of the two consecutive prices). Table 4.5 shows the summary statistics of the tick-by-tick returns.

The number of ticks in our sample varies considerably. The average return of the tick-by-tick data is negligible with respect to its standard deviation. The returns are skewed slightly to the left. The kurtosis is most higher than 3, the correspondant kurtosis of a Normal distribution. The Jarque-Bera test obtained with the confidence interval of 5% rejects the hypothesis of normal distribution of returns.

Figure 4.4.1 shows the trend of S&P 500 index price, the S&P 500 index log returns and the frequency distribution log returns when the sample frequency is one minute and the analyzed period range from January 2, 1990 to December 29, 2006.

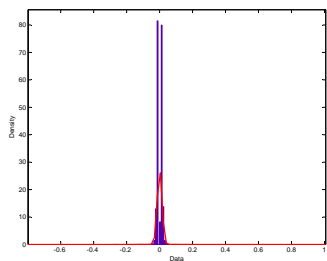
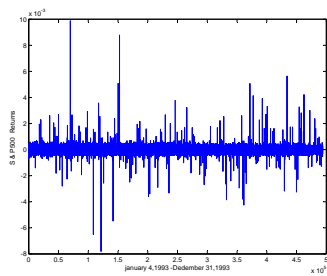
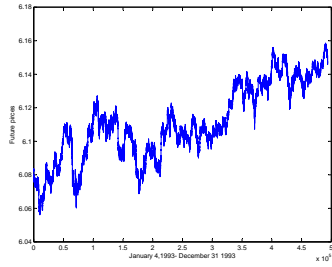


Figure 4.4.1: Top Panel: S&P 500 price index one minute trend for the period January 2, 1990 to December 29, 2006. Middle Panel S&P 500 one minute log returns index from the period January 2, 1990 to December 29, 2006. Bottom Panel Frequency distribution of S&P 500 returns index from the period January 2, 1990 to December 29, 2006 compared with the Normal distribution.

4.5 Results

The main subject of this section is to investigate the properties of the different spot return volatility, we use intradaily return on the S&P 500 range over the January 2, 1990 to December 29, 2006. We graph the corresponding spot volatility for the S&P 500 over the period January 2, 1990 to December 29, 2006. This variance are obtained using 1-minute observation frequency. The different estimators show to reconstruct in almost the same way the volatility as presented in the Figure 4.5.1

We have analysed the properties of the different realized volatility estimators, the Fourier estimator evidences the highest standard deviation; suggesting that it varies a lot over time, the Ogawa and Sanfelici estimator is more stable through time because it has the smallest standard deviation. When breaking down skewness and kurtosis, the distributions variances and standard deviations are clearly non-normal both are skewed right and leptokurtic. The fourier estimator shows to have the smallest skewness and kurtosis, while the rolling estimator counterpart presents the biggest one.

4.6 conclusion

In this chapter, we compared several spot volatility models, in term of their ability to reconstruct the true volatility. Successively, we have used these spot volatilities to estimate the diffusion coefficient function in the equation (4.2.4) by means of the Nadayara-Watson estimators. The estimators are compared with the other counterparts available in the literature, via Montecarlo simulations of the Chan et al. model.

The performance of the volatility models has been measured at different sampling frequency using three loss functions, the realized volatility proposed by Comte and Renault reveals to be the most accurate among the studied volatility.

We use the studied spot volatility to compute for a high frequency time series of S & P 500

When turn to the nonparametric estimator of the diffusion coefficient our findings

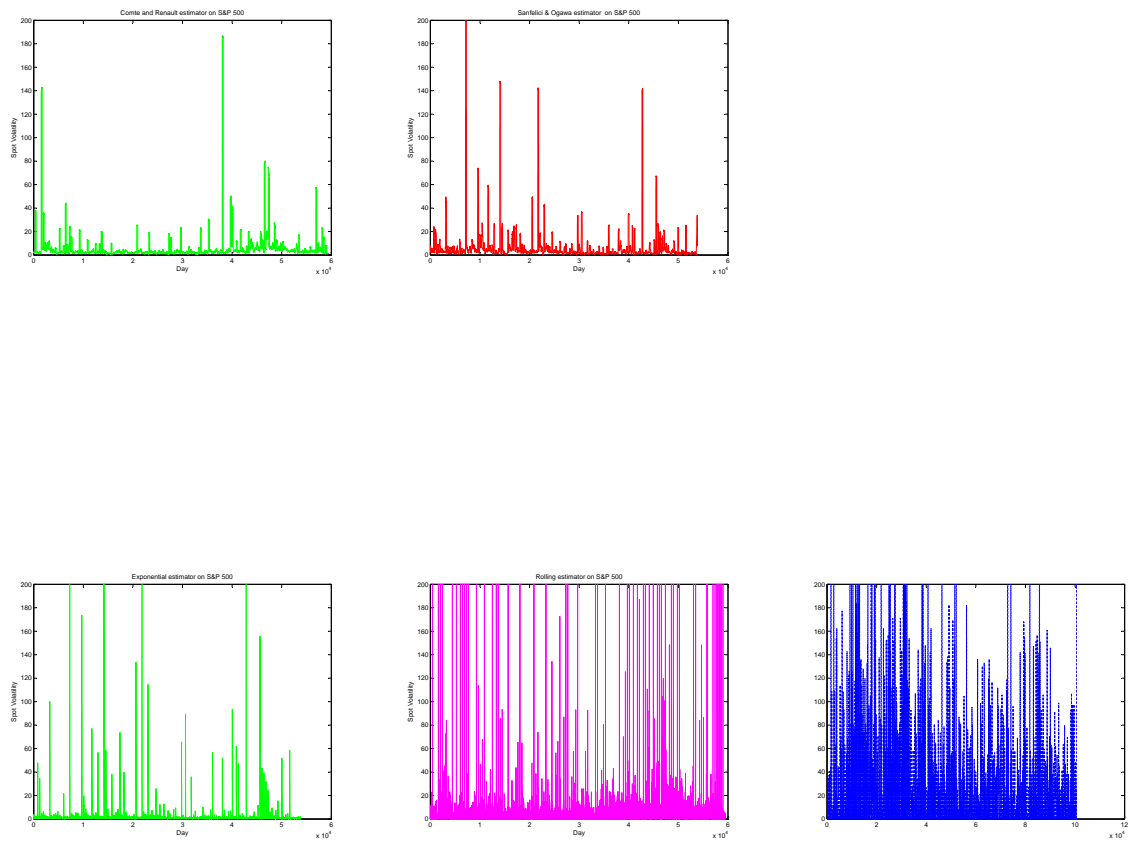


Figure 4.5.1: Estimated variance paths: Top Left Comte and Renault estimator; Top Right Ogawa and Sanfelici Estimator; Bottom Left Exponential Estimator; Bottom middle rolling Estimator; Bottom Right Fourier Estimator constructed on S&P 500 index from the period January, 2 1990 and December, 29 2006.

are in line with those presented in the literature, the estimator of Comte and Renault is the most accurate and evidence no bias, while the exponential counterpart is the less accurate and shown to be very biased. However, those nonparametric estimators can be used to variety of applications

Chapter 5

Option pricing with nonparametric diffusion coefficient

5.1 Introduction

More than three decades ago, F. Black and M. Scholes published a fundamental paper on options pricing, where they introduced the so called the *Black-Scholes equation* and the *Black-Scholes model*. In the same year, Merton published a paper on the same topic independently. The work of Black and Scholes had positive effects on the world of finance regarding the problem of options pricing. Nevertheless, given the success of that model in parsimoniously describing market option prices, it's not immediately obvious what the benefits of making such a modeling choice might be.

Some of the important assumptions of the *Black-Scholes model* are that the underlying asset's price process is continuous, the ability to hedge without transaction costs, the independent Gaussian returns, and the volatility is constant. Only this last parameter is not directly observable but it can be estimated from time series data. If the latter were true, this last assumption would lead to the conclusion that if we plot volatility against the strike price we would obtain a straight line, parallel to the horizontal axis. Equalizing the *Black-Scholes model* with the market observed option price and solving for volatility give us the implied volatility. Moreover, for

options with different maturities and different strikes which are written on the same stock, one should find the same implied volatility, meaning that the volatility of the stock is unique. Such it not the case. When plotting implied volatility using real market data one typically obtains a convex curve, known as the “smile curve” or the “volatility smile” for the stock price, with minimum price “at the money” i.e. where the strike price is equal to the underlying spot.

From empirical studies, we can infer that the distribution of stock price returns is highly peaked and fat-tailed relative to the Normal distribution and squared log returns evidence volatility clustering. Fat tails and the high central peak are characteristics of mixtures of distributions with different variances. This motivates some authors to model variance as a random variable. The volatility clustering feature implies that volatility (or variance) is auto-correlated. In the model, this is a consequence of the mean reversion of volatility.

In order to have a more realistic approach to the problem of option pricing, jump models and stochastic volatility models have been introduced. Jump models deal with the assumption of continuity by allowing the spot asset’s process to jump Merton (1976). When studying stochastic volatility models the volatility is described by a stochastic process. These models are used in order to price options where volatility varies over time. Stochastic volatility models are useful because they explain in a self-consistent way why it is that options with different strikes and expirations have different Black-Scholes implied volatilities or the “volatility smile”. In particular, traders who use the *Black-Scholes model* to hedge must continuously change the volatility assumption in order to match market prices. Their hedge ratios change accordingly in an uncontrolled way. More interestingly for us, the prices of exotic options given by models based on Black-Scholes assumptions can be wildly wrong and dealers in such options are motivated to find models which can take the volatility smile into account when pricing these.

Given the computational complexity of stochastic volatility models and their very strict difficulty of fitting the current prices vanilla options, academicians and practitioners sought a simpler model of pricing exotic options consistently with the volatility smile: the local volatility, the model has been studied first by Dupire,

Derman and Kani. In the local volatility model, we can posit a volatility that is a function of the spot process itself. Therefore, the spot process is a Markov process solution of a stochastic differential equation.

