AFFINE FRACTIONAL STOCHASTIC VOLATILITY MODELS

F. COMTE*, L. COUTIN*, AND E. RENAULT**

ABSTRACT. By fractional integration of a square root volatility process, we propose in this paper a long memory extension of the Heston (1993) option pricing model. Long memory in the volatility process allows us to explain some option pricing puzzles as steep volatility smiles in long term options and co-movements between implied and realized volatility. Moreover, we take advantage of the analytical tractability of affine diffusion models to clearly disentangle long term components and short term variations in the term structure of volatility smiles. In addition, we provide a recursive algorithm of discretization of fractional integrals in order to be able to implement a method of moments based estimation procedure from the high frequency observation of realized volatilities.

AMS Classification 2000: 91B70, 62P05, 62K99

1. INTRODUCTION

As extensively discussed by Sutton (2000) in his recent methodological lectures, the empirical option pricing literature, which is based on the Black and Scholes (1973)(BS) model, has focused heavily on problems surrounding the measurement of the volatility parameter σ. Hull and White (1987)(HW) pioneered the use of the continuous-time stochastic volatility models to capture the effect of stochastic variations in this parameter. Renault and Touzi (1996) have shown that a HW model with a stochastic volatility process $\sigma(t)$ independent from the standardized Brownian innovation of the stock price process implies a U-shaped symmetric volatility curve, whereby the volatility $\sigma_{\mathrm{imp,h}}(t)$ extracted at time t from the BS option pricing formula given the observed (HW) option price (for a given time to maturity h) is graphed against the log-moneyness of the option.

More generally, the empirical biases of the BS model have been dubbed the smile effect in reference to a symmetric implied volatility curve, but numerous distorted smiles in the shape of smirks or frowns are inferred more frequently from market data. Renault (1997) and Garcia, Luger, Renault (2001) discuss some extensions of the HW option pricing model which account for the presence of an implied volatility smile, smirk or frown in option data. The basic idea is to explain asymmetric smiles by an instantaneous correlation between returns and volatility in the line of Heston (1993) option pricing model.

Irrespective of its interpretation, the stochastic feature of the volatility process and its instantaneous correlation with the return of the underlying asset appear to be relevant for explaining the

Key words and phrases. Fractional integrals, Long memory processes, Integrated volatility, Stochastic volatility.

* Université René Descartes-Paris V, Laboratoire MAP5.

** University of North Carolina, Chapel Hill.

variation in strike of the pricing performance of standard (BS) option pricing models. However, as pointed out by Sundaresan (2000) in his survey of the performance of continuous-time methods for option valuation, the remaining puzzle is the so-called term structure of volatility smiles, that is, the fact that the volatility smile effect appears to be dependent in a systematic way on the maturity structure of options. Actually, Sundaresan (2000) first acknowledges that the volatility smile appears to be stronger in short term options than in long term ones, which is consistent with the stochastic volatility interpretation. When volatility is stochastic, the option price appears to be an expectation of the BS price with respect to the probability distribution of the so-called "integrated volatility" $(1/h)\int_t^{t+h}\sigma^2(u)du$ over the lifetime of the option, indeed only a fraction of it in the Heston model (see Renault and Touzi (1996) for the HW case, Romano and Touzi (1997) and Section 5 below for the extension of the Heston model). In other words, the fixed volatility parameter σ^2 of BS appears to be replaced by the average of the volatility process random path over the lifetime of the option. By a simple application of the law of large numbers to the volatility process (assumed to be stationary and ergodic), one realizes that the effects of the randomness of the volatility should vanish when the time to maturity of the option increases and therefore the volatility smile should be erased.

Sundaresan (2000) is nevertheless right to conclude that the term structure of implied volatilities still appears to have patterns that cannot be so easily reconciled. The main issue to address is to reconcile short term and long term observed patterns of the term structure of implied volatilities. On the one hand, stochastic volatility effects appear to be still significant for very long maturity options as documented by Bollerslev and Mikkelsen (1999). On the other hand, observed prices for very short term option cannot be explained inside the stochastic volatility framework without introducing huge volatility risk premia which would become explosive in longer terms.

We propose in this paper a continuous time stochastic volatility model with long memory, as in Comte and Renault (1998)(CR). It allows us to endow the volatility process with high persistence in the long run in order to capture the steepness of long term volatility smiles without increasing the short run persistence. Actually, as in CR, the volatility process appears to be less persistent in the very short run than any standard diffusion volatility process while it is much more persistent in the long run: the autocovariance function of the volatility process decreases at an hyperbolic rate for infinitely large lags instead of the standard exponential rate.

The main contribution of this paper is to propose a long memory specification of the volatility process and an associated option pricing model which, while maintaining the appealing features of the CR model, appears to be much more tractable for financial applications. The square root volatility process and namely the long memory version of it we introduce in this paper is in several important respects the best suited for option pricing and hedging.

While the square root process, as first studied by Feller (1951) was popularized for interest rates modelling by Cox, Ingersoll and Ross (1985), its relevance for volatility modelling in the context of quadratic-variation-based strategies of portfolio insurance was put forward by Geman and Yor (1993). As stressed by these auhtors, a first advantage of the square root process is its interpretation as a Bessel squared process, since the class of laws of squares of Bessel processes is stable by convolution operation, a property which is not shared by the commonly used models of positive processes, like the log-normal one. This may justify to see the volatility of the return on a portfolio as a square root process since it is consistent with a similar assumption for the volatility process of its components. This nice aggregation property will be maintained for the long memory extension we propose in this paper.

Actually, we are going to show that when the volatility process is LMSR, the integrated volatility (for a given time to maturity) is endowed by this long memory feature only through its conditional expectation, while higher order conditional centered moments are short memory. This result is important because it means that the model captures the two stylized facts of long memory dynamics of implied volatilities and deep volatility smiles in long term options without introducing some unpalatable long range dependence in the stochastic process of the volatility smiles. In other words, only one long memory state variable (which is well proxied by at the money BS implied volatilities) is needed to compute option prices in practice; besides, conditionally to this variable, one recovers nice Markovian properties of option prices.

Moreover, the LMSR model is specified by a parsimonious set of parameters which are easy to interpret, in relation to both short and long term behavior of the volatility process, and not so difficult to estimate. In these two respects, the great advantage of the LMSR model is to allow us to derive explicit formula of the various moments of interest as simple functions of these parameters. Actually, the model of the volatility process that we propose here can be interpreted, in the short memory extreme case, as a translation of a standard square root model. This is the reason why we recover the simplicity of analytical computations in an affine model as recently put forward by Duffie, Pan and Singleton (2000). The price to pay for this simplicity is to work with a process, the square root one, the diffusion coefficient of which is not Lipschitz with respect to the state. This leads us to set in this paper a self-contained theory of fractional integration of the square root process which is not a direct application of CR (see Hu and Nualart (2005)) or of the more comprehensive theory which has been developed in the last ten years (see Cheridito and Nualart (2005) and the references therein). However, the discretization scheme we propose for practical implementation of this continuous time model with discrete time data heavily rests upon the recent works of Carmona, Coutin and Montseny (2000) and Coutin and Pontier (2001).

The paper is organized as follows. The general Markov setting is presented in section 2, as well as the relevant pieces of fractional integration theory. This allows us to propose a continuous time stochastic volatility model where the LMSR volatility process is defined, up to an additive constant, as fractional integration of a square root process. The resulting features of volatility persistence are illustrated by a characterization of the autocorrelation function of the volatility process as well as by continuous record asymptotics for rolling sample volatility estimators. Section 3 is devoted to the study of integrated volatility in terms of term structure, memory property and quadratic variation approximations. The issue of statistical inference is addressed in Section 4, starting with the problem of the discretization of fractional operators both for simulation and estimation purpose. Then it is shown that all the volatility model parameters can be consistently estimated by simple methods of moments completed with a step of fractional derivation. A short scale Monte Carlo study shows the feasibility of the procedure and discusses the choice of some discretization parameters. It is worth noticing that the statistical approach described in Section 4 is only based on high frequency observations of a return process, the volatility of which is the focus of our interest. Section 5 gathers the proofs of all previous sections.

2. The general framework

2.1. **The Markov setting.** Let $S(t)$ denote the date t level of an equity index, a foreign currency or some other asset. Assume that $S(t)$ is generated by the process:

$$dS(t) \quad = \quad r(t)S(t)dt + \sigma(t)S(t)dW^S(t) \tag{2.1}$$

where $W^S(t)$ is a standard Brownian motion and $r(t)$ and $\sigma(t)$ denote respectively the short term interest rate, and a volatility process.
We assume that there exists a state variable $\tilde{\sigma}^2(t)$ such that $(S(t), \tilde{\sigma}^2(t))$ is jointly stationary Markovian with respect to: $\mathcal{I}(t) = \sigma[S(\tau), \tilde{\sigma}^2(\tau), \quad \tau \leq t]$. For sake of expositional simplicity, we also assume that $\tilde{\sigma}^2(t)$ is a univariate diffusion process:

$$d\tilde{\sigma}^2(t) = f(\tilde{\sigma}^2(t))dt + g(\tilde{\sigma}^2(t))dW^{\tilde{\sigma}}(t), \tag{2.2}$$

with: $W^{\tilde{\sigma}}(t)$ standard Brownian motion, $\text{cov}[dW^S(t), dW^{\tilde{\sigma}}(t)] = \rho(t)dt$, $\sigma(t), \rho(t), r(t)$ deterministic functions of the path $\tilde{\sigma}^2(\tau)$, $0 \leq \tau \leq t$.

Under regularity assumptions, equations (2.1) and (2.2) define the stationary stochastic process $(S(t), \tilde{\sigma}^2(t))$ without ambiguity and $\mathcal{I}(t) = \sigma\{W^S(\tau), W^{\tilde{\sigma}}(\tau), \tau \leq t\}$. In the following, we shall denote $\mathbb{E}_t(.) = \mathbb{E}(.|\mathcal{I}(t))$ and similarly for conditional variances.

2.2. **The volatility process.**

2.2.1. *Fractional integration.* Let X be a centered continuous stochastic process. Following Samko et al. (1993), we can define for $0 < \alpha < 1$ the fractional integration operator:

$$I^{(\alpha)}(X)(t) = \int_{-\infty}^{t} \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)} X(s)ds \tag{2.3}$$

for any given sample path $s \to X(s)$ such that $|X(s)|^p$ is Lebesgue integrable on $\mathbb{R}$ for some p in $[1, 1/\alpha[$. Under regularity conditions, this would allow an a.s. definition of a stochastic process $X^{(\alpha)}(t) = I^{(\alpha)}(X)(t)$ that would be strongly stationary when $X(t)$ is.

The basic intuition behind this definition is to consider a stochastic process $X^{(\alpha)}(t)$ whose sample paths, deduced from those of $X(t)$ by fractional integration, are smoother than the ones of $X(t)$. In the limit case $\alpha \to 1$, $X^{(\alpha)}(t)$ would be the unit root process whose first difference coincides with the mean value of $X(t)$ on the unit time interval. The opposite limit case $\alpha \to 0$ amounts to $X^{(\alpha)}(t)$ equal to $X(t)$. To see this, note that a "formal" integration by part on (2.3) gives, by assuming that $\lim_{s\to-\infty}|s|^{\alpha}X(s)=0$:

$$I^{(\alpha)}(X)(t)=\int_{-\infty}^{t}\frac{(t-s)^{\alpha}}{\Gamma(\alpha+1)}dX(s)\longrightarrow_{\alpha\to0}X(t).$$

Yet, we focus in this paper on a $\mathbb{L}^2$-definition of $X^{(\alpha)}(t)$. We first note that, for $\omega>0$ given,

$$\int_{-\omega}^{t}\frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}X(s)ds,\quad t>-\omega,$$

is well defined for $\alpha>0$. $X^{(\alpha)}(t)$ can be constructed as the mean-square limit of this integral when ω goes to infinity, insofar as $\alpha<\frac{1}{2}$ and the process X is short memory:

Proposition 2.1. *For $0\le\alpha<\frac{1}{2}$, and if X is mean-square stationary and with an exponentially decaying autocovariance function ($|\gamma_X(u)|\le\gamma_X(0)e^{-k|u|}$), then $X^{(\alpha)}(t)=I^{(\alpha)}(X)(t)$ is mean square stationary and for $t\to+\infty$: $\left\|X^{(\alpha)}(t)-\int_0^t[(t-s)^{\alpha-1}/\Gamma(\alpha)]X(s)ds\right\|_2=O\left(t^{\alpha-1/2}\right)$.*

Conversely, fractional derivation corresponds to the operator $I^{(\alpha)}(X)(t)$ for negative α. In this case too, one may consider either the a.s. definition on the sample paths or the mean-square definition. In both cases, the integral (2.3) is not defined for negative α and must be replaced by:

$$I^{(\alpha)}(X)(t)=\frac{d}{dt}\left[\int_{-\infty}^{t}\frac{(t-s)^{\alpha}}{\Gamma(1+\alpha)}X(s)ds\right]. \tag{2.4}$$

It follows from Samko et al. (1993) and from the mean-square definition that:

Proposition 2.2. *For $0<\alpha<\frac{1}{2}$, if X is mean-square stationary with an exponentially decaying autocovariance function,*

$$I^{(-\alpha)}\left[I^{(\alpha)}(X)(t)\right]=X(t).$$

For functions, the result is easy to see by a simple application of Fubini's theorem.