Assuming that the underlying asset price follows diffusion processes, by imposing suitable conditions on the kernel, we can obtain the nonparametric volatility function of the underlying asset-return process. The constructed volatility will be the continuous-state analog to the implied binomial tree proposed by Rubinstein (1994) and Derman and Kani (1994), and the implied volatility functions in Dupire (1994), and Dumas, Fleming and Whaley (1995). The prices obtained with the nonparametric are consistent with options pricing theory. Furthermore, the estimated volatility prevailing in our models evidences the main features of the true implied volatility and is in line generally with those in the literature for the six classes of options in which our sample is divided.

In this chapter, we compare the competing nonparametric volatility models introduced in chapter 3 in term of pricing of the European call option written on the S&P 500. To our knowledge, this is the first investigation in this direction. Although many important contributions have been done on the stochastic volatilities, authors have concentrated on parametric estimations of the volatility and then used it for the options pricing, see Hull and White (1987) and Heston (1993) among others. Santon (1997), Jiang and Knight (1997), and Renò (2008) proposed nonparametric paradigm in estimating the diffusion coefficient, but they use it to evaluate the dynamic of the spot rate of interest rate. Jiang (1998) used nonparametric estimation of the drift, volatility and market price of risk for pricing the options written on interest rates.

We have structured this chapter as follows. In Section 5.2, we borrow from several authors and present a brief review of stochastic volatility models. We present the implementation technique for computing option price using nonparametric approach in Section 5.3. In section 5.4, we recall some well-known models used to confront our proposed model. In Section 5.5 we present the implementation of numerical method, while in Section 5.6 we present empirical results and compare our results with other models available in the literature. Finally, we summarise our findings in section 5.7.

5.2 Stochastic volatility

In this section, we use the one dimension price process defined on a complete probability space, $(\Omega, \mathbb{F}, \mathbb{P})$, evolving in continuous time over the interval $[0, T]$, where T denotes a positive integer. We further consider an information filtration, i.e., an increasing family of sigma-fields, $(\mathcal{F}_t)_{t \in [0, T]} \subseteq \mathbb{F}$, which satisfies the usual conditions of \mathbb{P} -completeness and right continuity. Finally, we assume that the asset prices through time t , including the relevant state variables, are included in the information set \mathcal{F}_t .

For tackling the drawbacks of the original option pricing, several authors proposed competing models. The first popular model of stochastic volatility was proposed in Cox (1975), namely the CEV model. In that framework, the same one-dimension Brownian motion governs the stock price and the stochastic volatility of this price. Strictly speaking, it assumes that the relative stochastic volatility of the underlying asset price is state dependent, so the $\sigma_t = \sigma(S_t)$. In the model proposed by Cox, the stock price follows the corresponding dynamic under the risk neutral probability \mathbb{P}^*

$$dS_t = S_t r dt + \alpha S_t^\beta dW^* \quad (5.2.1)$$

where $\alpha > 0$, and $0 \leq \beta \leq 1$ are constants. For $\beta = 1$, (5.2.1) yields the Black and Scholes model. Cox (1975) assumed that the filtration \mathcal{F}_t is generated by the driving Brownian motion W^* . Then the CEV model is complete, so that any European contingent claim that is \mathcal{F}_t -measurable and \mathbb{P}^* -integrable possesses a unique arbitrage price given by the risk-neutral expectation

$$\pi_t(X) = B_t \mathbb{E}_{\mathbb{P}^*}(B_T^{-1}(h(S_T))/\mathcal{F}_t).$$

where h is the pay off and B is the price of the risk free asset.

The most interesting advantage of CEV model over the classical *Black and Scholes model* derives from the presence of the additional parameter β that renders the formal model flexible than the other one. By their approximate formula, Hagan and Woodward (1999a) conclude that the CEV model has an inherent flaw of incoherently predicting the future movements of the Black-Scholes implied volatility.

In the same vein, Hagan et al. (2002) examined the dynamics of the implied volatility smile. In their model, the authors analysed under the martingale measure \mathbb{P}^* the forward price which is assumed to obey the SDE

$$dF_t = \hat{\alpha}_t F_t^\beta dW_t^* \quad (5.2.2)$$

where

$$d\hat{\alpha}_t = \nu \hat{\alpha}_t d\tilde{W}_t, \quad \hat{\alpha}_0 = \alpha \quad (5.2.3)$$

where F_t is the forward value, $\hat{\alpha}$ is the volatility, ν is the volatility of the volatility and W^* and \tilde{W} are Brownian motions with respect to a common filtration \mathbb{F} , with a constant correlation ρ , the previous given model is termed by the SABR model. It is an extension of the classical CEV model. We will not focus on this type of model in this work.

We are rather interested in the following type of volatility in which we suppose that the stock price dynamic is of the form

$$dS(t) = \mu(t)S(t)dt + \sigma(t)S(t)dW_1 \quad (5.2.4)$$

has a stochastic dynamic for his own, i.e

$$d\sigma(t) = \alpha(S, \sigma, t)dt + \eta\beta(S, \sigma, t)dW_2 \quad (5.2.5)$$

with

$$\text{corr}(dW_1, dW_2) = \rho dt$$

In that setting, the new stochastic quantity that we are modeling, namely the volatility, is a nontraded asset. Thus we are confronted to the problem of having a source of randomness that cannot be trivially hedged away. Because we have two sources of randomness the option have to be hedged with two other contracts, one being the underlying asset as usual, however we need another option to hedge the volatility risk. Applying standard Ito formula, and following the standard

Black&Scholes *pdf* derivation we obtain the following equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma S^2(t)\frac{\partial^2 V}{\partial S^2} + \rho\eta\sigma\beta S(t)\frac{\partial^2 V}{\partial\sigma\partial S} + \frac{1}{2}\eta^2\sigma\beta^2\frac{\partial^2 V}{\partial\sigma^2} = -(\alpha - \varphi\beta)\frac{\partial V}{\partial\sigma} \quad (5.2.6)$$

for some function $\varphi(S, \sigma, t)$. Conventionally, $\varphi(S, \sigma, t)$ is called the *market price of volatility risk* because it tells us how much of the expected return of V is explained by the risk (i.e standard deviation) of σ in the Capital asset price framework.

This nice SDE, however don't always have a closed form solution and when the solutions are available is not easy to figure out. Moreover, given the extreme difficulty of fitting existing parameters to the current prices of vanilla options, practitioners sought a simpler way of pricing exotic options consistently with the volatility skew. Breeden and Litzenberger (1978) proved that the risk neutral *pdf* could be derived from the market price of European options. Inspired by that result, Dupire (1994) and Derman and Kani (1994) noted that under the risk-neutrality, there was a unique diffusion process consistent with this distribution. The corresponding unique state-dependent diffusion coefficient $\sigma_L(S, t)$ consistent with the European option prices is known as the *local volatility function*.

Noting that the local volatility will in general be a function of the current stock price S_0 , the price process can be written as

$$\frac{dS_t}{S_t} = \mu(t)dt + \sigma_L(S, t; S_0)dW \quad (5.2.7)$$

Before using our model, we will modified the lognormal model supposing that the volatility is a deterministic positive function of time and stock price: $\sigma = \sigma(t, X_t)$.

The stochastic differential equation modeling the asset price now is

$$dX_t = \mu X_t dt + \sigma(t, X_t)dW_t \quad (5.2.8)$$

and the function $C(t, x)$, given the no-arbitrage price of the European derivatives at time t when the corresponding asset price is $X_t = x$, satisfies the generalized

Black-Scholes partial differential equation

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 C}{\partial x^2} + rx \frac{\partial C}{\partial x} - rC = 0 \quad (5.2.9)$$

where the derivation is identical to the constant- σ case present in the seminal work of Black-Scholes. The coefficient is substituted by $\sigma(t, X_t)$. The terminal condition is the pay off function: $C(T; x) = h(x)$.

In this new model, the market is still complete since the randomness of the volatility was introduced as a function of the existing randomness of the lognormal model. There exists a unique risk-neutral martingale measure \mathbb{P}^* under which the stock price is a geometric Brownian motion with drift rate r and the same volatility $\sigma(t, X_t)$ thanks to Girsanov theorem:

$$dX_t = rX_t dt + \sigma(t, X_t) dW_t \quad (5.2.10)$$

Although the risk-neutral valuation is still possible, we know that using the non-parametric estimators, the Black and Scholes formula is no longer relevant in the sense that the new partial differential equation cannot have explicit solution any more.

Applying the Itô Lemma together with the risk neutrality, give rise to the Dupire equation which is defined as

$$\frac{\partial C}{\partial T} = \frac{\sigma_L^2 K}{2} \frac{\partial^2 C}{\partial K^2} + (r(T) - D(T)) \left(C - K \frac{\partial C}{\partial K} \right) \quad (5.2.11)$$

where $r(t)$ is the risk-free rate, $D(t)$ is a dividend yield and C is the short for $C(S_0, K, T)$ the undiscounted option prices of different strikes.

Proof. From Gatheral (2006) , □

Suppose the stock price diffuses according to the equation

$$\frac{dS}{S} = \mu_t dt + \sigma_L(S_t, t) dW. \quad (5.2.12)$$

where $\mu_t = r_t - D$ and the local volatility $\sigma_L(S, t)$

The undiscounted risk-neutral value $C(S_0, K, T)$ of a European option with strike K and expiration T is given by

$$C(S_0, K, T) = \int_K^\infty dS_T \varphi(S_T, T; S_0) (S_T - K), \quad (5.2.13)$$

where $\varphi(S_T, T; S_0)$ is the risk-neutral probability density of the spot at time T . It evolves according to the Fokker-Planck or Kolmogorov Forward equation

$$\frac{1}{2} \frac{\partial^2}{\partial S_T^2} (\sigma_L^2 S_T^2 \varphi) - \frac{\partial}{\partial S_T} (\mu S_T \varphi) = \frac{\partial \varphi}{\partial T}. \quad (5.2.14)$$

Differentiating (5.2.13) with respect to K gives

$$\frac{\partial C}{\partial K} = - \int_K^\infty dS_T \varphi(S_T, T; S_0) \quad (5.2.15)$$

and

$$\frac{\partial^2 C}{\partial K^2} = \varphi(S_T, T; S_0) \quad (5.2.16)$$

Successively differentiating with respect to T we obtain

$$\frac{\partial C}{\partial T} = \int_K^\infty dS_T \left\{ \frac{\partial}{\partial T} \varphi(S_T, T; S_0) \right\} (S_T - K) \quad (5.2.17)$$

by using (5.2.14) in (5.2.17)

$$\frac{\partial C}{\partial T} = \int_K^\infty dS_T \left\{ \frac{1}{2} \frac{\partial^2}{\partial S_T^2} (\sigma_L^2 S_T^2 \varphi) - \frac{\partial}{\partial S_T} (\mu S_T \varphi) \right\} (S_T - K) \quad (5.2.18)$$

Integrating the first part of (5.2.18) by parts, $\int_a^b f g' dx = [fg] - \int_a^b f' g dx$, gives

$$\begin{aligned} & \int_K^\infty dS_T \left\{ \frac{1}{2} \frac{\partial^2}{\partial S_T^2} (\sigma_L^2 S_T^2 \varphi) \right\} (S_T - K) = \\ & = \left[\left\{ \frac{1}{2} \frac{\partial}{\partial S_T} (\sigma_L^2 S_T \varphi) \right\} (S_T - K) \right]_K^\infty - \int_K^\infty dS_T \frac{1}{2} \frac{\partial}{\partial S_T} (\sigma_L^2 S_T^2 \varphi) \end{aligned} \quad (5.2.19)$$

using the fact that $\lim_{K \rightarrow \infty} S_T = 0$ the term in the braces vanishes and the second

term becomes $\frac{\sigma_L^2 K^2}{2} \varphi$. Integrating the second term of (5.2.18) by part yields

$$\begin{aligned}
 & - \int_K^\infty dS_T \left\{ \frac{1}{2} \frac{\partial}{\partial S_T} (\mu S_T \varphi) \right\} (S_T - K) = \\
 & \qquad \qquad \qquad [-\mu S_T \varphi (S_T - K)]_K^\infty - \int_K^\infty dS_T (\mu S_T \varphi)
 \end{aligned} \tag{5.2.20}$$

following the same method as before, the term in the braces vanishes. This leads to the following equation

$$\frac{\partial C}{\partial T} = \frac{\sigma_L^2 K^2}{2} \varphi + \int_K^\infty dS_T \mu S_T \varphi \tag{5.2.21}$$

The second equation can be written as

$$\int_K^\infty dS_T \mu S_T \varphi = \mu \left[\int_K^\infty dS_T \varphi (S_T - K) + K \int_K^\infty dS_T \varphi \right] \tag{5.2.22}$$

We can remark that the first term on the right hand side above is exactly the undiscounted option value from (5.2.13). By using (5.2.15) to the second term of (5.2.22) and (5.2.16) to the first term of (5.2.21) we finally obtain the (5.2.11).

When expressing the option price as a function of forward price $F_T = S_0 \exp\{\int_0^T \mu(t) dt\}$, we would get the same expression except for the drift. That is

$$\frac{\partial C}{\partial T} = \frac{\sigma^2 K}{2} \frac{\partial^2 C}{\partial K^2}$$

where C now represents $C(F_T, K, T)$. Inverting this gives

$$\sigma^2(K, T, S_0) = \frac{\frac{\partial C}{\partial T}}{\frac{1}{2} K^2 \frac{\partial^2 C}{\partial K^2}} \tag{5.2.23}$$

The right hand side of equation (5.2.23) can be computed from known European option prices.

5.3 Implementation method

The nonparametric estimate of the diffusion coefficient can be seen as a local volatility since the estimated volatilities are constructed on the local trajectories of the underlying process. In this section we will use all the nonparametric coefficient presented in the previous chapter to derive the corresponding price of European call option.

In this section we try to face up to Dupire problems contained in the Dupire approach. As we introduce previously, suppose that the dynamic of an underlying asset has the following equation,

$$\frac{dS_t}{S_t} = \mu_t dt + \sigma(t, S_t) dW$$

where the local volatility function is parametrized as $\sigma : [0, T] \times]0, \infty[\rightarrow]0, \sigma_{max}[$, and the function $C(T, K)$, given the no-arbitrage price of the European derivatives at time T when the corresponding asset price is $X_t = K$, the strike price, then satisfies the generalized Dupire partial differential equation

$$\begin{cases} \frac{\partial C}{\partial T} = \frac{1}{2} \sigma^2(T, K) K^2 \frac{\partial^2 C}{\partial K^2} + (r - D)(C - K \frac{\partial C}{\partial K}) & \forall (T, K) \in]0, T[\times \mathbb{R}_+^* \\ C(0, K) = (S - K)^+ & \forall K \in \mathbb{R}_+^* \end{cases} \quad (5.3.1)$$

where the derivation has been explicitly discussed in the previous section. The initial condition is the pay off function: $C(\tau, K) = h(K)$, with $\tau = T$. Theoretically, we can use this equation to find out the local volatility $\sigma(., .)$ from a collection of option prices $(T, K) \rightarrow C(T, K)$ observed for a continuum of values of (T, K) : this is the Dupire formula already mentioned in the previous section:

$$\sigma(T, K) = \sqrt{\frac{2(C_T + rKC_K)}{K^2 C_{KK}}}$$

and this volatility is unique as Dupire assumes that there is a bijection between the call price $C(T_i, K_i)$ and the local volatility $\sigma(T_i, K_i)$.

Despite the fact that the theory ensures that there exists a unique local volatility, it is non-trivial problem to recover it from real option data. This drawback derives from the assumption of a well defined European option price space, which is not the case on real markets. In practice, only a finite number of options $[C^l(T_i, K_j)]_{i,j}$ is available for different maturities T_i and strike K_j . This renders the problem strongly underdeterminate. For being close to the theory, practitioners resort to smoothness procedures or interpolation techniques of the implied volatility surface in order to obtain a continued and smoothed collection $\hat{C}(T, K)$ to which the Dupire formula will be applied.

The set of price $\hat{C}(T, K)$ on which the estimation of the volatility is based may not be a perfect observation of the market price, but a reconstruction (non uniqueness). Therefore, even in the theoretical case where the observed option prices come from a diffusion model, the reconstituted collection $\hat{C}(T, K)$ can differ from the theoretical price $C(T, K)$ obtained from the diffusion model. The difference between the reconstructed collection price $\hat{C}(T, K)$ using smoothness or interpolation can have a complex dependence in (T, K) . We will not focus on that aspect here.

An alternative way to determine the local volatility can be reached by following DK approach (Derman and Kani, 1994). Although the two approaches are conceptually tantamount, the method presented in this section allows estimation of rather much smoother local volatility functions. From a numerical standpoint, these two approaches are further different from our nonparametric in the sense that we are not imposing the recombination of a computational tree, nor constrained to evaluate numerically the differential of the prices of the option $C_{K,T}(t, S)$.

Whatever the method we use to reconstruct the local volatility function, it is not possible to recover σ but its estimator $\tilde{\sigma}(\cdot, \cdot)$ which will be in the best case close to $\sigma(\cdot, \cdot)$. If we denote the bias as $\varepsilon(\cdot, \cdot)$ which depends on expiration T and strike K , the observation error or the error of the reconstruction of the collection of option prices can be then defined as

$$\tilde{\sigma}(t, S) = \sigma(t, S) + \varepsilon(t, S).$$

The estimator $\tilde{\sigma}(\cdot, \cdot)$ of the local volatility function is then used into equation (5.3.1) to evaluate every type of options and the derivatives on other financial instruments. In Section 5.6 we use our competing nonparametric estimators to determine European call price of options written on the S&P 500, however we are interested in measuring our nonparametric volatility estimators in term of their capability of deriving option prices that is close to the option market prices.

5.4 Competitive methods

In order to assess the performance of our model in option pricing we will compare it with popular well-known models; The original Black-Scholes model (B&S) (1973), Merton's jump diffusion model (M) (1976) and the Heston's stochastic volatility model (H) (1993).

In the B&S model the underlying stock price is assumed to have the dynamic of the geometric Brownian motion diffusion process of the form

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

where σ and μ are volatility (constant) and the drift of the underlying asset.