2.2.2. *Fractional integration of a square root process.* From now on, we assume that for some positive k,γ and $\tilde{\theta}$, $\tilde{\sigma}^2(t)$ is the square root process:

$$d\tilde{\sigma}^2(t)=k(\tilde{\theta}-\tilde{\sigma}^2(t))dt+\gamma\tilde{\sigma}(t)dW^X(t). \tag{2.5}$$

Then (see e.g. Lamberton and Lapeyre (1996)): $k\tilde{\theta}\ge\gamma^2/2\ \Rightarrow\mathbb{P}(\tau_0^x=+\infty)=1$, where $\tau_0^x=\inf\left\{t\ge0\ /\ \tilde{\sigma}_x^2(t)\le0\right\}$ and $\tilde{\sigma}_x^2(t)$ denotes the solution of (2.5) starting from $x\ge0$. The assumption $k\tilde{\theta}\ge\gamma^2/2$ is maintained in all the rest of the paper. We consider the stationary solution of (2.5), its mean is $\tilde{\theta}$ and denote the centered version of the process by

$$\tilde{\sigma}_c^2(t)=\tilde{\sigma}^2(t)-\tilde{\theta}. \tag{2.6}$$

Note that since $\tilde{\sigma}^2$ remains positive with probability one, we can take $\tilde{\sigma}(t)=\sqrt{\tilde{\sigma}^2(t)}$. It is worth noticing that the variability of $\tilde{\sigma}^2(t)$ is increasing with the diffusion coefficient γ, decreasing with its mean-square reversion parameter k, but also increasing with the parameter $\tilde{\theta}$. These properties are maintained for the fractionally integrated process $(\tilde{\sigma}_c^2)^{(\alpha)}(t)$ and explicit formulas are available:

Proposition 2.3. *Let* $0 < \alpha < \frac{1}{2}$, $Var((\tilde{\sigma}_c^2)^{(\alpha)}(t)) = \dfrac{\tilde{\theta}\gamma^2}{k^{2\alpha+1}} \dfrac{\Gamma(1-2\alpha)\Gamma(2\alpha)}{\Gamma(1-\alpha)\Gamma(\alpha)}$.

In the limit case $\alpha \to 0$, using that $\Gamma(\alpha) \sim_{\alpha \to 0} 1/\alpha$, one gets the variance of the square-root process:

$$\text{Var } [\tilde{\sigma}^2(t)] = \text{ Var } [\tilde{\sigma}_c^2(t)] = \frac{\tilde{\theta}\gamma^2}{2k}. \tag{2.7}$$

To define a long memory volatility process, we propose to set

$$\sigma^2(t) = |\theta + (\tilde{\sigma}_c^2)^{(\alpha)}(t)|. \tag{2.8}$$

Note that, since $(\tilde{\sigma}_c^2)^{(\alpha)}(t)$ is not lower bounded, the positivity of $\theta + (\tilde{\sigma}_c^2)^{(\alpha)}(t)$ will never be guaranteed, irrespective of the value of θ. However, if θ is large with respect to Var $[(\tilde{\sigma}_c^2)^{(\alpha)}(t)]$, one may consider for all practical purpose that the following process

$$\breve{\sigma}^2(t) = \theta + (\tilde{\sigma}_c^2)^{(\alpha)}(t) \tag{2.9}$$

is a good approximation of $\sigma^2(t)$, and therefore, this is the process we study all along the paper. Actually, it is well documented on common stock price data (see e.g. Jacquier, Polson and Rossi (1994)) that the mean value of the volatility process is often between 10% and 20% while the coefficient of variation $\gamma^2/(2k)$ is much smaller than one. This allows us to consider, for the sake of expositional simplicity in all the rest of the paper, that the volatility process is given by (2.9) for $\theta = \mathbb{E}\breve{\sigma}^2(t)$ of the same order of magnitude as $\tilde{\theta} = \mathbb{E}\tilde{\sigma}^2(t)$.

2.2.3. *Volatility persistence.* The autocovariance function $c_{\breve{\sigma}^2}(h) = \text{cov}(\breve{\sigma}^2(t), \breve{\sigma}^2(t+h))$ of the volatility process is easily deduced from the one of $\tilde{\sigma}^2$ denoted by $c_{\tilde{\sigma}^2}(h) = \text{cov}(\tilde{\sigma}^2(t), \tilde{\sigma}^2(t+h))$, by

$$\begin{aligned} c_{\breve{\sigma}^2}(h) &= \int_{-\infty}^{t+h} \int_{-\infty}^{t} \frac{(t+h-s)^{\alpha-1}(t-u)^{\alpha-1}}{\Gamma(\alpha)^2} \text{cov}(\tilde{\sigma}^2(u), \tilde{\sigma}^2(s)) du ds \\ &= \int_0^{+\infty} \int_0^{+\infty} \frac{x^{\alpha-1} y^{\alpha-1}}{\Gamma(\alpha)^2} c_{\tilde{\sigma}^2}(h+y-x) dx dy \end{aligned} \tag{2.10}$$

Then, we have just to plug into (2.10) the well-known formula for the function c_X:

$$c_{\tilde{\sigma}^2}(h) = \text{cov}(\tilde{\sigma}^2(t+h), \tilde{\sigma}^2(t)) = \frac{\tilde{\theta}\gamma^2}{2k} e^{-k|h|}, \tag{2.11}$$

to get $c_{\breve{\sigma}^2}(h)$ as an explicit function of the parameters $k, \tilde{\theta}, \gamma$ and α. Proposition 2.4 makes explicit the volatility persistence property in the very short term as well as in the long term:

Proposition 2.4. *For* $0 < \alpha < 1/2$ *and* $c_{\breve{\sigma}^2}(0) = Var(\breve{\sigma}^2(t)) (= Var(I^{(\alpha)}(\tilde{\sigma}_c^2)(t)))$ *given by Proposition 2.3:*

$$\frac{c_{\breve{\sigma}^2}(h)}{c_{\breve{\sigma}^2}(0)} = 1 - \frac{(kh)^{2\alpha+1}}{2\alpha(2\alpha+1)\Gamma(2\alpha)} + O(h^2) \text{ when } h \text{ tends to zero,}$$

$$\frac{c_{\breve{\sigma}^2}(h)}{c_{\breve{\sigma}^2}(0)} \sim \frac{(kh)^{2\alpha-1}}{\Gamma(2\alpha)} \text{ when } h \text{ tends to infinity.}$$

The first striking property of long memory is that, when the horizon h goes to infinity, the autocovariance function decays to zero with an hyperbolic rate instead of the common exponential rate for short memory processes. In this respect, there is some discontinuity when $\alpha \to 0$. On the other hand, the fractional Brownian motion-like behavior of the process $I^{(\alpha)}(\tilde{\sigma}_c^2)(t)$ implies that in the very short term, the autocorrelation function reaches 1 with the speed $(kh)^{2\alpha+1}$ (with the factor $2\alpha(2\alpha+1)\Gamma(2\alpha))$ instead of kh. Continuity for $\alpha \to 0$ is ensured by the fact that $\Gamma(\alpha) \sim 1/\alpha$ when $\alpha \to 0$. In both cases, $h \to 0$ or $h \to +\infty$, it is worth noticing that the common role of the mean reversion parameter k is maintained through fractional integration.
As a consequence, the global long memory properties of the long memory process $\breve{\sigma}^2(t)$ can be characterized through its spectral density:

Proposition 2.5. *For all positive λ, $f_{\breve{\sigma}^2}(\lambda) = \int e^{-i\lambda h} c_{\breve{\sigma}^2}(h)dh = f_{\tilde{\sigma}^2}(\lambda)\lambda^{-2\alpha}$, with $f_{\tilde{\sigma}^2}(\lambda) = (\tilde{\theta}\gamma^2)/(\lambda^2 + k^2)$.*

The higher degree of volatility persistence implied by higher values of the integration parameter α explains the better performance of continuous record asymptotics to filter the volatility process. To see this, let us consider an interval of time $[0, T]$ in which we can observe $(N+1)$ high frequency realizations $R(t_k) = \log(S(t_k))$ of the log-prices at dates $t_k = kT/N$, $k = 0, 1, \dots, N$. By computing the rolling sample estimator

$$\hat{\sigma}^2_{N,p}(t) = \frac{N}{pT} \sum_{k=[tN/T]-p+1}^{[tN/T]} (R(t_k) - R(t_{k-1}))^2, \tag{2.12}$$

we are faced with a standard bias-variance trade off. The larger p, the smaller the variance of $\hat{\sigma}^2_{N,p}(t)$ but the larger its bias with respect to $\sigma^2(t)$ since it is based on more lagged observations. However, thanks to volatility persistence, this is all the less detrimental when α is large:

Proposition 2.6. *Let $R(t) = R(0) + \int_0^t \mu(s)ds + \int_0^t \breve{\sigma}(s)dW^S(s)$ with σ defined as in (2.9) and $\mathbb{E}(|\mu(s)|^4) < M, \forall s \geq 0$. Let T be fixed and N and p such that p/N is small, then*

$$\mathbb{E}[\hat{\sigma}^2_{N,p}(t) - \breve{\sigma}^2(t)]^2 \leq \frac{2\mathbb{E}(\sigma^4)}{p} + K\left(p\frac{T}{N}\right)^{2\alpha+1} + \frac{K'}{N} + O\left(\left(\frac{pT}{N}\right)^2\right)$$

where K is a constant and K' is a constant proportional to $\sqrt{M\mathbb{E}(\sigma^4)}$.

Proposition 2.6 extends to the case of affine long memory volatility processes some previous results of Nelson (1991) for short memory processes and Comte and Renault (1998) for log-normal long memory processes.
In the affine setting, the role of the various parameters can be characterized analytically. Moreover, the optimal bias-variance trade off is reached when p goes to infinity as $[N/p]^{2\alpha+1}$ that is $p = [N^{\frac{2\alpha+1}{2\alpha+2}}]$, leading to a rate of convergence $N^{-\frac{2\alpha+1}{2\alpha+2}}$ for the mean squared error. In other words, when the degree of the fractional process increases from 0 to 1/2, p can be increased from $N^{1/2}$ to $N^{2/3}$ with an inversely proportional rate of convergence for the mean squared error.

3. Integrated volatility

We study in this section several properties of integrated volatility, defined as $\int_{t_1}^{t_2} \breve{\sigma}^2(s)ds$ for any pair $t_1 < t_2$ of dates. Three issues are addressed.

First, the term structure of integrated volatilities, that is the function $h \mapsto \int_t^{t+h} \breve{\sigma}^2(s)ds$.

Second, the memory property of the integrated volatility process, defined as $Z(t+1) = \int_t^{t+1} \breve{\sigma}^2(s)ds$.

Third, the sample counterpart of integrated volatility provided by quadratic variation asymptotic theory.

3.1. **Term structure.**

The term structure of integrated volatilities at time t, that is the function $h \mapsto \int_t^{t+h} \sigma^2(s)ds$ raises some important issues. It is already worth reminding that:

First, Black-Scholes implied volatilities at time t for option maturing at time $(t+h)$ are tightly related to the conditional expectation $\mathbb{E}_t\left[(1/h)\int_t^{t+h} \breve{\sigma}^2(s)ds\right]$ of time-averaged integrated volatility. This so-called expectation hypothesis has been well-documented in empirical option pricing studies including Campa and Chang (1995) and Byoun, Kwok and Park (2003).

Second, the shape of the volatility smile at time t for a given maturity $(t+h)$ must be explained in relation with the conditional probability distribution given $\mathcal{I}(t)$ of the unexpected part

$$U_t(h) = \int_t^{t+h} \breve{\sigma}^2(s)ds - \mathbb{E}_t\left[\int_t^{t+h} \breve{\sigma}^2(s)ds\right] \tag{3.1}$$

of the integrated volatility. In particular, since the steepness of the volatility smile has much to do with a Jensen effect (see Renault and Touzi (1996)), the size of the conditional variance $V_t\left[\int_t^{t+h} \breve{\sigma}^2(s)ds\right]$ plays a crucial role in its explanation.

Note that the conditional probability distribution of the integrated volatility given $\mathcal{I}(t)$ is actually a conditional probability distribution given the sub-σ-field of the state variables: $\mathcal{F}(t) = \sigma\left[\tilde{\sigma}^2(\tau), \tau \le t\right]$. Then, since the volatility process is long memory, one might be afraid that this conditional probability distribution depends upon the whole path $\tilde{\sigma}^2(\tau), \tau \le t$ of state variables, making option pricing in this context a daunting task. But the main result of this subsection is that $U_t(h)$ inherits the Markov property of the state variables process $\tilde{\sigma}^2(t)$. More precisely, we are going to show that the probability distribution of $U_t(h)$ given $\mathcal{F}(t)$ depends upon $\mathcal{F}(t)$ only through the current state $\tilde{\sigma}^2(t)$. It is actually determined by the conditional probability distribution given $\tilde{\sigma}^2(t)$ of the path $\tilde{\sigma}^2(\tau), t \le \tau \le t+h$ of the state variables over the lifetime of the option.

This very convenient result is not specific to the affine stochastic volatility model but more generally a consequence of a kind of commutativity property between two integration operators: the fractional integration operator which defines the volatility process on the one hand and the common integration operator which defines the integrated volatility on the other hand. Thus the integrated volatility is decomposed in three parts:

$$\int_t^{t+h} \breve{\sigma}^2(s)ds = \theta h + f_{\alpha,h}(\mathcal{F}_t) + \frac{1}{\Gamma(\alpha+1)}\int_0^h (h-s)^\alpha \tilde{\sigma}_c^2(t+s)ds. \tag{3.2}$$

where: θh is its unconditional expectation,

$$f_{\alpha,h}(\mathcal{F}_t) = \frac{1}{\Gamma(\alpha+1)} \int_{-\infty}^{t} \left[(t+h-s)^{\alpha} - (t-s)^{\alpha}\right] \tilde{\sigma}_c^2(s) ds$$

belongs to the information $\mathcal{F}_t$ available at time t, while, by virtue of the Markovianity of the process $\tilde{\sigma}_c^2(t)$, the optimal forecast at time t of the third part is

$$\frac{1}{\Gamma(\alpha+1)} \int_0^h (h-s)^{\alpha} \mathbb{E}_t \tilde{\sigma}_c^2(t+s) ds = g_{\alpha,h}(\tilde{\sigma}_c^2(t)) \tag{3.3}$$

for some deterministic function $g_{\alpha,h}$. In the particular case of an Ornstein-Uhlenbeck like state variables process, as for instance the affine model,

$$\mathbb{E}_t \tilde{\sigma}_c^2(t+h) = e^{-kh} \tilde{\sigma}_c^2(t) \tag{3.4}$$

and then

$$g_{\alpha,h}(\tilde{\sigma}_c^2(t)) = G_{\alpha}(h) \tilde{\sigma}_c^2(t) \quad \text{with} \quad G_{\alpha}(h) = \frac{1}{\Gamma(\alpha+1)} \int_0^h (h-s)^{\alpha} e^{-ks} ds. \tag{3.5}$$

We are then able to state

Proposition 3.1. *The conditional distribution of $U_t(h)$ defined by (3.1) given $\mathcal{F}(t)$ is a deterministic function of the conditional probability distribution of $(\tilde{\sigma}^2(\tau))_{t\leq\tau\leq t+h}$ given $\tilde{\sigma}_c^2(t)$. This deterministic function is defined by*

$$U_t(h) = \frac{1}{\Gamma(\alpha+1)} \int_0^h (h-s)^{\alpha} \tilde{\sigma}_c^2(t+s) ds - g_{\alpha,h}(\tilde{\sigma}_c^2(t))$$

with $g_{\alpha,h}(\tilde{\sigma}_c^2(t))$ given by (3.3).

It is worth noticing that the linear formula $g_{\alpha,h}(\tilde{\sigma}_c^2(t)) = G_{\alpha}(h)\tilde{\sigma}_c^2(t)$ is actually valid for any Ornstein-Ulhenbeck like state variables process $\tilde{\sigma}_c^2$ conformable to (3.4). In addition, the affine structure allows us to get closed-form formulas for all higher moments $\mathbb{E}_t[U_t(h)]^k$, $k = 2, \ldots$. The nice consequence of the affine structure is that not only these moments depend on $\mathcal{F}(t)$ only through $\tilde{\sigma}_c^2(t)$, but this dependence is linear:

Proposition 3.2. *For all $p = 2, 3, \ldots$, there exist deterministic quantities $A_p(\alpha, h)$ and $B_p(\alpha, h)$ such that*

$$\mathbb{E}_t \left[U_t(h)\right]^p = A_p(\alpha, h) \tilde{\sigma}_c^2(t) + B_p(\alpha, h). \tag{3.6}$$

The proof provided in Appendix gives explicit formulas for the coefficients $A_p(\alpha, h)$ and $B_p(\alpha, h)$ (see equations (5.9) and (5.10)). Moreover, it is also proved that for any $p \geq 2$, for $h \to +\infty$,

$$A_p(\alpha, h) \sim \frac{p!}{k\Gamma(\alpha+1)^p} \left(\frac{\gamma^2}{2k^2}\right)^{p-1} h^{p\alpha} \text{ and } B_p(\alpha, h) \sim \frac{p!}{k\Gamma(\alpha+1)^p} \frac{\tilde{\theta}}{2} \left(\frac{\gamma^2}{k}\right)^{p-1} h^{p\alpha+(p-1)} \tag{3.7}$$

In particular: $V_t\left[\int_t^{t+h}\breve{\sigma}^2(s)ds\right] = A_2(\alpha,h)\tilde{\sigma}_c^2(t)+B_2(\alpha,h)$. In other words, the conditional variance of the average volatility over the lifetime of the option is in probability equivalent to its deterministic part:

$$V_t\left[\frac{1}{h}\int_t^{t+h}\breve{\sigma}^2(s)ds\right] \sim_{h\to+\infty} \frac{\gamma^2\tilde{\theta}h^{2\alpha-1}}{k^2\Gamma(\alpha+1)^2} \tag{3.8}$$

This result is important since it shows that, with a moderate level of long memory in the volatility process, $\alpha = 1/4$ say, this conditional variance is divided by ten when the time to maturity h of the option contract is multiplied by 100. By contrast, the same factor 100 would divide the variance in the short memory case. Since we know (see Renault and Touzi (1996)) that the volatility smile is produced by a Jensen effect due to the randomness of the average volatility $\frac{1}{h}\int_t^{t+h}\breve{\sigma}^2(s)ds$, the long memory model of volatility dynamics appears much better suited to reproduce observed volatility smiles in long term options (say with a time to maturity larger than six months) than the common short memory models.

Moreover, since we can rewrite (3.8) as:

$$V_t\left[\frac{1}{h}\int_t^{t+h}\breve{\sigma}^2(s)ds\right] \sim_{h\to+\infty} \frac{\gamma^2\tilde{\theta}}{k^{2\alpha+1}}\frac{(hk)^{2\alpha-1}}{\Gamma(\alpha+1)^2} \tag{3.9}$$

we can clearly disentangle two effects in the explanation of the volatility smile:

a) The first one, independent of the maturity, is simply produced by the unconditional variance $\gamma^2\tilde{\theta}/k^{2\alpha+1}$ of the spot volatility process.
b) The second one captures the erosion of the volatility smile when maturity increases. It is given by the term $(hk)^{2\alpha-1}$ where, for a given long memory parameter α, the time to maturity h is scaled by the mean reversion parameter k. The stronger is the mean reversion in the volatility process, the less pronounced the volatility smile will be for long horizons.

On the opposite, we have for infinitely short times to maturity:

$$A_2(h,\alpha) \sim_{h\to 0} \frac{\gamma^2 h^{2\alpha+3}}{(2\alpha+3)\Gamma(\alpha+2)^2}, \text{ and } B_2(h,\alpha) \sim_{h\to 0} \frac{\gamma^2\tilde{\theta}h^{2\alpha+3}}{2(2\alpha+3)\Gamma(\alpha+1)\Gamma(\alpha+3)}. \tag{3.10}$$

Therefore, the conditional variance $V_t\left[(1/h)\int_t^{t+h}\breve{\sigma}^2(s)ds\right]$ is, when h goes to zero, an infinitely small of order $h^{2\alpha+1}$. In other words, by contrast with the long horizon case, the volatility smile for very short term options will be, ceteris paribus, less pronounced in the case of long memory stochastic volatility than is the common short memory case .

About the possible explanations of observed volatility smiles, another warning is in order. It is often claimed that the convexity of the volatility smile is produced by the unconditional excess kurtosis of log-returns. While volatility persistence does produce unconditional kurtosis in general, the study of very short term options precisely shows that volatility persistence and unconditional kurtosis are two distinct phenomena and that the volatility smile does correspond to the first phenomenon.

To see this, let us consider for the sake of notational simplicity that the log-price has a zero deterministic drift and there is no leverage effect, that is the two Wiener processes W^S and W^X

are independent. By using the stationarity of σ^2, we obtain that the unconditional kurtosis of the return over the period $[t, t+h]$ is given by:

$$K(h) = \frac{\mathbb{E}\left[R_t^4(h)\right]}{\left(\mathbb{E}\left[R_t^2(h)\right]\right)^2} = 3\left[1 + \frac{\mathrm{Var}\left[\frac{1}{h}\int_t^{t+h}\sigma^2(u)du\right]}{(\mathbb{E}(\sigma^2))^2}\right]. \tag{3.11}$$

The limit cases $h \to 0$ and $h \to \infty$ are easy to deduce from (3.11):
a) First, since $\frac{1}{h}\int_t^{t+h}\sigma^2(u)du$ converges in mean-square towards $\sigma^2(t)$ when h tends to zero,

$$\lim_{h\to 0} K(h) = 3\left[1 + \frac{\mathrm{Var}(\sigma^2)}{(\mathbb{E}(\sigma^2))^2}\right] \tag{3.12}$$

where $\mathrm{Var}(\sigma^2)$ denotes the unconditional variance of the mean-square stationary spot volatility process. Formula (3.12) is actually an application to very short time intervals of a result well-known since Clark (1973): the excess kurtosis is equal to 3 times the squared coefficient of variation of the stochastic variance. This excess kurtosis effect persists in the very short term even though the volatility smile erases at rate $h^{2\alpha+1}$ as the conditional variance $V_t\left[\frac{1}{h}\int_t^{t+h}\sigma^2(u)du\right]$. This rate is actually also the rate of convergence of the kurtosis coefficient $K(h)$ towards its limit value (3.12).
b) Second, since $\frac{1}{h}\int_t^{t+h}\sigma^2(u)du$ converges in mean-square towards $\mathbb{E}(\sigma^2)$ when h tends to infinity, $\lim_{h\to+\infty} K(h) = 3$. This limit is reached at the speed $h^{2\alpha-1}$, that is the one which also governs the convergence to zero of the conditional variance $V_t\left[\frac{1}{h}\int_t^{t+h}\sigma^2(u)du\right]$. In other words, volatility smile and excess kurtosis assessments lead to the same conclusion in the very long term. Thanks to long memory, the convergence towards the log-normal case when the time to maturity increases is much slower.

The following proposition states that the kurtosis results described above remain valid in the general case of non-zero drifts and leverage effects. These results are actually deduced from the long memory properties of the volatility process and do not depend on the specific affine structure:

Proposition 3.3. *Let $\breve{K}(h)$ denote the approximate unconditional kurtosis coefficient of the log-return $R_t(h) = \log(S(t+h)/S(t))$ with σ^2 replaced by $\breve{\sigma}^2$. Then*

(i) $K(h) = 3\left[1 + \frac{\mathrm{Var}(\breve{\sigma}^2)}{(\mathbb{E}(\breve{\sigma}^2))^2}\right] + O(h^{2\alpha+1})$ *when h goes to zero,*
(ii) $K(h) = 3 + O(h^{2\alpha-1})$ *when h goes to infinity.*

3.2. **Memory properties.** The integrated volatility process is defined as

$$Z(t+1) = \int_t^{t+1} \breve{\sigma}^2(u)du = \theta + \int_t^{t+1}\left[\int_{-\infty}^{u} \frac{(u-s)^{\alpha-1}}{\Gamma(\alpha)}\tilde{\sigma}_c^2(s)ds\right]du.$$

Z(t) is a second order stationary process the spectral density f_Z of which is easily deduced from the spectral density $f_{\tilde{\sigma}^2}$ of $\tilde{\sigma}^2$:

Proposition 3.4. *Let f_Z and $f_{\tilde{\sigma}^2}$ be the spectral density of Z and $\tilde{\sigma}^2$ respectively. Then*

$$f_Z(\lambda) = \lambda^{-2\alpha} f_{\tilde{\sigma}^2}(\lambda)\frac{|e^{i\lambda}-1|^2}{\lambda^2}.$$

In particular for $\lambda \to 0$, $f_Z(\lambda) \sim \lambda^{-2\alpha} f_{\tilde{\sigma}^2}(0)$.

This implies that Z is a stationary long memory process of order α: its autocovariance function at lag h decays to zero with hyperbolic rate $h^{2\alpha-1}$ when h goes to infinity. While simple albeit tedious computations would allow to deduce from Proposition 3.4 the analog of Proposition 2.4 for characterizing the asymptotic behavior of the autocovariance function of the integrated volatility process $Z(t)$, we prefer to focus here on the autocovariance function of the expected integrated volatility process

$$Y(t) = \mathbb{E}_t Z(t+1) = \int_0^1 \mathbb{E}_t \left[\breve{\sigma}^2(t+u)\right] du.$$

As already announced, we expect that the memory properties of the process $Y(t)$ mimic the ones of Black-Scholes implied volatilities computed on fixed maturities.

It is worth noticing that the three processes $\breve{\sigma}^2(t)$ of spot volatility, $\mathbb{E}_t\breve{\sigma}^2(t+1)$ of expected spot volatility and $Y(t) = \mathbb{E}_t Z(t+1)$ of expected integrated volatility share the same equivalent of the autocovariance function for infinite lags:

Proposition 3.5. *When* $h \to +\infty$ *and for* $\alpha \in]0, 1/2[$, $cov[\breve{\sigma}^2(t), \breve{\sigma}^2(t+h)]$, $cov[\mathbb{E}_t\breve{\sigma}^2(t+1), \mathbb{E}_{t+h}\breve{\sigma}^2(t+h+1)]$ *and* $cov[Y(t), Y(t+h)]$ *can all be written*

$$Ch^{2\alpha-1} + o(h^{2\alpha-1}) \text{ with } C = \frac{\tilde{\theta}\gamma^2}{k^2} \frac{\Gamma(1-2\alpha)}{\Gamma(1-\alpha)\Gamma(\alpha)}.$$

3.3. **Quadratic variation.** A number of recent papers (see Andersen, Bollerslev and Diebold (2003) and references therein) have put forward non parametric measurement of volatility through quadratic variation estimated from high frequency data. More precisely, for a diffusion-type model of log-prices $R(t) = \ln S(t)$ as deduced from (2.1):

$$dR(t) = \mu(t)dt + \sigma(t)dW^S(t)$$

the realized volatility $\sum_{i=1}^n [R(t_i) - R(t_{i-1})]^2$ over a time grid $t = t_0 < t_1 < \cdots < t_i < t_{i+1} < \cdots < t_n = t+1$ converges in mean square when $\max_{1 \le i \le n} |t_i - t_{i-1}|$ goes to zero, towards $\int_t^{t+1} \sigma^2(u)du$. Barndorff-Nielsen and Shephard (2002) characterize the rate of convergence and the asymptotic probability distribution of: $\int_t^{t+1} \sigma^2(u)du - \sum_{i=1}^n [R(t_i) - R(t_{i-1})]^2$.
In this respect, realized volatility is a non parametric estimator of $\int_t^{t+1} \sigma^2(u)du$ with a well settled asymptotic theory. This may be helpful in particular to make inference about the volatility process since it means that integrated volatility is something we can asymptotically "observe", with a well characterized measurement error distribution.