In the Merton model the stock price underlying is assumed to follow a jump diffusion process

$$dS_t = (\mu - \lambda_j k_j) S_t dt + \sigma S_t dW_t + (J_t - 1) S_t dq$$

where q is a Poisson process uncorrelated with W and the intensity λ_j , which is the rate at which jumps occur, J_t is proportional increase in the stock price at time t and $k_j = E(J_t - 1)$ stands for the average jump size. The closed-form solution for the price of European call exists in the special case that the logarithmic of J_t is Normally distributed, with standard deviation δ_j ,

$$C_M = \sum_{n=0}^{\infty} \frac{e^{-\lambda' \tau} (\lambda' \tau)^n}{n!} C_{BS}(S, \tau; \sigma_n, r_n) \quad (5.4.1)$$

where $\lambda' = \lambda_j(1 + k_j)$, $\sigma_n^2 = \sigma_j^2 + n\delta_j^2/\tau$ and $r_n = r - \lambda_j k_j + n\log(1 + k_j)/\tau$

In the end, in the Heston model it is assumed that the spot price at time t follows the diffusion

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_t^1$$

where W^1 is a Wiener process. The volatility V_t follows an Ornstein-Uhlenbeck process

$$d\sqrt{V_t} = -\gamma\sigma(t)dt + \delta(t)dW_t^2$$

and W^2 is another Wiener process such that W^1 and W^2 are correlated with correlation ρ . Let $x = \sqrt{V_t}$ and apply the Ito's formula in $f(x) = x^2$. The result is

$$dV_t = [\delta^2 - 2\gamma V_t]dt + 2\delta\sqrt{V_t}dW_t^2$$

Then if we let, $k = 2\gamma$, $\theta = \frac{\delta^2}{2\gamma}$, and $\sigma = 2\delta$ we end up with the Heston model where

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_t^1 \tag{5.4.2}$$

$$dV_t = k[\theta - V_t]dt + \sigma\sqrt{V_t}dW_t^2 \tag{5.4.3}$$

$$\text{corr}(dW^1, dW^2) = \rho dt \tag{5.4.4}$$

V is the implied spot variance of the returns, k is the mean-reversion speed, θ is the long-run variance and σ is the volatility parameter of the diffusion volatility V_t . If $\theta < V$ then the process decreases until the volatility goes under the theta parameter. Then it goes up again and so on. A closed form solution for this model is available in Heston (1993) and can be implemented by numerical integration of the characteristic function.

5.5 Implementation of numerical method

In this section we use the simplified Dupire equation presented in the previous section where we suppose that the dividend yield $D(T)$ and the risk-free rate $r(T)$ are zero, namely we consider future prices. The same consideration can be used in the original Black & Scholes equation, but for computational convenience, we have chosen the Dupire equation. Given the above restriction, the Dupire equation can be written as

$$\frac{\sigma^2(K, T)}{2} \frac{\partial^2 C}{\partial K^2} = \frac{\partial C}{\partial T} \quad (5.5.1)$$

where $C(S, t, K, T)$ denotes the premium at time t for a spot S of an European call of strike K and maturity T . Both derivatives should be positive to avoid arbitrage. We will use equation (5.5.1) to determine the option prices using our nonparametric estimates for $\sigma^2(K, T)$. One advantage of this forward equation-type is that all cross-section option series with the same maturity can be valued instantaneously. A different approach is to solve the Black & Scholes PDE as many time as options of different exercise price are; for the two approaches we obtain the same results. But the second approach will be much time consuming.

The new equation can be interpreted in another way. If $\sigma^2(K, T)$ is known, it establishes a relationship between the price as of today of call options of varying maturities and strikes.

Equation (5.5.1) can be seen as the opposed of the classical Black & Scholes partial differential equation which involves, for a fixed option (K and T fixed), derivatives with respect to the current time and value of the spot price. This happens if we set the interest rate equal to zero in the Black & Scholes equation, therefore we retrieve the following:

$$-\frac{\sigma^2(S, t)}{2} \frac{\partial^2 C}{\partial S^2} = \frac{\partial C}{\partial t} \quad (5.5.2)$$

Equation (5.5.1) and (5.5.2) can be thought as operating in the same space of functionals (dual). However, the relationship is not always true, as (5.5.2) applies to any contingent claim, though (5.5.1) holds because the intrinsic value of a call happens to be the second integral of a Dirac function.

We know from standard results that the arbitrage-free price for a contingent claim according to the Dupire approach is given by the following formula

$$C(F, t, K, T) = \int_K^{+\infty} (F_T - K)^+ \varphi(F_T) dF_T \quad (5.5.3)$$

Where $r = 0$, $F = F_t$ is the future price.

Following the Fokker-Planck equation our problem can be reduced to the following:

$$\begin{cases} \frac{\partial C}{\partial \tau} - \frac{\sigma^2(K, \tau)}{2} \frac{\partial^2 C}{\partial K^2} = 0, & \tau \geq 0 \\ C(K, 0) = (F - K)^+ \end{cases} \quad (5.5.4)$$

$$\tau = T - t \text{ with } t = 0$$

this problem is parabolic equation with initial condition $C(K, 0)$ which can be interpreted as the price today of a call with strike K and immediate maturity.

5.5.1 Discretization method

We now describe the numerical procedure we have used to compute the approximated solution of the previous PDE.

In order to discretize the Dupire equation, we should work on a bounded open set $\mathcal{O} = (K_{min}, K_{max})$ where K is a constant to be chosen carefully in order to reduce the approximation error of the algorithm. We also need to specify the boundary conditions. In our case, we shall impose the linearity conditions. These conditions are suitable for instance when a finite difference solution is required (e.g., when there are discrete dividends, when we want to use local volatility surface $\sigma(S, t)$ in our model, or when the strike price resets periodically), and simple, yet consistent boundary conditions are strong to define. It is found to be that, however, in a large number of option structures far from the strike price, or other such 'interesting' regions, the option value is nearly linear with respect to the spot. This observation is true for many exotic and path-dependent options, and plain vanilla options.

Once the problem has been localized, we restrict K to belong to the interval

$[K_{min}; K_{max}]$ and τ varying in the interval $[0; T]$, we successively obtain the semi-discretized equation

$$\begin{cases} \frac{\partial C}{\partial \tau} - \frac{\sigma^2(K, \tau)}{2} \frac{\partial^2 C}{\partial K^2} = 0, & \text{in }]K_{min}; K_{max}[\times]0; T[, \\ \frac{\partial^2 C}{\partial K^2}(K_{min}, \tau) = 0, & \text{if } \tau \in]0; T[, \\ \frac{\partial^2 C}{\partial K^2}(K_{max}; \tau) = 0, & \text{if } \tau \in]0; T[, \\ C(K, 0) = f(K) = \max(0, S_0 - K) & \text{for } K \in]K_{min}; K_{max}[. \end{cases} \quad (5.5.5)$$

Let $N = \frac{(K_{max} - K_{min})}{dx}$ the number of elements of the grid and h a number to be chosen carefully which is the step of the discretization in space, we set for i varying from 0 to N

$$K_i = K_{min} + ih$$

$$f_i = f(K_i)$$

Consider the vector $c_h = (c_h^i)_{1 \leq i \leq N}$ in \mathbb{R}^N as the discrete approximation of a function C (i.e. $c_h^i \approx C(K_i)$) and replace the first derivative $\frac{\partial C(K_i)}{\partial K}$ with

$$\partial_h c_h^i = \frac{c_h^{i+1} - c_h^i}{h}.$$

Similarly, replace the second derivative $\frac{\partial^2 C(K_i)}{\partial K^2}$ with

$$\partial_h^2 c_h^i = \frac{c_h^{i+1} - 2c_h^i + c_h^{i-1}}{h^2}.$$

Set the vector $\tilde{A}_h c_h = \frac{\hat{\sigma}^2(K_i)}{2} \frac{c_h^{i+1} - 2c_h^i + c_h^{i-1}}{h^2}$

We set

$$\alpha_i = \frac{\hat{\sigma}^2(K_i)}{2h^2}$$

$$\beta_i = \frac{\hat{\sigma}^2(K_i)}{h^2}$$

$$\lambda_i = \frac{\hat{\sigma}^2(K_i)}{2h^2}$$

We have then, for every $i \in \{1, \dots, N-1\}$ and for every τ :

$$\left(\tilde{A}_h c_h(\tau) \right)_i = \alpha_i c_h^{i-1}(\tau) + \beta_i c_h^i(\tau) + \lambda_i c_h^{i+1}(\tau)$$

when $i = 0$, we have for whatever T a Dirichlet type boundary conditions $c_0(\tau) = S_0 e^{-r(\tau-t_0)}$ and therefore, the first equation for $i = 1$

$$\left(\tilde{A}_h c_h(\tau) \right)_1 = \alpha_1 c_h^0(\tau) + \beta_1 c_h^1(\tau) + \lambda_1 c_h^2(\tau).$$

when $i = N$ ($K = K_{max}$), we have for every τ the linearity boundary condition $c_h^N(\tau) = 0$, whereof, the last equation for $i = N-1$

$$\left(\tilde{A}_h c_h(\tau) \right)_N = \alpha_N c_h^{N-2}(\tau) + \beta_N c_h^{N-1}(\tau) + \lambda_N c_h^N(\tau).$$

under the specified linearity conditions, \tilde{A}_h is an intermediate operator of \mathbb{R}^{N-1} which is represented by the following matrix:

$$\left((\tilde{A}_h)_{ij} \right)_{1 \leq i, j \leq N} = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & \beta_1 & \lambda_1 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & \alpha_2 & \beta_2 & \lambda_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \alpha_3 & \beta_3 & \lambda_3 & \cdots & \vdots & 0 \\ \vdots & 0 & \vdots & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & 0 & \cdots & \alpha_{N-1} & \beta_{N-2} & \lambda_{N-2} & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \alpha_{N-1} & \beta_{N-1} & 0 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}$$

The space discretization conduct to the ordinary equation

$$\begin{cases} \frac{dc_h(\tau)}{d\tau} - \tilde{A}_h c_h(\tau) = 0 & \text{if } \tau \in]0, T] \\ c(\tau_0) = f \end{cases} \quad (5.5.6)$$

where $f = (f(K_i))_{1 \leq i \leq N-1}$.

Using the θ -schemes, the time discretization of this equation can be defined. Suppose that $\theta \in [0, 1]$, and let k be the time-step such that $T = Mk$. We approximate the solution c_h at time nk by the c_i^n , where the sequence $(c_i^n)_{n=0, \dots, M}$ are the solution of the recursive equation:

$$\begin{cases} c^0 = f \\ \frac{c_i^{n+1} - c_i^n}{k} + \theta \tilde{A}^n c_i^n + (1 - \theta) \tilde{A}^{n+1} c_i^{n+1} = 0 \text{ if } 0 \leq n \leq M - 1 \end{cases} \quad (5.5.7)$$

where $\tilde{A}^n = \tilde{A}_h(\tau + nk)$

the previous system is solved by forward induction.

We can obtain two different schemes type according to the value of θ .

- if $\theta = 1$, the scheme is explicit,
- if $0 \leq \theta \leq 1$, the scheme is implicit.

Now we have to resolve a linear system of the form

$$M^n c^{n+1} = q^n \quad (5.5.8)$$

where

$$\begin{aligned} M &= I + \frac{\theta \Delta \tau}{\Delta K^2} \tilde{A}_h \\ q &= I - \frac{(1-\theta) \Delta \tau}{\Delta K^2} \tilde{A}_h \end{aligned}$$

with M an $(N - 1, N - 1)$ matrix being tridiagonal for any n . To solve this system, we can triangularize it at every time-step using the pivoting method.

Up to now, we have not mentioned the estimation of the drift coefficient in this dissertation. The principale raison being that it is not influenced in evaluating the option prices when using equation (5.5.1). Nevertheless, Jiang and Knight (1998) demonstrated that with additional assumptions, the drift function can be estimated as

$$\hat{\mu}(x) = \frac{1}{2} \left[\frac{d\hat{\sigma}^2(x)}{dx} + \hat{\sigma}^2(x) \frac{p'(x)}{p(x)} \right]$$

by using the information contained in the marginal density function $p(x)$, along with the estimated $\hat{\sigma}^2(x)$.

5.6 Empirical findings

In this section, we concentrate on the comparative analysis of all the nonparametric volatility estimators introduced in the previous chapter and successively, we make the comparison of our method with the popular well-known methods available in the literature, namely the Black & Scholes, the Jump-Diffusion model and the Heston model. All our empirical results will be based on the S&P 500 data set.

5.6.1 Data description and characteristics

We focus on the options of the S&P 500 index, which are the most actively traded European-styled contracts and have the daily dividend distribution of the index available (from the S&P 500 Bulletin). The nice characteristics of this index have permitted to many researchers to test their theoretical work on it. Therefore, S&P 500 options and options on S&P 500 futures have been analysed by Bates (1998), Dumas *et al*, Rubinstein (1994), Sanfelici (2007). These considerations provide us a motivation to apply on it our nonparametric estimators and successively we will compare our results to the S&P 500 call options obtained from the CBOE (Chicago Board Option Exchange). The analysed sample spans from the period January 4, 1993 to December 31, 1993 (253)

Table 5.1 describes the main characteristics of our data set. During the one year considered period the variation exhibited by short-term interest rates is small in size: they range from 2.85 to 3.21%. The options in our sample vary considerably in price and terms; for instance, the time-to-maturity vary from 1 to 350 days, with a median of 66 days.

The average total daily volume during the considered period was 65476 contracts. Following the CBOE practice, the expiration months are the three near term months followed by three additional months from the March quarterly cycle (March, June, September, December). The options are European, they expire on the third Friday of the month and the underlying asset is an index, the most likely case for which a lognormal assumption (with continuous dividend stream) can be justified. By the simple effect of diversification, jumps are less likely to occur in the index than in the

Variable	Mean	Std.dev	Min	Max
Call price, $C(\$)$	18.29	17.32	0.0249	68.60
Implied volatility (%)	11.01	2.98	5.07	36.83
τ (days)	79.96	66.56	1	350
K (index points)	445.52	27.48	375	550
F (index points)	454.87	10.21	429.18	474.21
r (%)	3.06	0.08	2.85	3.21

Table 5.1: Summary statistics for the sample of trade CBOE daily call option prices on the S&P 500 index in the period January 4,1993 to December 31,1993 (13 078). τ denotes the times to maturity, r the riskless rate, K the strike price and F the S&P 500 futures value implied from the call and put prices. 'Std.dev.' denotes the sample standard deviation of the variable. During this period, the average daily value of the variable of the S&P 500 index was 451.66

individual equities. This feature allows us to think that the market is as close as the theoretical assumptions underlying the Black & Scholes model.

The beginning sample contains 16 963 call options, we take the average of bid and ask price as our raw data. We consider observations with the time-to-maturity longer or equal to one day. The implied volatility greater than 70% and price less than 0.02 or greater than 70.00 were cancelled out. After that, we remain with a final sample of 13078 observations.

When applying our nonparametric approach to the raw data, we have to face three important problems. Firstly, in-the-money options are very infrequently traded with respect to at-the-money and out-the-money options, The former is notoriously unreliable. There is an unbalance daily volume for out-of-the-money contracts and the volume of in-the-money contract of the same magnitude.

Secondly, the underlying index price is very hard to observe at the exact times that the options prices are recorded. Especially, there is not guarantee that the closing index value is reported at the same time as the closing transaction for each option. S&P 500 index are traded on the Chicago Mercatile Exchange(CME), not on the CBOE and the time-stamped recorded quotes may not necessarily be perfectly synchronized across the two markets. Mismatch of little size can lead to economically significant but spurious pricing anomalies.

Thirdly, the index typically pays dividend and the future rate of dividend payment is difficult, if not impossible, to determine. The daily dividend provided by Standard and Poor's on the S&P 500 is by nature forward-looking, and there is no reason to assume that the actual dividends recorded ex post correctly reflect the expected future dividend at the time the options is priced.

We can tackle this uncomfortable drawback following Ait-sahalia and Lo (1998) and Sanfelici (2007). Based on the fact that all options are recorded at the same time on each day, we require only the temporally matched index price per day. To get around the unobservability of the dividend rate $\delta_{t,\tau}$, we deduce the futures price $F_{t,\tau}$ and S_t for each maturity τ . By the spot-futures parity, $F_{t,\tau}$ are linked through the relationship

$$F_{t,\tau} = S_t e^{(r_{t,\tau} - \delta_{t,\tau})\tau}. \quad (5.6.1)$$

To derive the implied futures, we resort to the put-call parity relation, which must be satisfied if arbitrage opportunity are to be eliminated, independently of any parametric option-pricing model.

$$C(S_t, K, \tau, r_{t,\tau}, \delta_{t,\tau}) + Ke^{-r_{t,\tau}\tau} = P(S_t, K, \tau, r_{t,\tau}, \delta_{t,\tau}) + F_{t,\tau}e^{-r_{t,\tau}\tau} \quad (5.6.2)$$

where C and P denote respectively the call and the put price of actively traded options with the same strike K and the time-to-expiration τ . From this expression, we require reliable call and put prices at the same strike price K and time-to-expiration τ . For this purpose, we must use calls and puts that are closest to at-the-money. It is well known that in-the money options are illiquid relative to the out-of-the-money counterparts, hence any matched pair that is not at-the-money would have one potentially unreliable price.

We use the formula introduced in the previous chapter to derive our nonparametric local volatility function. Successively we evaluate the different approach by substituting each estimate in the equation (5.5.1) for pricing European call.

We divide our data set into several categories as in Baskhi *et al.* (1997) and Sanfelici (2007). Following these authors the division criteria is according to moneyness S_t/K or the time-to-expiration τ . We say that a call option is *at-the-money* (ATM)

if $0.98 < S_t/K \leq 1.02$, *out-of-the money* (OTM) if $S_t/K \leq 0.98$ and *in-the-money* (ITM) if $S_t/K > 1.02$. A finer repartition resulted in six moneyness categories. When classifying our data by term of expiration, we say an option contract has a *short term maturity* if $\tau < 60$ days, *medium term maturity* if $60 \leq \tau \leq 180$ days and *long term maturity* if $\tau > 180$ days. From empirical data, the proposed moneyness and maturity classification produce 18 categories which are reported in the Table 5.2 . ITM and ATM options account for respectively 50% and 24% of the total sample, while the short-term and the medium term take up respectively around 47% and 43%. The average price varies from \$0.19 for short-term deep OTM options to \$47, 29 for long-term deep ITM calls. For empirical study, each class can be used since the options are partitioned quite uniformly in the different moneyness categories and included nearly 1500 observations each and every.

Many approaches have been proposed to extract local volatility function: Derman and Kani (1994), and Rubinstein (1994) proposed the non arbitrage binomial or trinomial tree model where the volatility function is obtained at the end of each node inverting the corresponding call price. In the continuous time approach proposed by Dupire, B (1994), it is retrieved by means of the equation (5.5.1), where it is assumed to know the prices of options of all strikes and maturities via the implied volatility surface. Loosely speaking, we know the quantity $C_{K,T}(0, S)$ as a function of K and T , and therefore it is quite immediate to evaluate (numerically) the derivatives of the observed option prices with respect to the maturity and the strike price.

Our approach is completely different, in the sense that the volatility is estimated from the underlying index dynamics and does not rely on option prices, we construct our nonparametric estimator directly on the S&P 500 data using the Nadayara-Watson formula which has been explained in the previous chapter, we derive the nonparametric estimator by substituting in the latter the corresponding spot volatility estimators reported in the Table 4.1 (chapter 4). It is well known that index price stock are nonstationary, this particular feature of stock options renders the nonparametric estimation very challenging. Therefore we need to choose cleverly the parameter used to construct our local volatility function by means of our nonparametric estimator. We can assume that what we observe in the finite interval

<i>Moneyness</i>					
<i>S/K</i>		<i><60</i>	<i>60-180</i>	<i>>180</i>	<i>Subtotal</i>
OTM	<0,96	0.19	0.81	2.35	1.01
		478	979	436	1893
	0.96-0.98	0.78	2.90	10.43	2.42
		738	640	114	1492
ATM	0,98-1,00	2.66	6.52	13.50	4.66
		861	647	58	1566
	1.00-1.02	7.56	11.78	17.75	9.66
		874	656	56	1586
ITM	1.02-1,04	14.64	18.09	25.21	16.99
		788	611	143	1542
	>1.04	33.46	39.01	47.29	36.97
		2416	2192	391	4999
Subtotal		16.56	19.42	21.78	18.29
		6155	5725	1198	13078

Table 5.2: The reported numbers are, respectively, the average quoted bid-ask mid point price and the total of observations for each moneyness-maturity category. S denotes the spot S&P 500 index level and K is the exercise price.