In our case, the problem is that the diffusion process we are interested in for statistical inference is the underlying (short memory) state variable process $\tilde{\sigma}^2$. Thus, a prerequisite to use statistical inference strategies based on quadratic variation is to get a measurement of $\int_t^{t+1} \tilde{\sigma}^2(u)du$. This is actually possible thanks to the following result:

Proposition 3.6. *Let* $Z(t+1) = \int_t^{t+1} \breve{\sigma}^2(s)ds$ *where* $\sigma^2(t) = \theta + I^{(\alpha)}(\tilde{\sigma}_c^2)(t)$ *is defined as in Section 2, then* $I^{(-\alpha)}(Z-\theta)(t+1) = \int_t^{t+1}(\tilde{\sigma}^2(s) - \tilde{\theta})ds$.

In other words, if high frequency data allow to observe Z and then to estimate its mean θ and its memory parameter α, a fractional derivation will lead to recover the integrated volatility associated to the standard short memory underlying model:

$$\int_t^{t+1} \tilde{\sigma}^2(s)ds = \tilde{\theta} + \int_t^{t+1} \tilde{\sigma}_c^2(s)ds.$$

Then statistical methods well suited for this affine model (see in particular Bollerslev and Zhou (2002) for a method of moments) can be used.
Once more, this convenient result is made possible by the commutation property between fractional integration/derivation operator and standard integration. Such a commutation does not work in a log-Ornstein-Uhlenbeck volatility model as described in Comte and Renault (1998).

4. Statistical inference

4.1. **Discretization methods.**

4.1.1. *Discretization of the CIR equation (2.5).* A simple Euler discretization scheme is not well-suited in the case of the CIR equation. We resort here to a method proposed by Rogers (1995), which is well-founded and user-friendly. Its only drawback is to impose some constraints on the CIR parameters k, $\tilde{\theta}$ and γ. It can only accommodate parameter values such that $d = 4k\tilde{\theta}/\gamma^2$ is an integer (see also Diop (2003)) for the general issue of simulating CIR processes).

4.1.2. *Discretization of the fractional integration.* We present in this section a discretization scheme for fractional integrals, studied by Carmona, Coutin and Montseny (2000) for Gaussian processes. The aim of all the transformations below are to exhibit a *recursive* discretization method. Indeed, a naive Euler scheme for fractional integrals leads to non-recursive formulas, see Comte and Renault (1998).
The method is the result of four consecutive ideas which can be described as follows:

(1) Approximate the operator $I^{(\alpha)}$ by the truncated operator $I_0^{(\alpha)}$ which associates to a process X the process: $I_0^{(\alpha)}(X)(t) = \int_0^t [(t-s)^{\alpha-1}/\Gamma(\alpha)]X(s)ds$. By Proposition 2.1, this approximation is consistent when t goes to infinity, with an absolute error of order $t^{\alpha-1/2}$.
(2) Use Laplace inverse transform to write $(t-s)^{\alpha-1} = (1/\Gamma(1-\alpha))\int_0^\infty x^{-\alpha}e^{-x(t-s)}dx$, and apply Fubini's theorem:

$$I_0^{(\alpha)}(X)(t) = \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\int_0^\infty x^{-\alpha}\left(\int_0^t e^{-x(t-s)}X(s)ds\right)dx.$$

This representation, known as the diffusive representation of fractional integrals, is used by Carmona, Coutin and Montseny (2000) to write that any function f, continuous on $[0,T]$ satisfies:

$$I_0^{(\alpha)}(f)(t) = \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\int_0^\infty x^{-\alpha}\Psi(x,t,f)dx, \tag{4.1}$$

where $\Psi(x,t,f)$ is the linear operator defined in Coutin and Pontier (2001) for continuous functions on $[0,T]$ by

$$\Psi(x,t,f) = \int_0^t e^{-x(t-s)} f(s)ds. \tag{4.2}$$

(3) Use a geometric subdivision of $\mathbb{R}^+$, $x_i = r^i$, $i = -n, -n+1, \ldots, 0, 1, \ldots, n-1$ (for some $r \in]1,2[$ and n going to infinity) to discretize the frequency integral (the integral w.r.t. x):

$$\int_0^{+\infty} x^{-\alpha}\psi(x,t,f)dx \# \sum_{i=-n}^{n-1} \int_{x_i}^{x_{i+1}} x^{-\alpha}\psi(x,t,f)dx.$$

Then, by considering the probability density function over the interval $[x_i, x_{i+1}]$ defined by:

$$g_i(x) = \frac{x^{-\alpha}}{c_i}, \quad c_i = \int_{x_i}^{x_{i+1}} x^{-\alpha}dx = \frac{(r^{1-\alpha}-1)}{1-\alpha} r^{(1-\alpha)i} \tag{4.3}$$

and the mean value η_i over this interval for this density function:

$$\eta_i = \frac{1}{c_i}\int_{x_i}^{x_{i+1}} x^{1-\alpha}dx = \frac{1-\alpha}{2-\alpha}\frac{r^{2-\alpha}-1}{r^{1-\alpha}-1} r^i, \tag{4.4}$$

we use the mean value approximation for the frequency integral of interest:

$$\int_{x_i}^{x_{i+1}} x^{-\alpha}\psi(x,t,f)dx \# c_i\psi(\eta_i,t,f).$$

Then the volatility process on a discrete time grid will be recovered from the path of the process X by:

$$(\sigma^2)^{r,n}(t_j) = \theta + \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\sum_{i=-n}^{n-1} c_i\Psi(\eta_i, t_j, X).$$

(4) Use a time discretization to compute recursively $\psi(\eta_i, t_j, X)$ along a discrete time sample t_j, $j = 1, 2, \ldots, N$. In order to do this, it is worth noticing that:

$$\Psi(x, t_{j+1}, f) = e^{-x(t_{j+1}-t_j)}\Psi(x,t_j,f) + \int_{t_j}^{t_{j+1}} e^{-x(t_{j+1}-s)} f(s)ds.$$

This suggests the following recursive discretization:

$$\psi(x,t_{j+1},f) \# e^{-x\Delta}\Psi(x,t_j,f) + f(t_j)\frac{1-e^{-x\Delta}}{x},$$

written, for sake of simplicity, in the simplest case of regularly spaced observations: $t_{j+1} = t_j + \Delta$, $j = 0, 1, \ldots, N$.

To summarize, from observed values $X(t_j)$, $j = 0, 1, \ldots, N$ of the underlying process X, we recover a time discretization of the function ψ:

$$\Psi^\Delta(x, t_{j+1}, f) \quad = \quad \Psi^\Delta(x,t_j,f)e^{-x\Delta} + f(t_j)\frac{1-e^{-x\Delta}}{x}, \; \Psi^\Delta(x,t_0,f) = 0. \tag{4.5}$$

and by plugging this discretization in step (3), a discretization of the volatility process

$$(\sigma^2)^{r,n,\Delta}(t_j) = \theta + \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\sum_{i=-n}^{n-1} c_i\Psi^\Delta(\eta_i, t_j, X) \tag{4.6}$$

Then we have the following assessment of the rate of convergence of the discretization (4.6):

Proposition 4.1. *For any $\beta \in]0, \frac{1}{2}[$, the random variable*

$$\sup_{(r,n,\Delta)\in[1,2]\times\mathbf{N}^*\times]0,1]} \frac{1}{(r-1)^2(1+\Delta^\beta)+r^{-\alpha n}} \sup_{j=1,\dots,N} |(\sigma^2)^{r,n,\Delta}(t_j) - I_0^{(\alpha)}(\tilde{\sigma}^2-\tilde{\theta})(t_j)|$$

belongs to $\mathbb{L}^p$ for all $p \geq 1$.

Let us emphasize that such a recursive discretization scheme is a necessary tool if one wants to do statistical inference from a large discrete time sample of observations $S(t_j)$, $j = 0, 1, \dots, N$. It is well-know indeed that the main problem of fractional calculus is its non-recursive aspects and the fact that a standard discretization would require the use of all the previous points to compute the last one, at each time. Formula (4.6) uses only deterministic stored values η_j, c_j, and recursively computable $\Psi^\Delta(\eta_i, t_j, \tilde{\sigma}^2 - \tilde{\theta})$, $\tilde{\sigma}(t_j)$.

4.1.3. *Discretization of the fractional derivation.* For estimation purpose, we shall need approximations of the derivation operator, for which similar schemes can be developed.

Let f be a function Hölder continuous of order β. According to formula (5.6) p.99 and (5.57) p.109 in Samko et al. (1993) the fractional derivatives of order α of f, are given by

$$I^{(-\alpha)}(f)(t) = \frac{\alpha}{\Gamma(1-\alpha)} \int_{-\infty}^t \frac{f(t)-f(s)}{(t-s)^{\alpha+1}} ds$$

and

$$I_0^{(-\alpha)}(f)(t) = \frac{\alpha}{\Gamma(1-\alpha)} \left[\frac{f(t)}{\alpha t^\alpha} + \int_0^t \frac{f(t)-f(s)}{(t-s)^{\alpha+1}} ds \right].$$

Then we shall use the approximate derivation to be applied to $Z - Z_0$:

$$Z^{\Delta,-\alpha,r,n}(t_j) = \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \sum_{i=-n}^{n-1} c'_i \Xi^\Delta(\eta'_i, t_j, Z - Z_0),$$

where Ξ^Δ is computed by

$$\Xi^\Delta(x, 0, f) = 0,\ \Xi^\Delta(x, t_{j+1}, f) \quad = \quad e^{-x\Delta}\Xi^\Delta(x, t_j, f) + (f(t_{j+1}) - f(t_j)).$$

and c'_i and η'_i are given by

$$c'_j = r^{\alpha j}\frac{r^\alpha - 1}{\alpha},\quad \eta'_j = r^{(\alpha+1)j}\frac{r^{\alpha+1}-1}{c'_j(\alpha+1)}. \tag{4.7}$$

We can prove then (see Appendix) that there exists a constant C such that for any f Hölder continuous of index $\beta > \alpha$, for $r \in]1, 2]$, $n \in \mathbf{N}^*$ and $\Delta \in]0, 1[$

$$\sup_{j=0,\dots,N} |\sum_{i=-n}^{n-1} c'_i \Xi^\Delta(\eta'_i, t_j, f) - I_0^{(-\alpha)}(f)(t_j)| \leq C[(1+\Delta^{(\beta+\alpha)/2})(r-1)^2 + r^{-n\min(\alpha,(\beta-\alpha)/2)}]. \tag{4.8}$$

Since f in our problem is $Z_t - Z_0$ (or $\sigma^2(t) - \sigma^2(0)$), we know that the considered paths are $\alpha + 1/2 - \epsilon$-Hölder, for any $\epsilon > 0$ so that $\beta - \alpha = 1/2 - \epsilon > 0$.

4.2. **Statistical strategy.** We describe here a statistical strategy based on observations $\sigma^2(t_j)$, $j = 1, \ldots, N$ of the volatility process obtained from high-frequency data under the assumptions of Proposition 2.6. A similar strategy could easily be settled from observations of the integrated volatility process $\int_{t_j}^{t_{j+1}} \sigma^2(u)du$, as described in Section 3.4. In both cases, a comprehensive asymptotic theory should also take into account the measurement error as in Barndorff-Nielsen and Shephard (2002). This issue is left for future work. For sake of expositional simplicity, we focus here on the case without leverage ($\rho = 0$). The approach could easily be extended to accommodate the estimate of ρ.

1. The mean θ and the fractional integration parameter α of the stationary long memory process $\sigma^2(t)$ can be estimated by standard semiparametric procedures. We choose here to simply use the empirical mean $\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n} \sigma^2(t_j)$ and the Geweke and Porter-Hudak (1983) log-periodogram regression estimator $\hat{\alpha}$ as revisited by Robinson (1996). Let us recall that the idea of log-periodogram regression is based on the equivalent $f(\lambda) \sim C\lambda^{-2\alpha}$ of the spectral density in the neighborhood of zero. For a choice of Fourier frequencies $\lambda_k = 2k\pi/n$, for $k = \ell, \ell+1, \ldots, m$ and an estimation of the log-periodogram as

$$I_n(\lambda_k) = \frac{1}{2\pi n}\left|\sum_{j=1}^{n}(\sigma^2(t_j) - \hat{\theta})e^{it_j\lambda_k}\right|^2$$

the approximated regression model $\ln(I_n(\lambda_k)) \sim \ln(C) - 2\alpha\ln(\lambda_k)$, $\quad k = \ell, \ell+1, \ldots, m$ provides an OLS estimator $\hat{\alpha}$ of α. According to Robinson (1996), fine tuning of the trimming parameters ℓ and m is required to get good properties of $\hat{\alpha}$. Moreover, it is worthwhile to realize that the standard asymptotic theory for the Geweke and Porter-Hudak estimator has only been developed for Gaussian processes, which is not the case of the volatility process. However, in the same way as Velasco (2000) provides results of consistency of log-periodogram based estimators of the short memory parameters for very general linear processes, we can hope to get by this way a consistent estimator $\hat{\alpha}$ of α. The short scale Monte Carlo study of Section 4.3 gives some support to this claim.