$[0, T]$ is a part of stationary process. Moreover, the other correction is to choose the bandwidth, we find that the smoothness parameter that provide acceptable result is contained in the interval $17 \leq h \leq 25$, out of this band we obtain unsatisfying results. An higher value of h will lead to a smoother estimate of the density in the tails of the distribution where fewer data are available. We finally maintain $h = 20$ because is the value of the bandwidth which give appealing outcomes when applying on every our six classes. Furthermore, if the kernel function is locally lipschitz, it can be applied to the nonstationary data. We have used tick-by-tick data of S&P 500 from which we have extracted all price with observation frequency one minute in order to mitigate the effect of microstructure noise. Futhermore, the theory assumes that when the distance between two observations is close to zero, an arbitrary precision in the estimate of the spot volatility can be reached. Broadly speaking, these facts contribute to the improvement of our local volatility function. The estimator $\tilde{\sigma}(K, \tau)$ evidences a strong volatility smile in general. We will turn on this aspect later.

We want to check if our nonparametric estimators of volatility verify the empirical facts exhibited by implied volatility when applied in the stock data. The Table 5.3 summarizes the main features of the implied volatility, which is obtained by inverting the Black & Scholes formula from each option price in our sample and successively produce an average 'implied' Black volatility for each moneyness-maturity category. The results show that the average B-S implied volatility tend to decrease monotonically as the call option move from deep ITM to ATM and then this variation is strong for the short-term options, which evidences a slight smile; this suggests that the short-term options are proned to severe mispricing. From the smile evidence, we can observe a negatively skewed implicit return distributions with excess kurtosis. Therefore, any acceptable model intended to price options written on S&P 500 have to be consistent with these features.

We describe the procedure used to estimate the spot volatility and the other structural parameters of the alternative models. A well consolidated practice (Bakshi et al. 1997, Sanfelici 2007) is to compute option-implied parameters by implementing each model in the two steps as explained below.

1. Collect m option prices taken from the same point in time t (or same day) for any m greater than or equal to one plus the number of parameters to be estimated. For each $j = 1, \dots, m$, let τ_j and K_j be respectively the time to expiration and the strike price of the j th option; let $C_j(t; \tau_j, K_j)$ be its observed price and $\hat{C}_j(t; \tau_j, K_j)$ its price worked out by the model with S_t and $r_{t,\tau}$ taken from the market. The difference $C_j - \hat{C}_j$ is a function of the values taken by $\Phi = \{\sigma\}$ in the B-S model, by $\Phi = \{\sigma_j, \lambda_j, \delta_j, k_j\}$ in the J-D model and $\Phi = \{\sqrt{V}, \sigma_\nu, k_j, \theta_\nu, \rho\}$ in the Heston model. For each j , we define

$$\epsilon_j(\Phi) = C_j(t; \tau_j, K_j) - \hat{C}_j(t; \tau_j, K_j).$$

2. Find the parameter vector Φ to solve the nonlinear least-squares problem at time t

$$SSE(t) = \min_{\Phi} \sum_{j=1}^m [\epsilon_j(\Phi)]^2 \quad (5.6.3)$$

Go back to Step 1 until the two steps have been repeated for each day in the sample.

The objective function (5.6.3) is defined as the sum of squared pricing error and may force the estimation to assign more weight to relative expensive options (e.g., ITM and long-term options).

In the calibration procedure, when the $\frac{F_{t,\tau_j}}{K_j}$ does not belong to the grid G , the value C_j is approximated by the quadratic interpolation of the nearest grid points.

5.6.2 Model Comparison Results

Turn to the behaviour exhibited by our nonparametric approach, the noteworthy fact is that for all the spot volatilities used in our model, the relative local volatility function presents a monotonically decreased trend from deep in the money and to the deep out-of-the-money, including at-the-money options. This is consistent with the empirical fact underlying the options written on stock index and it is in line with other methods existing in the literature. If we focus our analysis on all options

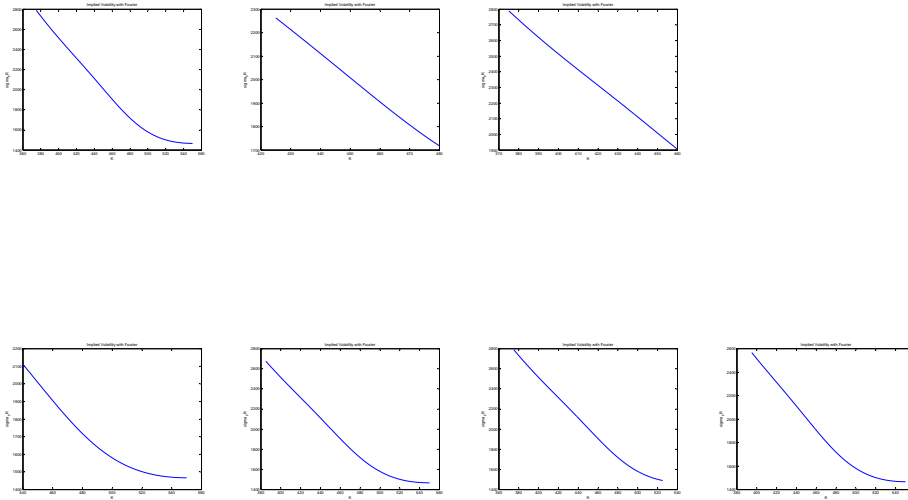


Figure 5.6.1: Local volatility Curves for Nonparametric Estimator with Fourier spot volatility. The curve relative to all classes are reported

available, the trend of any estimator is almost the same, albeit the Fourier estimator tend to generate the most highest skew with respect to the striking price when confronting with other estimators. We can notice in particular that our estimate smile is exaggeratedly asymmetric, pointing to the empirical fact that the European call options written on the stock index loss their U-shape after the 1987 crash. This particular characteristic inherent of our models is consistent with empirical findings and contrasts stochastic volatility models which typically produce symmetric smiles. This is an other example among others that nonparametric approaches can suggest an information for constructing parametric models, if not to valid them. Much evidence can be found in Figure 5.6.1

Table 5.4 summarises the prices of call options on our considered sample computed using the nonparametric models. When performing the simulation we use the discretization techniques exposed below, we assume the interest rate equal to zero,

Moneyness S/K	Time to maturity			Total	
	< 60	60-180	> 180		
OTM	<0.96	8.51	8.00	8.53	8.25
	0.96-0.98	7.98	8.53	9.91	8.36
ATM	0.98-1.00	8.57	9.21	10.03	8.89
	1.00-1.02	9.80	9.98	10.45	9.90
ITM	1.02-1.04	11.43	10.80	11.04	11.15
	>1.04	11.78	13.02	12.49	13.83
Total		11.47	10.64	10.41	11.01

Table 5.3: Average B-S implied volatilities for different moneyness maturity categories. S denotes the spot S&P 500 index level and K is the exercise price.

this does not harm very much our results. We start our analysis focusing on the whole dataset available in our sample. The value of the call options is almost of the same range for each maturity. A noteworthy fact is that the price figured out by the Florens-Zmirou estimator and the Comte and Renault estimator are almost the same. This behaviour can be found also when analysing the ATM option call prices. The ITM options call prices are the same for all maturity, the value is around \$49. Not surprisingly, our prices are consistent with the features of actual market prices and the theoretical assumption of no-arbitrage underlying options pricing: that is ITM options price is higher with respect to ATM options price which are in their turn more expensive than OTM which are the cheapest in the market.

Table 5.5 reports the daily average absolute (APE), the percentage averaged error (PPE) and the daily averaged mispricing index (MISP)

$$APE = \frac{\sum_{i=1}^N |C_i - \hat{C}_i|}{N} \quad (5.6.4)$$

$$PPE = \frac{\sum_{i=1}^N |C_i - \hat{C}_i|}{N * C_i} \quad (5.6.5)$$

Nonparametric options price					
Maturity					
Categories	First	Second	Third	Four	Fifth
All	24.357	24.812	25.267	26.745	28.191
	24.358	24.813	25.269	26.750	28.197
	24.258	24.580	24.903	25.952	26.997
	24.336	24.763	25.190	26.579	27.943
	24.318	24.721	25.126	26.445	27.748
	24.328	24.741	25.166	26.745	27.448
ATM	16.739	17.713	18.445	19.821	29.387
	16.741	17.716	18.448	19.823	20.388
	16.479	17.255	17.873	19.232	19.957
	16.686	17.622	18.336	19.719	20.321
	16.642	17.552	18.255	19.648	20.274
	16.354	17.675	18.654	19.256	20.345
ITM	49.064	49.064	49.064	49.064	49.064
	49.064	49.064	49.064	49.064	49.064
	49.064	49.064	49.064	49.064	49.064
	49.064	49.064	49.064	49.064	49.064
	49.064	49.064	49.064	49.064	49.064
	49.064	49.064	49.064	49.064	49.064
OTM	3.792	4.462	5.053	6.596	7.789
	3.793	4.464	5.057	6.603	7.789
	3.636	4.131	4.586	5.825	6.822
	3.759	4.394	4.959	6.444	7.602
	3.675	4.321	4.487	6.555	7.802
	3.731	4.333	4.870	6.287	7.396

Table 5.4: Estimated nonparametric call options on the S&P 500 index for different moneyness. The six elements of each cell are from top to bottom the price obtained with the following estimator: Florens-Zmirou, Comte and Renault, Rolling Extimator, Rolling Exponential, Ogawa & Sanfelici and the Fourier

$$MISP = \frac{\sum_{i=1}^N (C_i - \hat{C}_i)/C_i}{\sum_{i=1}^N |(C_i - \hat{C}_i)/C_i|} \quad (5.6.6)$$

where C_i is the observed call price available on the market, \hat{C}_i is the call price worked out using nonparametric techniques, and N is the sample dimension. The mispricing index ranges from -1 and 1 and indicates on average, the overpricing (when it is negative) and the underpricing (when it is positive) induced by the model. In general, our model tends to overprice the options irrespectively on the class where it is applied. When the model underprices options, the bias is higher in magnitude. The APE index tends to be high when applied to all data available. The ATM and OTM classes have the lowest APE, but when moves to ITM class this index increases considerably, this is probably due to the fact that ITM options are illiquid, viz difficult to price. The PPE index presents a very alternated behaviour. When we confront our result using MISP index, we notice that, our estimators tend to overprice call price in general in all the analysed classes. But things change when we analyse the ITM class. In this case the MISP is positive for all the estimators, suggesting that this loss function underprices in-the-money options.