2. We use the estimations $\hat{\theta}$ and $\hat{\alpha}$ and the approximation of the fractional derivation operator described in the previous subsection to get implied values of $\tilde{\sigma}_c^2(t_j) = I^{(-\alpha)}(\sigma^2 - \theta)(t_j)$. Since the process $\tilde{\sigma}^2$ follows an affine process

$$d\tilde{\sigma}^2(t) = -k\tilde{\sigma}^2(t)dt + \gamma\sqrt{\tilde{\sigma}^2(t) + \tilde{\theta}}dW^{\tilde{\sigma}^2}(t),$$

it is easy to derive consistent asymptotically normal estimators of the unknown parameters $k, \tilde{\theta}$ and γ. We choose here to estimate these three parameters by a very simple method of moments based on the following three theoretical moments:

$$m_2 = \mathbb{E}(X^2(t)) = \frac{\gamma^2\tilde{\theta}}{2k}, \quad c_q = \operatorname{cov}[X(t), X(t+q\Delta)] = \frac{\gamma^2\tilde{\theta}}{2k}e^{-kq\Delta}, \quad m_3 = \mathbb{E}(X^3(t)) = \frac{\gamma^4\tilde{\theta}}{2k^2}.$$

By denoting by $\hat{m}_2, \hat{c}_q$ and $\hat{m}_3$ the empirical counterparts of m_2, c_q and m_3 respectively, this gives the following estimators:

$$\hat{\tilde{\theta}} = \frac{2\hat{m}_2^2}{\hat{m}_3}, \quad \hat{k} = \frac{1}{q\Delta} \ln \left[\frac{\hat{m}_2}{\hat{c}_q} \right] \quad \text{and} \quad \hat{\gamma} = \left(\frac{2\hat{m}_2 \hat{k}}{\hat{\tilde{\theta}}} \right)^{1/2}. \tag{4.9}$$

4.3. **Simulation results.**

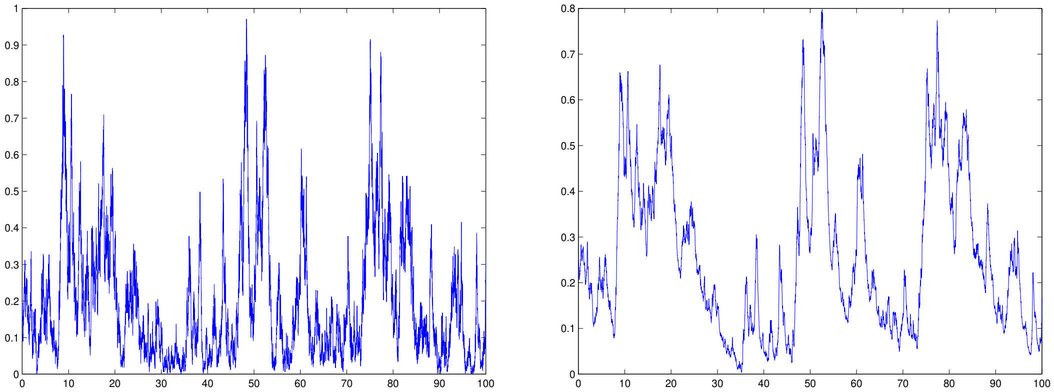

FIGURE 1. Left: A path of CIR generated with the multivariate method ($d = 3$) and parameters $k = 1, \gamma = 0.5, \Delta = 0.01, n = 10000$. Right: Simulated volatility by fractional integration with $\alpha = 0.25$ of the above path, $\theta = 0.2$.

The CIR process $\tilde{\sigma}^2$ has been generated with time intervals Δ using the multidimensional method with dimension $d = 3$. For a choice of d, γ and k, the mean is given by $\tilde{\theta} = d\gamma^2/(4k)$ and the variance by $c_{\tilde{\sigma}^2}(0) = \tilde{\theta}\gamma^2/(2k) = d\gamma^4/8k^2$. Fractional integration is performed to generate σ^2 with the same step Δ using formula (4.6) applied to $\tilde{\sigma}^2$ centered by its empirical mean (instead of the theoretical mean) with a value of θ that ensures that all paths are positive. Comparison of both path in Figure 1 and illustrates the smoothing properties of fractional integration.

The discretized fractional integration operator has been tested on deterministic functions and appears to crucially depend on the choice of r (the step of the geometric subdivision): when the sample is small with large step, r must be small (around 1.02-1.04) and for large samples with smaller steps, r must be larger (around 1.3). For a well chosen r, it performs better than the naive method. Besides, we tested the discretized fractional derivation by checking that when applied to a fractionally integrated process, it delivers the original process. Here again, the choice of r must be done carefully. But in any case, the recursive method is considerably faster than the naive one.

The paths of σ^2 are used to construct the paths of the log prices $R(t) = \ln S(t)$ with constant mean $\mu = 0$ with the recursion:

$$R^\Delta((k+1)\Delta) = R^\Delta(k\Delta) + \sqrt{\sigma^2(k\Delta)}\sqrt{\Delta}\varepsilon_{(k+1)\Delta}$$

where $\sqrt{\Delta}\varepsilon_{(k+1)\Delta} \stackrel{\mathcal{L}}{=} W^S((k+1)\Delta) - W^S(k\Delta)$ are i.i.d. $\mathcal{N}(0, \Delta)$. Figure 2 plots the resulting paths computed and the one of σ^2.

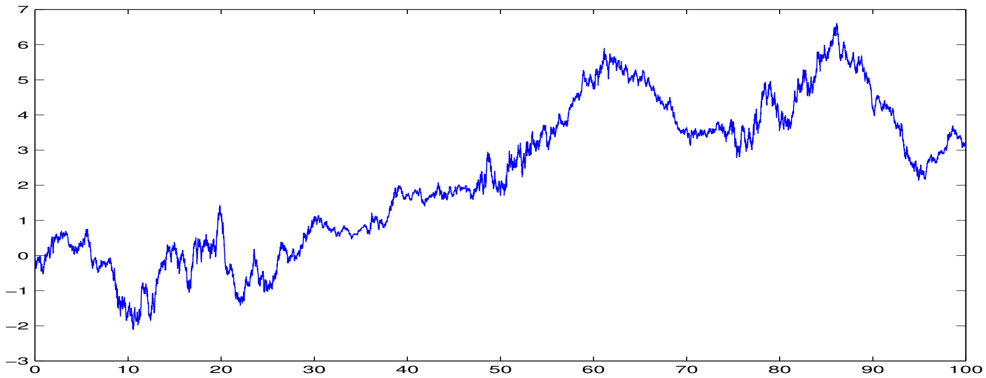

FIGURE 2. Simulated log-prices with the above volatility and mean $\mu = 0$.

Next, the method of estimating the volatility or the integrated volatility by using sums by block of squared increments of the log-prices is studied. More precisely, Figure 3 shows the difference between the true volatility $\sigma^2(k\Delta)$, for $k = 1, \dots, n$ and the values for $t = \Delta, 2\Delta, \dots, n\Delta$

$$\hat{\sigma}^2_{n,p}(t) \quad = \quad \frac{1}{p\Delta} \sum_{j=([t/(p\Delta)]-1)p+1}^{[t/(p\Delta)]p} \left[R^{\Delta}(j\Delta) - R((j-1)\Delta)\right]^2$$

for $k = 1, \dots, n/p$, with $p = 50$ and $n = 10000$.

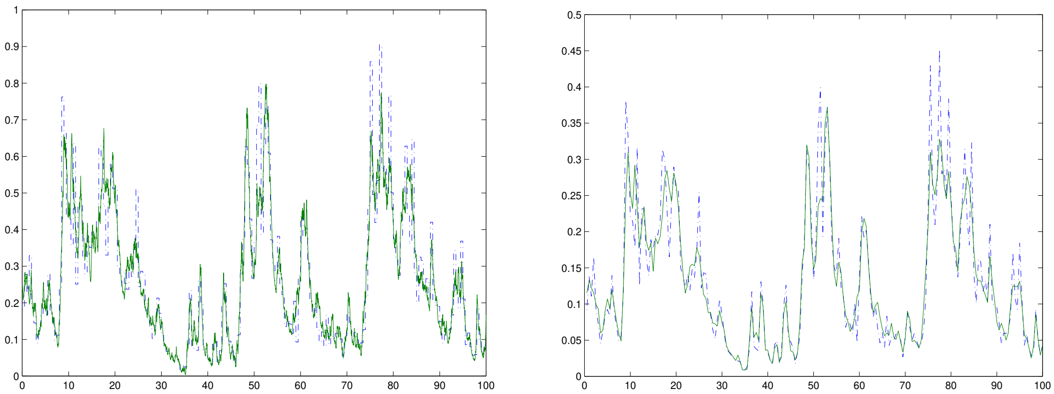

FIGURE 3. Left: Estimated volatility using blocks of size $p = 50$ (dotted line) compared with the true volatility (full line). Right: Estimated integrated volatility using blocks of size $p = 50$ (dotted line) compared with the true (cumulated) volatility (full line).

Figure 3 shows the difference between the following approximation of the integrated volatility: $\sum_{j=(k-1)p+1}^{kp} \sigma^2(j\Delta)$ for $k = 1, \dots n/p$ and the approximation of the quadratic variation $\sum_{j=(k-1)p+1}^{kp} \left[R^{\Delta}((j+1)\Delta) - R(j\Delta)\right]^2$ for $k = 1, \dots, n/p$, still with $p = 50$ and $n = 10000$ and on the same path.

In both cases, further simulations show that a smaller value of p ($p = 5, 10, 15$) would produce a too noisy measurement volatility. We also applied the fractional derivation of order α to the estimated volatility process and compared the paths to the initial paths of the CIR (with the same step). The curves coincide quite well and the results are very convincing.

To test the robustness of the log-periodogram regression to non-gaussianity, we estimated α from the complete samples of simulated volatilities using this method. We plot in Figure 4 the histogram of the estimated values of α for 100 samples of length 10 000. The results are quite convincing but the method is very sensitive to the choice of the trimming parameter m. We chose $m/n = 36.35\%$ while the $\ell = 1806$ first terms of all samples have been dropped out. This experiment allows to conclude that the procedure is robust to our case of non-gaussianity.

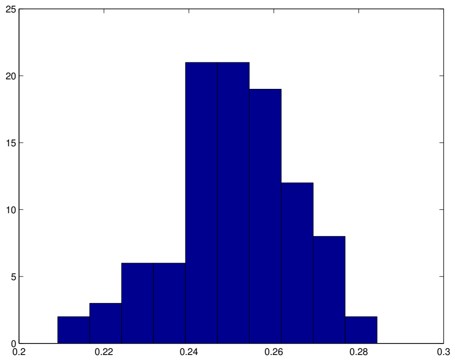

FIGURE 4. Histogram of estimated α over 100 samples of length $n = 10000$ with step $\Delta = 0.04$ for a true value $\alpha = 0.25$, $(k, \gamma, \theta, \tilde{\theta}) = (1, 0.5, 0.3, 0.3125)$, $d = 5$, $r = 1.3$. Mean $= 0.2506$, Standard deviation $= 0.015$.

The results of the global procedure are reported in Table 1 below for $n = 10000$ observations corresponding to 400 days of observations with 25 observations per day. Here, the CIR is generated with $d = 5$, discretized fractional integration is done with $r = 1.3$, quadratic variations are computed using blocks of size 25 to estimate the centered volatility (and not integrated volatility), the log-periodogram regression uses the complete samples with $m = 0.39$ times the number of observations (about 400 remaining), and discretized fractional derivation takes $r = 1.025$.

Parameter	θ	α	γ	k	$\tilde{\theta}$
True Value	0.3	0.25	0.5	1	0.3125
Mean	0.304	0.252	0.318	1.227	0.3240
Standard Deviation	0.013	0.059	0.080	0.430	0.242

TABLE 1. RESULTS OF THE ESTIMATION OF THE PARAMETERS OF EQUATIONS n =10 000 AND TIME INTERVAL $\Delta = 0.04$.

The results are satisfactory even though the true unknown values of parameters define a volatility process with less mean reversion (smaller k) and more volatility (higher γ). This kind of finite sample bias is common for persistent time series (see e.g. Pastorello et al. (2000)). Moreover, the rather high values of standard errors could easily be reduced by using a larger set of moment conditions than the minimum one (4.9).

While this Monte Carlo experiment has been performed with high mean level of volatility (around 30 percents), it is worth considering also a case of a volatility process with more mean reversion, less instantaneous variance and a smaller mean level. The true unknown values considered in Table 2 below were provided by an empirical study of Moreaux et al. (1998).

With these values, we first simulate 100 paths of the underlying short memory volatility processes $\tilde{\sigma}^2$ of length 10000 with step 0.04. We notice that in this low variance case, the size of the blocks used for the quadratic variation procedure had to be chosen much smaller than in the previous higher variance case: namely, we took blocks of size 4 (instead of 25 before). This implies longer samples with smaller step than previously, so that the choice of r in the fractional derivation must be slightly increased to $r = 1.035$. We tested the moment method (see formula (4.9) on the simulated path of $\tilde{\sigma}^2$ and found perfect results for $q = 1$ given in Table 2.

Parameter	γ	k	$\tilde{\theta}$
True Value	0.85	15	0.06
Mean	0.850	15.025	0.060
Standard Deviation	0.019	0.471	0.002

Table 2. Results of the estimation of the parameters of the CIR using the (unobservable) 100 paths with length 10 000 and step 0.04, of $\tilde{\sigma}^2$ generated by the multivariate procedure with $d = 5$.

After application of (2.9) on the 100 simulated paths, all the computed values of $\sigma^2(t)$ appear to be positive. The global procedure gives the results stated in Table 3. In other words, the global procedure performs quite well, and testing simulated data nearer of real data leads to some practical recommendations on the size of blocks and the choice of r.

Parameter	θ	α	γ	k	$\tilde{\theta}$
True Value	0.06	0.25	0.85	15	0.06
Mean	0.06	0.252	0.997	14.513	0.048
Standard Deviation	0.049	0.020	0.083	1.470	0.005

Table 3. Results of the estimation of the parameters over 100 samples with length 10 000 and step 0.04.

5. Appendix: Proofs

We first give a lemma summarizing some classical properties of the CIR process that will be useful in the following proofs.

Lemma 5.1. *Let $\tilde{\sigma}^2$ denote the associated solution of (2.5) with initial condition $X_0 = x_0$, then*
1. $\tilde{\sigma}^2$ admits a modification with sample paths Hölder continuous for any index a strictly less than $\frac{1}{2}$.