To gain a better insight into the different nonparametric models performance, we apply our model each time using one of the six alternative sets from the whole sample: ITM, ATM, OTM, short-term, meduim-term, long-term categories. The error tend to be small in the whole sample with respect to the other categories in terms of APE and MISP, while being high relative to PPE. When ATM options are priced, the resulting estimates do not significantly differ from their counterparts for the whole data set. OTM call options are associated with relative low pricing error, while ITM options correspond to higher mispricing error, indicating that, for the illiquid ITM calls to be priced properly, the volatility of the underlying needs to be higher than for all options of any maturity to be priced.

When analysing with respect to time to expiration, short-term option seem to be more challenging, they are associated with the highest pricing error. However, the

	Florens-Zmirou	All options	ITM	ATM	OTM	SHORT	MEDIUM	LONG
APE		0.69	1.34	0.93	0.59	0.44	0.68	0.83
PPE		0.40	0.04	0.19	1.06	0.44	0.43	0.24
MISP		-0.82	0.83	0.41	-0.52	-0.90	-0.88	-0.39
	Comte-Renault	All options	ITM	ATM	OTM	SHORT	MEDIUM	LONG
APE		0.69	1.34	0.93	0.59	0.44	0.69	0.83
PPE		0.40	0.04	0.19	1.07	0.44	0.43	0.24
MISP		-0.82	0.83	-0.41	-0.53	-0.90	-0.88	-0.39
	Rolling Estimato	All options	ITM	ATM	OTM	SHORT	MEDIUM	LONG
APE		0.86	1.54	1.08	0.53	0.38	0.84	1.33
PPE		0.15	0.05	0.14	0.50	0.14	0.16	0.27
MISP		0.016	0.99	0.51	0.11	-0.32	0.01	0.02
	Exponential Est	All options	ITM	ATM	OTM	SHORT	MEDIUM	LONG
APE		0.68	1.37	0.90	0.55	0.41	0.66	0.90
PPE		0.33	0.04	0.17	0.92	0.37	0.35	0.24
MISP		-0.73	0.91	-0.99	-0.45	-0.85	-0.80	-0.24
	Ogawa & San Est	All options	ITM	ATM	OTM	SHORT	MEDIUM	LONG
APE		0.91	1.77	0.91	0.540	0.27	0.63	0.47
PPE		0.34	0.04	0.18	0.95	0.32	0.89	0.19
MISP		-0.63	0.59	-0.90	-0.55	-0.95	-0.56	-0.25
	Fourier Estimator	All options	ITM	ATM	OTM	SHORT	MEDIUM	LONG
APE		0.69	0.18	0.89	0.53	0.39	0.64	0.96
PPE		0.29	0.05	0.16	0.80	0.32	0.30	0.21
MISP		0.64	0.92	-0.21	-0.36	-0.82	-0.71	-0.04

Table 5.5: Average absolute (APE) Mispricing index (MISP), and percentage (PPE) pricing errors between the market price and the kernel based-price.

structure of mispricing by term to expiration is very similar; this can be due to the fact that our model does not vary according to the time parameter. This can limit our nonparametric approach to be a true pricing engine.

In this last part we compare the results obtained with the nonparametric approach with those computed with some popular methods used by academicians and practitioners. That is the Black & Scholes, the Jump diffusion and the Heston model. In this comparison we have considered only the Fourier estimator. We do not compare them to the other estimators studied in this thesis, because numerical results

	Fourier Est	All options	ITM	ATM	OTM	SHORT	MEDIUM	LONG
APE	Fourier	0.69	0.18	0.89	0.53	0.39	0.64	0.96
	B-S	0.53	0.32	0.41	0.26	0.33	0.69	1.15
	J-D	0.20	0.14	0.20	0.11	0.12	0.17	0.19
	H	0.14	0.01	0.20	0.11	0.13	0.15	0.18
PPE	Fourier	0.29	0.05	0.16	0.80	0.32	0.3042	0.21
	B-S	0.40	0.01	0.12	0.53	0.27	0.54	0.90
	J-D	0.10	0.006	0.05	0.20	0.08	0.07	0.10
	H	0.08	0.006	0.05	0.19	0.10	0.07	0.08
MISP	Fourier	0.64	0.92	-0.21	-0.36	-0.82	-0.71	-0.04
	B-S	0.90	-0.09	0.59	0.74	0.84	0.88	0.82
	J-D	0.40	0.32	-0.003	0.13	0.19	0.39	0.43
	H	0.17	0.03	-0.03	0.02	0.39	0.45	0.29

Table 5.6: Average absolute (APE) Mispricing index (MISP), and percentage (PPE) pricing errors between the market price and the kernel based-price for the Fourier and B-S; J-D; H model.

obtained from them have the same features. Therefore, the comments relative to Fourier estimator are also relevant for the remainder estimators

The analysed loss function suggest that for the all four models, there is a positive mispricing indices for all categories except for the ATM and the ITM classes in general, However when using the MISP the price worked out using nonparametric approach tend to underestimate the option price for almost all categories apart from All options and ITM classes, but for the considered nonparametric model the overpricing is observed only in the all options and ITM classes, as shown in table 5.6. The result however shows that the models systematically overprice call price in general, except for the ITM calls for B-S and the ATM calls for J-D and the H model. Nevertheless as one can imagine, both the J-D and the H model are more accurate. To sum up, the nonparametric model represents a real improvement with respect to B-S in term of pricing properties. However, the performance is worse than the J-D and H.

The Table 5.7 reports the results computed with daily data for the Rolling estimator. This estimator has been chosen because it has been originally constructed

	Rolling Est	All options	ITM	ATM	OTM	SHORT	MEDIUM	LONG
APE	Rolling	28.66	4.79	5.42	6.24	26.24	23.42	19.05
	B-S	0.53	0.32	0.41	0.26	0.33	0.69	1.15
	J-D	0.20	0.14	0.20	0.11	0.12	0.17	0.19
	H	0.14	0.01	0.20	0.11	0.13	0.15	0.18
PPE	Rolling	31.12	0.21	2.46	19.53	35.27	19.06	9.14
	B-S	0.40	0.01	0.12	0.53	0.27	0.54	0.90
	J-D	0.10	0.006	0.05	0.20	0.08	0.07	0.10
	H	0.08	0.006	0.05	0.19	0.10	0.07	0.08
MISP	Rolling	-0.99	-0.26	-0.94	-0.70	-0.99	-0.99	-0.98
	B-S	0.90	-0.09	0.59	0.74	0.84	0.88	0.82
	J-D	0.40	0.32	-0.003	0.13	0.19	0.39	0.43
	H	0.17	0.03	-0.03	0.02	0.39	0.45	0.29

Table 5.7: Average absolute (APE) Mispricing index (MISP), and percentage (PPE) pricing errors between the market price and the kernel based-price for the Rolling estimator using daily data and B-S; J-D; H model.

for low frequency data (daily, weekly and monthly). From the results, we notice that the pricing error increase exaggeratedly showing that when daily data are used for computing option prices much information contained in high frequency data are neglected.

The computational cost of the nonparametric is less than using the closed-form solution for the J-D model and which is cheaper than the H model. Therefore, The J-D model represents a good trade-off between performance and computational cost. The Black and Scholes model is the cheapest and the Fourier estimator uses the highest computational time among our nonparametric estimates, as shown in Table 5.8.

In the end, our empirical evidence indicates that taking stochastic volatility into account gives the best improvement over the B-S formula. However, we can conclude that the nonparametric model contribute to explaining from theoretical and quantitative standpoint the strong pricing biases inducted in the B-S model.

	All options
	CPUtime
F-Z	1.24e+03
CR	1.25e+03
Rolling	1.202+e03
EXP	1.203e+03
O-S	1.262e+03
FR	2.345e+03
B-S	685.29
J-D	4.062e+03
H	9.629e+03

Table 5.8: CPU time in seconds for the computed European Call price on S&P 500 for different nonparametric methods and the alternative counterparts

5.7 Conclusion

Since the seminal work of Black & Scholes on options pricing, many researchers proposed sophisticated works in order better to evidence the empirical facts exhibited by the market. The stochastic volatility model highly used by institutional financial firm fail in capturing empirical market features.

Following the local volatility approach pioneered by Derman and Kani (1994), Dupire (1994), and Rubinstein (1994) who supposed that the volatility is a deterministic function of asset price and time, we have studied a new approach and used it for evaluating European call option prices.

This chapter enriches the growing literature of the nonparametric estimation by using almost all types of spot volatility existing in the literature for computing the option price of European call options. We have shown that our method can be classified in the class of local volatility function which has the particularity of being complete. Relatively to other local volatility function proposed earlier, our method is easy to manage and the computation cost is very low.

The volatility curve obtained with our approach is consistent with the actual market. The Fourier estimator shows to exhibit better the market features compared with other competing volatility estimators analysed.