2. If the process is backwardly defined for negative times, $\sup_{s\in]-\infty,-1]}\tilde{\sigma}_s^2|s|^{-\frac{1}{2}}$ *belongs to* $\mathbb{L}^p$ *for any* p.

Proof of Proposition 2.1. Observe that

$$\|\Delta\|_2^2 := \left\|(\tilde{\sigma}^2)^{(\alpha)}(t) - \int_0^t \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}\tilde{\sigma}^2(s)ds\right\|_2^2 = \int_{-\infty}^0\int_{-\infty}^0 \frac{(t-s)^{\alpha-1}(t-u)^{\alpha-1}}{\Gamma(\alpha)^2}\gamma_{\tilde{\sigma}^2}(|s-u|)dsdu.$$

Setting $s = tx'$ and $u = ty'$ and then $x = 1-x', y = 1-y'$ gives with the assumption on $\gamma_{\tilde{\sigma}^2}$ that

$$\begin{aligned}\|\Delta\|_2^2 &\leq \frac{t^{2\alpha}\gamma_{\tilde{\sigma}^2}(0)}{\Gamma(\alpha)^2}\int_1^{+\infty}\int_1^{+\infty} x^{\alpha-1}y^{\alpha-1}e^{-kt|x-y|}dxdy \\ &\leq \frac{2t^{2\alpha}\gamma_{\tilde{\sigma}^2}(0)}{\Gamma(\alpha)^2}\int_1^{+\infty} x^{\alpha-1}e^{ktx}\left(\int_x^{+\infty} y^{\alpha-1}e^{-kty}dy\right)dx.\end{aligned}$$

One integration by parts shows that the integrals is of order $1/t$. □

Proof of Proposition 2.3 and 2.4. First note that

$$C(\alpha) := \int_0^{+\infty} u^{\alpha-1}(u+1)^{\alpha-1}du = \int_0^1 x^{-2\alpha}(1-x)^{\alpha-1}dx = \frac{\Gamma(1-2\alpha)\Gamma(\alpha)}{\Gamma(1-\alpha)} \tag{5.1}$$

using a well known formula linking the $\beta(a,b)$ function and the Γ function. Second, we prove that

$$c_{\breve{\sigma}^2}(h) = \frac{\tilde{\theta}\gamma^2}{2k}\frac{\Gamma(1-2\alpha)}{\Gamma(1-\alpha)\Gamma(\alpha)}\int_{-\infty}^{\infty}|h+z|^{2\alpha-1}e^{-k|z|}dz. \tag{5.2}$$

Setting $h = 0$ in (5.2) gives the result of Proposition 2.3 since the last integral equals $2\Gamma(2\alpha)/k^{2\alpha}$. To prove (5.2), note that (2.10) and (2.11) imply $c_{\breve{\sigma}^2}(h) = [\tilde{\theta}\gamma^2/2k\Gamma(\alpha)^2]\int_{-\infty}^t\int_{-\infty}^{t+h}(t-u)^{\alpha-1}(t+h-v)^{\alpha-1}e^{-k|u-v|}dudv$. Then according to the change of variables $z = u-v$ and $r = t-u$ we obtain

$$c_{\breve{\sigma}^2}(h) = \frac{\tilde{\theta}\gamma^2}{2k\Gamma(\alpha)^2}\int_{-\infty}^{+\infty}\left(\int_0^{+\infty}(r)^{\alpha-1}(r+h+z)_+^{\alpha-1}dr\right)e^{-k|z|}dz. \tag{5.3}$$

We observe that, with formula (5.1), $\int_0^{+\infty}(r)^{\alpha-1}(r+h+z)_+^{\alpha-1}dr = |h+z|^{2\alpha-1}(\Gamma(1-2\alpha)\Gamma(\alpha)/\Gamma(1-\alpha))$. Then plugging it into (5.3) yields (5.2).

To prove the result, it remains to prove that, near 0

$$\int_{-\infty}^{\infty}|h+z|^{2\alpha-1}e^{-k|z|}dz = \int_{-\infty}^{\infty}|z|^{2\alpha-1}e^{-k|z|}dz + \frac{k}{\alpha(2\alpha+1)}|h|^{2\alpha+1} + O(h^2). \tag{5.4}$$

Using the change of variable $x = z+h$, we obtain $\int_{-\infty}^{\infty}|h+z|^{2\alpha-1}e^{-k|z|}dz = \int_{-\infty}^{\infty}|x|^{2\alpha-1}e^{-k|x-h|}dx$. Splitting the previous integral with respect to the sign of $(x-h)$ we obtain

$$\begin{aligned}\int_{-\infty}^{\infty}|h+z|^{2\alpha-1}e^{-k|z|}dz &= 2\int_0^{+\infty}x^{2\alpha-1}e^{-kx}dx\cosh(kh) \\ &\quad + \int_0^h|x|^{2\alpha-1}e^{-k(h-x)}dx - \int_0^h|x|^{2\alpha-1}e^{+k(h-x)}dx.\end{aligned}$$

Using the change of variables $y = \frac{x}{h}$ in the two previous integrals over $[0,h]$ we obtain

$$\int_{-\infty}^{\infty}|h+z|^{2\alpha-1}e^{-k|z|}dz = 2\left[\int_0^{+\infty}x^{2\alpha-1}e^{-kx}dx\cosh(kh) - |h|^{2\alpha}\int_0^1|y|^{2\alpha-1}\sinh(kh(1-y))dy\right].$$

Then, formula (5.4) is obtained as a consequence of the Taylor expansion of cosh and sinh and the dominated convergence theorem.
We use the same tools for the order near infinity. It remains to prove that near ∞

$$\int_{-\infty}^{\infty} |h+z|^{2\alpha-1} e^{-k|z|} dz = \frac{2}{k}|h|^{2\alpha-1} + O(|h|^{2\alpha}). \tag{5.5}$$

But (5.5) is an consequence of the Taylor expansion of $|1+\frac{z}{h}|^{\alpha-1}$, the Lebesgue dominated convergence theorem and $\int_{-\infty}^{\infty} |h+z|^{2\alpha-1}e^{-k|z|}dz = |h|^{2\alpha-1}\int_{-\infty}^{+\infty}(1+z/h)^{2\alpha-1}e^{-k|z|}dz$. $\square$
Proof of Proposition 2.5: $f_{\breve\sigma^2}$ is easily computed as the Fourier Transform of $c_{\breve\sigma^2}(h)$ given by (2.11).

$$\begin{aligned}
\int e^{-i\lambda h} c_{\breve\sigma^2}(h)dh &= \int_0^{+\infty}\int_0^{+\infty} \frac{x^{\alpha-1}y^{\alpha-1}}{\Gamma(\alpha)^2}\left(\int e^{-i\lambda h} c_{\breve\sigma^2}(h+y-x)dh\right)dxdy \\
&= \int \frac{e^{-i\lambda x}x^{\alpha-1}}{\Gamma(\alpha)}dx \int \frac{e^{i\lambda y}y^{\alpha-1}}{\Gamma(\alpha)}dy \int e^{-i\lambda v} c_{\breve\sigma^2}(v)dv, \quad v = h+y-x \\
&= \left|\int \frac{e^{-i\lambda x}x^{\alpha-1}}{\Gamma(\alpha)}dx\right|^2 f_{\breve\sigma^2}(\lambda) = \frac{1}{\lambda^{2\alpha}}\left|\int \frac{e^{-iu}u^{\alpha-1}}{\Gamma(\alpha)}du\right|^2 f_{\breve\sigma^2}(\lambda) = \frac{f_{\breve\sigma^2}(\lambda)}{\lambda^{2\alpha}}
\end{aligned}$$

since from Samko et al. (1993) p.137 $\int_0^{+\infty} t^{\alpha-1}e^{-zt}dt = \Gamma(\alpha)/z^\alpha$, for $z \neq 0$ and $0<\alpha<1$ when $Re(z)=0$.$\square$
Proof of Proposition 2.6. We start with the case $\mu = 0$ and $\rho = 0$. If $R(t) = R_0(t) = \int_0^t \sigma(u)dW^S(u)$, the proof is given in Comte and Renault (1998), Proposition 5.1. This proof uses the independence of σ and W^S, but it can easily be extended as follows. If $m = [tN/T]$, just write

$$\mathbb{E}\left[\hat\sigma^2_{N,p}(t) - \breve\sigma^2(t)\right]^2 \le 2\left(\mathbb{E}\left[\hat\sigma^2_{N,p}(t) - \breve\sigma^2(t_m)\right]^2 + \mathbb{E}\left[\breve\sigma^2(t_m) - \breve\sigma^2(t)\right]^2\right).$$

Then the first term can be computed as in Comte and Renault (1998) using only predictability and not independence (by conditioning):

$$\begin{aligned}
\mathbb{E}[\hat\sigma^2_{N,p}(t) - \breve\sigma^2(t_m)]^2 &= \frac{2N^2}{p^2T^2}\sum_{k=m-p+1}^{m}\int\int_{[t_{k-1},t_k[^2}(c_{\breve\sigma^2}(u-v) - c_{\breve\sigma^2}(0))dudv + 2\frac{\mathbb{E}\sigma^4}{p} \\
+\frac{N^2}{p^2T^2}\int\int_{[t_{m-p},t_m[^2}&(c_{\breve\sigma^2}(u-v) - c_{\breve\sigma^2}(0))dudv - \frac{2N}{pT}\int_{[t_{m-p},t_m[^2}(c_{\breve\sigma^2}(t_{m-p+1}-s) - c_{\breve\sigma^2}(0))ds.
\end{aligned}$$

Using that $c_{\breve\sigma^2}(h) - c_{\breve\sigma^2}(0) = Ch^{2\alpha+1} + O(h^2)$ for h small leads to the order $2\mathbb{E}(\breve\sigma^4)/p + K(pT/N)^{2\alpha+1}(1+1/p^{2\alpha+2}) + O((pT/N)^2)$ with K proportional to C and using that $t_k - t_{k-1} = T/N$ and $|t_m - t_{m-p}| = pT/N$.
The additional term is $\mathbb{E}\left[\breve\sigma^2(t_m) - \breve\sigma^2(t)\right]^2 = 2c_{\breve\sigma^2}(0) - 2c_{\breve\sigma^2}(|t_m - t|)$ so that as $|t_m - t| \le T/N$, we find that the expectation is less than $C(T/N)^{2\alpha+1}$.
Next if $R(t) = \int_0^t \mu(u)du + R_0(t)$, then we have to bound

$$\mathbb{E}_1 = \mathbb{E}\left[\frac{N}{pT}\sum_{k=t_{m-p+1}}^{t_m}\left(\int_{t_{k-1}}^{t_k}\mu(u)du\right)^2\right]^2 \text{ and } \mathbb{E}_2 = \mathbb{E}\left[\frac{N}{pT}\sum_{k=t_{m-p+1}}^{t_m}(R_0(t_k) - R_0(t_{k-1}))\int_{t_{k-1}}^{t_k}\mu(u)du\right]^2.$$

Using that $\left(\int_{t_{k-1}}^{t_k} \mu(u)du\right)^4 \leq (t_k - t_{k-1})^3 \int_{t_{k-1}}^{t_k} \mu^4(u)du$ we easily find that (Cauchy-Scwarz)

$$\mathbb{E}_1 \leq \frac{N^2}{p^2T^2} \sum_{k,l} \mathbb{E}^{1/2}\left[\left(\int_{t_{l-1}}^{t_l} \mu(u)du\right)^4\right] \mathbb{E}^{1/2}\left[\left(\int_{t_{k-1}}^{t_k} \mu(u)du\right)^4\right] \leq \frac{MT^2}{N^2}.$$

Moreover, we have, by using twice Cauchy-Schwartz inequality

$$\mathbb{E}_2 \leq \frac{N^2}{pT^2} \sum_k \mathbb{E}^{1/2}\left[(R_0(t_k) - R_0(t_{k-1}))^4\right] \mathbb{E}^{1/2}\left[\left(\int_{t_{k-1}}^{t_k} \mu(u)du\right)^4\right] \leq \frac{T}{N}\sqrt{3M\mathbb{E}(\sigma^4)}$$

since $\mathbb{E}\left[\left(\int_{t_{k-1}}^{t_k} \mu(u)du\right)^4\right] \leq M(t_k - t_{k-1})^4 = MT^4/N^4$ and that

$$\mathbb{E}\left[(R_0(t_k) - R_0(t_{k-1}))^4\right] = 3\int\int_{[t_{k-1},t_k]^2} \mathbb{E}(\sigma^2(u)\sigma^2(v))dudv \leq 3\frac{T^2}{N^2}\mathbb{E}(\sigma^4).$$

Lastly, we mention that this result is robust to some leverage effect i.e. it still holds even if W^S and $W^{\tilde{\sigma}}$ are not independent, which implies only a slight increase in the constant C". All these results are straightforward consequences of Proposition 2.4. This ends the proof of Proposition 2.6. □

Proof of Proposition 3.2: Let $U_t(h)$ be defined by (3.1). From Proposition 3.1, it follows that:

$$\Gamma(\alpha+1)U_t(h) = \int_0^h (h-s)^\alpha(\tilde{\sigma}_c^2(t+s) - e^{-ks}\tilde{\sigma}_c^2(t))ds.$$

The conditional expectation can be written

$$\mathbb{E}_t(U_t^k(h)) = \frac{1}{\Gamma(1+\alpha)^k}\int_0^h \dots \int_0^h \prod_{i=1}^k (h-x_i)^\alpha \mathbb{E}_t\left(\prod_{i=1}^k \left[\tilde{\sigma}_c^2(t+x_i) - e^{-kx_i}\tilde{\sigma}_c^2(t)\right]\right) dx_1 \dots dx_k.$$

Then the result holds if we prove that

$$\mathbb{E}_t\left(\prod_{i=1}^k \left[\tilde{\sigma}_c^2(t+x_i) - e^{-kx_i}\tilde{\sigma}_c^2(t)\right]\right) = A_k(x_1,\dots,x_k)X(t) + B_k(x_1,\dots,x_k). \tag{5.6}$$

This can be checked by induction. Indeed (5.6) is true for $k=1$ with null coefficients and for $k=2$ since if for instance $x_1 \leq x_2$, by inserting $\mathbb{E}_{t+x_1}$, the conditional expectation is equal to $e^{-k(x_2-x_1)}\mathbb{E}_t\left[\left(\tilde{\sigma}_c^2(t+x_1) - e^{-kx_1}\tilde{\sigma}_c^2(t)\right)^2\right]$. This term is equal to $e^{-k(x_2-x_1)}\mathbb{E}_t\left[\tilde{\sigma}_c^4(t+x_1) - e^{-2kx_1}\tilde{\sigma}_c^4(t)\right]$. Then using the standard following formula

$$\mathbb{E}_t\left[\tilde{\sigma}_c^4(t+v)\right] = e^{-2kv}\tilde{\sigma}_c^4(t) + \gamma^2\frac{e^{-kv} - e^{-2kv}}{k}\tilde{\sigma}_c^2(t) + \gamma^2\tilde{\theta}\frac{1-e^{-2kv}}{2k} \tag{5.7}$$

gives $e^{-k(x_2-x_1)}\left[(\gamma^2/k)(e^{-kx_1} - e^{-2kx_1})\tilde{\sigma}_c^2(t) + (\gamma^2\tilde{\theta}/2k)(1-e^{-2kx_1})\right]$. This gives the result (5.6) for $k=2$, with A_2 and B_2 as defined by (5.9) and (5.10) applied with $p=2$. Next, assume that it holds for any product $k-1$ such terms and assume for simplicity that $x_1 \leq \dots \leq x_k$. Then insert

$\mathbb{E}_{t+x_1}$ inside the conditional expectation. The term $\left(\tilde{\sigma}_c^2(t+x_1) - e^{-kx_1}\tilde{\sigma}_c^2(t)\right)$ gets out $\mathbb{E}_{t+x_1}$, and there remains $k-1$ terms inside, to which we apply the induction property. We find

$$\begin{aligned}
&\mathbb{E}_t\left(\prod_{i=1}^{k}\left[\tilde{\sigma}_c^2(t+x_i) - e^{-kx_i}\tilde{\sigma}_c^2(t)\right]\right) \\
= \; & \mathbb{E}_t\left[\left(\tilde{\sigma}_c^2(t+x_1) - e^{-kx_1}\tilde{\sigma}_c^2(t)\right)\left(A_{k-1}(x_2-x_1,\dots,x_k-x_1)X(t+x_1) + B_{k-1}(x_2-x_1,\dots,x_k-x_1)\right)\right] \\
= \; & A_{k-1}(x_2-x_1,\dots,x_k-x_1)\mathbb{E}_t\left[\tilde{\sigma}_c^4(t+x_1)) - e^{-2kx_1}\tilde{\sigma}_c^4(t)\right] \\
& = A_{k-1}(x_2-x_1,\dots,x_k-x_1)\left(\gamma^2\frac{e^{-kx_1}-e^{-2kx_1}}{k}\tilde{\sigma}_c^2(t) + \gamma^2\tilde{\theta}\frac{1-e^{-2kx_1}}{2k}\right) \qquad (5.8)
\end{aligned}$$

using formula (5.7) again. Therefore (5.6) is proved by induction and the result of Proposition 3.2 follows.□

Note that (5.8) implies that

$$A_p(\alpha,h) = \int\dots\int_{0\le x_1\le\dots\le x_p\le h}\frac{p!}{\Gamma(\alpha+1)^p}\prod_{i=1}^{p}(h-x_i)^{\alpha}A_p(x_1,\dots,x_p)dx_1\dots dx_p$$

$$B_p(\alpha,h) = \int\dots\int_{0\le x_1\le\dots\le x_p\le h}\frac{p!}{\Gamma(\alpha+1)^p}\prod_{i=1}^{p}(h-x_i)^{\alpha}B_p(x_1,\dots,x_p)dx_1\dots dx_p$$

$$\text{with } A_p(x_1,\dots,x_p) = A_{p-1}(x_2-x_1,\dots,x_p-x_1)\frac{\gamma^2}{k}(e^{-kx_1}-e^{-2kx_1})$$

$$\text{and } B_p(x_1,\dots,x_p) = A_{p-1}(x_2-x_1,\dots,x_p-x_1)\frac{\gamma^2\tilde{\theta}}{2k}(1-e^{-2kx_1}).$$

Initial conditions are given by $A_2(x_1,x_2) = (\gamma^2/k)e^{-k(x_2-x_1)}(e^{-kx_1}-e^{-2kx_1})$ and $B_2(x_1,x_2) = (\gamma^2\tilde{\theta}/2k)e^{-k(x_2-x_1)}(1-e^{-2kx_1})$. Consequently, we can derive the following formulae:

$$A_p(x_1,\dots,x_p) = \left(\frac{\gamma^2}{k}\right)^{p-1}e^{-k(x_p-x_{p-1})}\prod_{i=1}^{p-1}\left(e^{-k(x_i-x_{i-1})} - e^{-2k(x_i-x_{i-1})}\right),$$

$$B_p(x_1,\dots,x_p) = \frac{\tilde{\theta}}{2}\left(\frac{\gamma^2}{k}\right)^{p-1}e^{-k(x_p-x_{p-1})}\prod_{i=1}^{p-1}\left(1 - e^{-2k(x_i-x_{i-1})}\right),$$

where by convention, $x_0 = 0$. In other words,

$$A_p(\alpha,h) = \frac{p!}{\Gamma(\alpha+1)^p}\left(\frac{\gamma^2}{k}\right)^{p-1}\int\dots\int_{0\le x_1\dots\le x_p\le h}e^{-k(x_p-x_{p-1})}\prod_{i=1}^{p-1}\left(e^{-k(x_i-x_{i-1})} - e^{-2k(x_i-x_{i-1})}\right)\prod_{i=1}^{p}(h-x_i)^{\alpha}dx_i \qquad (5.9)$$

$$B_p(\alpha,h) = \frac{p!}{\Gamma(\alpha+1)^p}\frac{\tilde{\theta}}{2}\left(\frac{\gamma^2}{k}\right)^{p-1}\int\dots\int_{0\le x_1\dots\le x_p\le h}e^{-k(x_p-x_{p-1})}\prod_{i=1}^{p-1}\left(1 - e^{-2k(x_i-x_{i-1})}\right)\prod_{i=1}^{p}(h-x_i)^{\alpha}dx_i. \qquad (5.10)$$

By using Lebesgue dominated Theorem, when $h\to+\infty$, we find that

$$A_p(\alpha,h) \sim \frac{p!}{\Gamma(\alpha+1)^p}\left(\frac{\gamma^2}{k}\right)^{p-1}h^{p\alpha}I_p(\alpha) \text{ and } B_p(\alpha,h) \sim \frac{p!}{\Gamma(\alpha+1)^p}\left(\frac{\gamma^2\tilde{\theta}}{2k}\right)^{p-1}h^{p\alpha}J_p(\alpha,h)$$

$$\text{with } I_p(\alpha) = \int\dots\int_{0\le x_1\dots\le x_p\le+\infty}e^{-k(x_p-x_{p-1})}\prod_{i=1}^{p-1}\left(e^{-k(x_i-x_{i-1})} - e^{-2k(x_i-x_{i-1})}\right)dx_1\dots dx_p$$

and $J_p(\alpha,h)=\int\dots\int_{0\le x_1\cdots\le x_p\le h} e^{-k(x_p-x_{p-1})}\prod_{i=1}^{p-1}\left(1-e^{-2k(x_i-x_{i-1})}\right)dx_1\dots dx_p$.

The change of variables $u_1=x_1, u_2=x_2-x_1,\dots,u_p=x_p-x_{p-1}$ gives

$$I_p(\alpha)=\int_0^{+\infty}e^{-ku_p}du_p\prod_{i=1}^{p-1}\int_0^{+\infty}(e^{-ku_i}-e^{-2ku_i})du_i=\frac{1}{k}\left(\frac{1}{2k}\right)^{p-1},$$

$$J_p(\alpha,h)=\int_0^{h}e^{-ku_p}du_p\prod_{i=1}^{p-1}\int_0^{h}(1-e^{-2ku_i})du_i=\frac{1-e^{-kh}}{k}\left(h-\frac{1-e^{-2kh}}{2k}\right)^{p-1}.$$

Therefore $I_p(\alpha,h)\sim_{h\to+\infty}(1/k)\,(1/2k)^{p-1}$ and $J_p(\alpha,h)\sim_{h\to+\infty}h^{p-1}/k$. Gathering all terms implies (3.7). □

Proof of Proposition 3.3: It follows from (3.11) that $K(h)=3\left[1+(1/h^2\mathbb{E}^2(\breve\sigma^2))\int_0^h\int_0^h c_{\breve\sigma^2}(u-v)dudv\right]$. From Proposition 2.4, for h small, $K(h)=3[1+(1/h^2\mathbb{E}^2(\breve\sigma^2))\int_0^h\int_0^h c_{\breve\sigma^2}(0)(1-\kappa(u-v)^{2\alpha+1}+o((u-v)^2))dudv]$, where $\kappa=k^{2\alpha+1}/[(2\alpha+1)\Gamma(2\alpha+1)]$. This leads, for h small, to

$$K(h)=3\left[1+\frac{1}{\mathbb{E}^2(\breve\sigma^2)}c_{\breve\sigma^2}(0)[1-2\kappa h^{2\alpha+1}/[(2\alpha+2)(2\alpha+3)]+o(h^2)\right].$$

This implies (i).

For h tending to infinity, write $K(h)=3\left[1+(2/h^2\mathbb{E}^2(\breve\sigma^2))\int_0^h(h-x)c_{\breve\sigma^2}(x)dx\right]$. From the second part of Proposition, 2.4, it follows that $\int_0^h(h-x)c_{\breve\sigma^2}(x)dx=O(h^{2\alpha+1})$ and this implies (ii). □

Proof of Proposition 3.4: It follows from the definition of Z that $cov(Z(t),Z(t+h))=\int_0^1\int_h^{h+1}c_{\breve\sigma^2}(u-v)dudv$. The Fourier transform of the previous equation yields: $f_Z(\lambda)=\int_{\mathbf{R}}e^{i\lambda h}\int_0^1\int_h^{h+1}c_{\breve\sigma^2}(u-v)dudvdh$. Then we use Fubini's Theorem

$$f_Z(\lambda)\;=\;\int_0^1\int_{\mathbb{R}}c_{\breve\sigma^2}(u-v)dudv\int_{u-1}^{u}e^{i\lambda h}dh=\frac{1-e^{-i\lambda}}{\lambda}\int_0^1e^{i\lambda v}dvf_{\breve\sigma^2}(\lambda).$$

The result follows then from Proposition 2.5 which implies $f_{\breve\sigma^2}(\lambda)=f_{\tilde\sigma^2}(\lambda)\lambda^{-2\alpha}$. □

Proof of Proposition 3.5. • The first result is proved in Proposition 2.4.

• Let $u_t=\mathbb{E}_t\breve\sigma^2(t+1)$. Since we can invert the conditional expectation and the integral

$$u_t=\int_{-\infty}^{t+1}\frac{(t+1-s)^{\alpha-1}}{\Gamma(\alpha)}\mathbb{E}_t(\tilde\sigma_c^2(s))ds+\theta.$$

According to Lemmas 5.1 1. and 2., we have

$$\Gamma(\alpha)^2cov(u_t,u_{t+h})=\int_{-\infty}^{t+1}\int_{-\infty}^{t+1+h}(t+1-s)^{\alpha-1}(t+1+h-r)^{\alpha-1}\mathbb{E}[\mathbb{E}_t(\tilde\sigma_c^2(s))\mathbb{E}_{t+h}(\tilde\sigma_c^2(r))]dsdr. \tag{5.11}$$

Observe that, using formula (2.11), if $s\notin[t,t+1]$ or if $r\notin[t+h,t+1+h]$, then

$$\mathbb{E}[\mathbb{E}_t(\tilde\sigma_c^2(s))\mathbb{E}_{t+h}(\tilde\sigma_c^2(r))]=c_{\tilde\sigma^2}(s-r)=\frac{\tilde\theta\gamma^2}{2k}e^{-k|r-s|}, \tag{5.12}$$

whereas if $(s,r)\in[t,t+1]\times[t+h,t+1+h]$

$$\mathbb{E}[\mathbb{E}_t(\tilde\sigma_c^2(s))\mathbb{E}_{t+h}(\tilde\sigma_c^2(r))]=\frac{\tilde\theta\gamma^2}{2k}e^{-k(s+r-2t)}. \tag{5.13}$$

Plugging formula (5.12) and (5.13) into (5.11) yields

$$\begin{aligned} cov(u_t, u_{t+h}) &= c_{\breve{\sigma}^2}(h) + \frac{\tilde{\theta}\gamma^2}{2k\Gamma(\alpha)^2}\int_t^{t+1}\int_{t+h}^{t+1+h}(t+1-s)^{\alpha-1}(t+1+h-s)^{\alpha-1}e^{-k(r+s-2t)}drds \\ &\quad -\frac{\tilde{\theta}\gamma^2}{2k\Gamma(\alpha)^2}\int_t^{t+1}\int_{t+h}^{t+1+h}(t+1-s)^{\alpha-1}(t+1+h-s)^{\alpha-1}e^{-k|r-s|}drds. \end{aligned}$$