When applying our method to real data any nonparametric estimator show to exhibit the same feature for every spot volatility considered in our analysis. The striking result to acknowledge is the fact that all call price to the ITM option have the same value, viz \$49. However, our result is consistent with the actual market price and the theoretical foundation of option pricing. As stated by several authors, nonparametric approaches can be used to suggest the functional form to parametric approaches and to valid them.

We compare the nonparametric model to well known popular models, such as Black & Scholes models itself, the Jump-Diffusion model and the Heston's model with stochastic volatility. The price worked out by nonparametric model is obtained using the Dupire's equation which the volatility is construted directly on the S&P 500 index future price spanning from january 4, 1993 to to December 31, 1993. To sum up, the nonparametric model represents a real improvement with respect to B-S in term of pricing properties. However, the performance is worse than the J-D and H.

Second, the computational cost of the nonparametric is less than using the closed-form solution for the J-D model and which is cheaper than the H model. Therefore, The J-D model represents a good trade-off between performance and computational cost. The Fourier estimator is the most time consuming among all the studied nonparametric estimators.

In the end, our empirical evidence indicates that taking stochastic volatility into account gives the best improvement over the B-S formula. However, we can conclude that the nonparametric model contribute to explaining from theoretical and quantitative standpoint the strong pricing biases inducted in the B-S model.

A serious limitation to this approach is that it is not varying with respect to the time parameter, including the time variations in the Nadayara-Watson estimator, may be crucial to obtain correct specification. This will permit our model to be a real pricing engin, further investigations in that sense will be the object of next researches.

Bibliography

- [1] Y. Ait-Sahalia and A.W Lo. A nonparametric estimation of state-price densities implicit in financial asset prices. *Econometrica*, 53(2):499–547, 1998.
- [2] Yacine Ait-Sahalia. A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econometrica*, 64(3):527–560, 1996.
- [3] T.G. Andersen, T. Bollerslev, and F.X. Diebold. Parametric and nonparametric volatility measurement. Technical report, National Bureau of Economic Research, 2002.
- [4] E. Andreou and E. Ghysels. Rolling-sample volatility estimators. *Journal of Business and Economic Statistics*, 20(3):363–376, 2002.
- [5] O.E. Barndorff-Nielsen. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280, 2002.
- [6] E. Barucci and R. Renò. On measuring volatility and the garch forecasting performance. *Journal of International Financial Markets, Institutions and Money*, 12:180–200, 2002a.
- [7] E. Barucci and R. Reno. On measuring volatility of diffusion processes with high frequency data. *Economics Letters*, 74(3):371–378, 2002.
- [8] T. Björk. *Arbitrage theory in continuous time*. Oxford University Press, USA, 2009.

- [9] F Black and M. Scholes. The pricing of options and corporate liabilities. *J.Polit. Econ.*, 81:637–659, 1973.
- [10] J.M Volle S. Bonnans, B. Cagnet. Estimation de la volatilité locale d’actifs financier par une methode numerique. 2002.
- [11] D.T. Breeden and R.H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651, 1978.
- [12] K.C. Chan, G.A. Karolyi, F.A. Longstaff, and A.B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, pages 1209–1227, 1992.
- [13] D.A. Chapman, J.B. Long Jr, and N.D. Pearson. Using proxies for the short rate: when are three months like an instant? *Review of Financial Studies*, 12(4):763–806, 1999.
- [14] F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.
- [15] E. Derman and I. Kani. Riding on a smile. *Risk*, 7(2):32–39, 1994.
- [16] B. Dumas, J. Fleming, and R.E. Whaley. Implied volatility functions: Empirical tests. *The Journal of Finance*, 53(6):2059–2106, 1998.
- [17] B. Dupire. Pricing with smile. 2004.
- [18] H. Ebens. Realized stock volatility. *Department of Economics, Johns Hopkins University*, 1999.
- [19] J. Fan. A selective overview of nonparametric methods in finance. *Statistical Science*, 20(4):317–337, 2005.
- [20] J. Fan, Y. Fan, and J. Jiang. Dynamic integration of time-and state-domain methods for volatility estimation. *Journal of the American Statistical Association*, 102(478):618–631, 2007.

- [21] J. Fan and Y. Wang. Spot volatility estimation for high-frequency data. *Statistics and its Interface*, 1:279–288, 2008.
- [22] D.P. Foster and D.B. Nelson. Continuous record asymptotics for rolling sample variance estimators. Technical report, National Bureau of Economic Research, 1994.
- [23] J. Gatheral. *The volatility surface; A Practitioner's guide*. John Wiley & Sons, Inc, 2006.
- [24] C.A.E. Goodhart and M. O'Hara. High frequency data in financial markets: Issues and applications. *Journal of Empirical Finance*, 4(2-3):73–114, 1997.
- [25] D. Lesniewski A. S. Hagan, P.S.Kumar and D. E. Woodward. Managing smile risk. *Wilmott Magazine*, 20:84–108, 2002.
- [26] P.R. Hansen, J. Large, and A. Lunde. Moving average-based estimators of integrated variance. *Econometric Reviews*, 27(1-3):79–111, 2008.
- [27] P.R. Hansen and A. Lunde. A forecast comparison of volatility models: does anything beat the garch (1,1)?. *Journal of applied econometrics*, 20:873–889, 2005.
- [28] P.R. Hansen and A. Lunde. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24(2):127–161, 2006.
- [29] L. Harris. Stock price clustering and discreteness. *Review of Financial Studies*, 4(3):389–415, 1991.
- [30] J.D. Hart. *Nonparametric smoothing and Lack of fit Test*. Springer Verlag, 1997.
- [31] S.L. Heston. A closed-form solution for options with stochastic volatility with application to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

- [32] A.L Hull, J.C White. The pricing of options on assets with stochastic volatility. *The Journal of finance*, 42:281–300, 1987.
- [33] J. Jacod and P. Protter. Asymptotic error distributions for the euler method for stochastic differential equations. *The Annals of Probability*, 26(1):267–307, 1998.
- [34] G Jiang. Nonparametric modeling of u.s. interest rate term structure dynamics and implications on the prices of derivative securities. *Journal of Financial and Quantitative Analysis*, 33(4):465–497, 1998.
- [35] G. Jiang and J. Knight. Nonparametric pricing of interest rate derivative securities. *Econometric Theory*, 13(5):615–645, 1997.
- [36] D. Kristensen. Nonparametric filtering of the realized spot volatility: a kernel-based approach. *Econometric Theory*, 26(01):60–93, 2010.
- [37] A. Madhavan. Market microstructure: A survey. *Journal of Financial Markets*, 3(3):205–258, 2000.
- [38] J.M. Maheu and T.H. McCurdy. Nonlinear features of realized fx volatility. *Review of Economics and Statistics*, 84(4):668–681, 2002.
- [39] P. Malliavin and M.E. Mancino. Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics*, 6(1):49–61, 2002.
- [40] C. Mancini and R. Renò. Threshold estimation of markov models with jumps and interest rate modeling. *Journal of Econometrics*, 160(1):77–92, 2011.
- [41] M. McAleer and M.C. Medeiros. Realized volatility: A review. *Econometric Reviews*, 27(3):10–45, 2008.
- [42] N. Meddahi. A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17(5):479–508, 2002.

- [43] R.C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144, 1976.
- [44] R.C. Merton. On estimating the expected return on the market* 1:: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361, 1980.
- [45] P.A. Mykland and L. Zhang. Inference for volatility-type objects and implications for hedging. Technical report, Tech. rep., Carnegie-Mellon University, 2001.
- [46] S. Ogawa and S. Sanfelici. An improved two-step regularization scheme for spot volatility estimation. *Economic Notes*, pages 105–132, 2011.
- [47] P.C.B. Phillips, J.Y. Park, and Yale University. Cowles Foundation for Research in Economics. Nonstationary density estimation and kernel autoregression. *Cowles Foundation Discussion Paper*, 1998.
- [48] P.E. Protter. *Stochastic integration and differential equations*, volume 21. Springer Verlag, 2004.
- [49] R Renò. Nonparametric estimation of diffusion coefficient of stochastic volatility model. *Econometric theory*, 24(5):1174–1206, 2008.
- [50] R. Renò, A. Roma, and S. Schaefer. A comparison of alternative nonparametric estimators of the short rate diffusion coefficient. *Economic Notes*, 35(3):227–252, 2008.
- [51] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, volume 4. Springer Verlag, 1998.
- [52] R. Riccardo. *Volatility And Correlation*. 1999.
- [53] M. Rubinstein. Implied binomial trees. *Journal of finance*, pages 771–818, 1994.
- [54] J.R. Russell and F.M. Bandi. Microstructure noise, realized volatility, and optimal sampling. In *Econometric Society 2004 Latin American Meetings*. Econometric Society, 2004.

- [55] S. Sanfelici. Calibration of a nonlinear feedback option pricing model. *Quantitative Finance*, 7(1):95–110, 2007.
- [56] D.W. Scott. *Multivariate density estimation*, volume 139. Wiley Online Library, 1992.
- [57] B.W. Silverman. *Density estimation for statistics and data analysis*, volume 26. Chapman & Hall/CRC, 1986.
- [58] R. Stanton. A nonparametric model of term structure dynamics and the market price of interest rate risk. *The Journal of Finance*, 52(5):1973–2002, 1997.
- [59] D. Tavella and C. Randall. *Pricing financial instruments: The finite difference method*. John Wiley & Sons, 2000.
- [60] P. Wilmott, S. Howison, and J. Dewynne. *The mathematics of financial derivatives: a student introduction*. Cambridge Univ Pr, 1995.
- [61] L. Zhang, P.A. Mykland, and Y. Ait-Sahalia. A tale of two time scales. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.
- [62] B. Zhou. High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52, 1996.
- [63] Y. ZU and P. BOSWIJK. Estimating realized spot volatility with noisy high frequency data. Technical report, Working Paper, 2009.