Then, since $\alpha < \frac{1}{2}$ it remains to prove that for h near ∞

$$(5.14) M(h) = \int_t^{t+1}\int_t^{t+1+h}(t+1-s)^{\alpha-1}(t+1+h-s)^{\alpha-1}e^{-k(r+s-2t)}drds = O(|h|^{\alpha-1}),$$

$$(5.15) \quad N(h) = \int_t^{t+1}\int_t^{t+1+h}(t+1-s)^{\alpha-1}(t+1+h-s)^{\alpha-1}e^{-k|r-s|}drds = O(|h|^{\alpha-1}).$$

The change of variables $x = s-t$, $y = r-t-h$ leads to $M(h) = e^{-kh}\int_0^1\int_0^1(1-x)^{\alpha-1}(1-y)^{\alpha-1}e^{-k(x+y)}dxdy$, $N(h) = \int_0^1\int_0^1(1-x)^{\alpha-1}(1-y)^{\alpha-1}e^{-k|-x+y+h|)}dxdy$. Taking $h>1$ yields $N(h) = e^{-kh}\int_0^1\int_0^1(1-x)^{\alpha-1}(1-y)^{\alpha-1}e^{-k|-x+y|)}dxdy$. Therefore, $M(h) = O(e^{-kh})$ and $N(h) = O(e^{-kh})$ when h goes to infinity. The behavior of $\mathrm{cov}(u_t, u_{t+h})$ is given by the one of $c_{\breve{\sigma}^2}(h)$ which from Proposition 2.4 is given by $c_{\breve{\sigma}^2}(h) \sim c_{\breve{\sigma}^2}(0)(kh)^{2\alpha+1}/\Gamma(2\alpha)$. □

• The proof follows the same lines as the previous one. Since we can invert the conditional expectation and the integral, we can write $Y(t) = \int_0^1\int_{-\infty}^{t+u}(t+u-s)^{\alpha-1}\mathbb{E}_t(\tilde{\sigma}_c^2(s))dsdu + \theta$. According to Lemmas 5.1 1. and 2., we have

$$\begin{aligned} &\Gamma(\alpha)^2 cov(Y(t), Y(t+h)) \\ =& \int_0^1\int_0^1 c_{\breve{\sigma}^2}(h+v-u)dudv \\ &+\int_0^1\int_0^1\int_t^{t+u}\int_{t+h}^{t+v+h}(t+u-s)^{\alpha-1}(t+v+h-r)^{\alpha-1}[e^{-k(s-t+r-t)}-e^{-k|r-s|}]dsdrdudv \\ =& \int_0^1\int_0^1\left[c_{\breve{\sigma}^2}(h+v-u)+\int_0^u\int_0^v(u-x)^{\alpha-1}(v-x)^{\alpha-1}[e^{-k(x+y+h)}-e^{-k(|y+h-x|)}]dsdr\right]dudv. \end{aligned}$$

Clearly for great values of h, the first integral behaves as $c_{\breve{\sigma}^2}(h)$ and for $h>1$, using that $k|y+h-x| = k(y+h-x)$ and with the same tools as above, we find that the second integral is of order $O(e^{-kh})$. □

Proof of Proposition 3.6. Since $Z(s+1)-\theta = F(s+1)-F(s)$ with $F'(s) = f(s) = I^\alpha(\tilde{\sigma}_c^2)(s)$, and since $I^{-\alpha}$ commutes with $f \mapsto \int_\cdot^{\cdot+1}(f)$ we have

$$I^{-\alpha}(Z-\theta)(t+1) = \int_t^{t+1}I^{-\alpha}(I^\alpha(\tilde{\sigma}_c^2))(s)ds = \int_t^{t+1}\tilde{\sigma}_c^2(s)ds = \int_t^{t+1}(\tilde{\sigma}^2(s)-\tilde{\theta})ds. \ \square$$

Proof of Proposition 4.1. First, we observe that $(\int_{-\infty}^0(t-s)^{\alpha-1}[\tilde{\sigma}_c^2(s)]ds,\ t\geq 0)$ goes to 0 when t goes to ∞. Indeed, according to the expression of $c_{\tilde{\sigma}^2}$ given in (2.11)

$$\mathbb{E}\left\{\left[\int_{-\infty}^0(t-s)^{\alpha-1}\tilde{\sigma}_c^2(s)ds\right]^2\right\} = \int_{-\infty}^0\int_{-\infty}^0(t-s)^{\alpha-1}(t-u)^{\alpha-1}e^{-k|u-s|}duds$$

is dominated by $c_{\breve{\sigma}^2}(t)$. Then according to Proposition 2.4, we have $\mathbf{E}\left\{\left[\int_{-\infty}^{0}(t-s)^{\alpha-1}\tilde{\sigma}_c^2(s)ds\right]^2\right\} = O(|t|^{2\alpha-1})$. As a consequence, the process $I^{(\alpha)}(\tilde{\sigma}_c^2)$ to be discretized is approximated by $I_0^{(\alpha)}(\tilde{\sigma}_c^2)$ where $I_0^{(\alpha)}$ is given by

$$I_0^{(\alpha)}(X)(t) = \int_0^t \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}X(s)ds. \tag{5.16}$$

Second, Formula (4.1) simply follows from $(t-s)^{\alpha-1} = (1/\Gamma(1-\alpha))\int_0^\infty x^{-\alpha}e^{-x(t-s)}dx$, and from Fubini's theorem. We refer to Coutin and Pontier (2001) for further properties of $\Psi(x,t,f)$.
Therefore, the discretization is performed as follows, by using representation (4.1). For $r \in [1,2]$ and $n \in \mathbb{N}^*$, let $\pi = (x_i, i = -n, \dots, n)$ be a geometric subdivision with mesh r and size $2n+1$ given by $x^i = r^i$. To each $i = -n, \dots, n-1$ we associate c_i and η_i given by (4.3) and (4.4) and the sum

$$I_0^{(\alpha,r,n)}(f)(t) = \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\sum_{i=-n}^{n-1} c_i \Psi(\eta_i, t, f).$$

Then, according to Lemma 13 of Carmona et al. (2000) applied to $g = \Psi(.,.,f)$, $\mu(dx) = \min(1, x^{-1})$ and $x^{-\alpha}$ and according to the properties of Ψ as given in Theorem 4.10 of Coutin and Pontier (2001), there exists a constant C such that for any f continuous from $[0,T]$ into $\mathbb{R}$, $\sup_{t\in[0,T]}|I_0^{(\alpha,r,n)}(f)(t) - I_0^{(\alpha)}(f)(t)| \le C[(r-1)^2 + r^{-n\alpha}]$.
It remains to compute a $\Psi(\eta_i, ., f)$. For that purpose, let $(t_j = j\Delta,\ j = 0,...,N)$ for $\Delta = \frac{T}{N}$ be a standard subdivision of $[0,T]$. We find the approximation (4.5) and

$$I_0^{(\alpha,r,n,\Delta)}(f)(t_j) = \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\sum_{i=-n}^{n-1} c_i \Psi^\Delta(\eta_i, t_j, f).$$

Applying Corollary 4.2 and Proposition 4.21 of Coutin and Pontier (2001), we obtain the following result:

Proposition 5.1. *There exists a constant C such that for any f β-continuous on $[0,T,]$ then for any $r \in [1,2]$, $n \in \mathbb{N}^*$ and $\Delta \in]0,1]$, $\sup_{j=0,\dots,N}|\Psi^\Delta(\eta_i,t_j,f) - \Psi(\eta_i,t_j,f))| \le C\Delta^\beta$.*

Therefore, using 1. of Lemma 5.1, we approximate $I_0^{(\alpha)}(\tilde{\sigma}^2 - \tilde{\theta})$ by (4.6). The result of Proposition 4.1 is then a consequence of Coutin and Pontier (2001).□
Proof of Formula (4.8). First observe that as previously

Lemma 5.2. *$I^{(-\alpha)}(\breve{\sigma}^2)(t) - I_0^{(-\alpha)}(\breve{\sigma}^2)(t)$ goes to 0 when t goes to ∞ in $\mathbb{L}^2(\Omega,\mathbb{P})$.*

Now, it remains to approximate $I_0^{(-\alpha)}(f)(t)$ for any f Hölder continuous of index $\beta > \alpha$. As in the previous section, we write $(t-s)^{-(\alpha+1)} = \int_0^{+\infty} x^\alpha e^{-x(t-s)}dx/\Gamma(1+\alpha)$, and we find, using Fubini's Theorem, that for $\alpha \in (0,1)$,

$$I_0^{(-\alpha)}(f)(t) = \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\int_0^{+\infty} x^{\alpha-1}\left(f(t)e^{-xt} + x\int_0^t e^{-x(t-s)}(f(t)-f(s))ds\right)dx.$$

Then we obtain the discretization $I^{(-\alpha),r,n}(f)(t) = \sum_{j=-n+1}^{n+1} c'_j \Xi(\eta'_j, t, f)$ where

$$\Xi(x,t,f) = e^{-xt} f(t) + \int_0^t x e^{-x(t-s)} (f(t) - f(s)) ds,$$

and c'_j, η'_j are given by (4.7). Then from Proposition 4.1 and Theorem 4.9 of Coutin and Pontier (2001), there exists a constant C such that for any f Hölder continuous of index $\beta > \alpha$,

$$\sup_{t\in[0,T]} |I_0^{(-\alpha)}(f)(t) - I^{(-\alpha),r,n}(f)(t)| \leq C[(r-1)^2 + r^{-n\min(\alpha,\beta-\alpha/2)}].$$

It remains to compute $\Xi(\eta_i, t, f)$. Note that $\Xi(x,t,f) = f(t) - \int_0^t x e^{-x(t-s)} f(s) ds$. For $t_j = j\Delta$, $\Delta = \frac{T}{N}$, we have (4.7). From Coutin and Pontier (2001), we obtain the estimation given by (4.8). □

Proof of Lemma 5.2. We compute

$$R_t = \mathbb{E}\left[\left(I^{(-\alpha)}(\breve{\sigma}^2)(t) - I_0^{(-\alpha)}(\breve{\sigma}^2)(t)\right)^2\right] = \left(\frac{\alpha}{\Gamma(1-\alpha)}\right)^2 \mathbb{E}\left\{\left(\int_{-\infty}^0 \sigma^2(s)/(t-s)^{\alpha+1} ds\right)^2\right\}.$$

Using Fubini's theorem, we obtain

$$R_t = \frac{\alpha^2}{\Gamma(1-\alpha)^2} \int_{-\infty}^0 \int_{-\infty}^0 \frac{c_{\breve{\sigma}^2}(u-s)}{(t-s)^{\alpha+1}(t-u)^{\alpha+1}} duds.$$

Then is clear from formula (5.3) that $|c_{\sigma^2}(h)| \leq c_{\breve{\sigma}^2}(0)$ for all nonnegative h. Therefore, $R_t = O\left(1/t^{2\alpha}\right)$. It follows that R_t goes to 0 when t goes to ∞ which completes the proof. □

References

[1] ANDERSEN, T.G., T. BOLLERSLEV and F.X. DIEBOLD (2003): "Parametric and Nonparametric Volatility Measurement," in *Handbook of Financial Econometrics,* eds Y. Ait-Sahalia and L.P. Hansen. Amsterdam : North Holland, forthcoming.

[2] BARNDORFF-NIELSEN, O.E. and N. SHEPHARD, N. (2002): "Econometric analysis of realised volatility and its use in estimating stochastic models." *Journal of the Royal Statistical Society*, Ser. B, 64, 253-280.

[3] BLACK, F. and M. SCHOLES (1973): "The pricing of options and corporate liabilities." *Journal of Political Economy,* 3, 637-654.

[4] BOLLERSLEV, T. and H.O. MIKKELSEN (1999): "Long-term equity anticipation securities and stock market volatility dynamics." *Journal of Econometrics,* 92, 75-99.

[5] BOLLERSLEV, T. and H. ZHOU (2002): "Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility." *Journal of Econometrics,* 109, 33-65

[6] BYOUN, S., KWOK, C.C.Y and PARK, H.Y. (2003): "Expectations Hypothesis of the Term Structure of Implied Volatility: Evidence from Foreign Currency and Stock Index Options." *Journal of Financial Econometrics,* 1, 126-151

[7] CAMPA, J.M. and CHANG, P.H.K. (1995): "Testing the Expectations Hypothesis on the Term Structure of Volatilities in Foreign Exchange Options." *Journal of Finance,* 50, 529-547.

[8] CARMONA, P., L. COUTIN and G. MONTSENY (2000): "Approximation of some Gaussian processes." *Statistical Inference for Stochastic Processes,* 3, 161-171.

[9] CHERIDITO, P. and D. NUALART (2005): "Stochastic integral of divergence type with respect to fractional Brownian motion with Hurst parameter $H \in (0, \frac{1}{2})$." Ann. Inst. H. Poincaré Probab. Statist., 41, 1049–1081.

[10] CLARK, P.K. (1973): "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices." *Econometrica,* 41, 135-155.

[11] COMTE, F. and E. RENAULT (1998): "Long memory in continuous time stochastic volatility models." *Mathematical Finance,* 8, 291-323.

[12] COUTIN, L. and M. PONTIER (2001): "Approximation d'un drap brownien fractionnaire." Working paper, University of Toulouse 3-Paul Sabatier.

[13] COX, J., J. Jr INGERSOLL and S. ROSS (1985): "An intertemporal general equilibrium model of asset prices." *Econometrica,* 53, 363-384.

[14] DAHLHAUS, R. (1989): Efficient parameter estimation for self-similar processes. *Annals of Statistics* 17, 1749-1766.

[15] DIOP, A. (2002): "Schéma d'Euler symétrisé pour des processus de type Cox-Ingersoll-Ross, de Hull-White et des processus de Bessel." PhD Thesis, INRIA.

[16] DUFFIE, D., R. PAN and K. SINGLETON (2000): "Transform analysis and asset pricing for affine jump-diffusion." *Econometrica,* 68, 1343-1376.

[17] FELLER, W. (1951): "Two singular diffusion problems." *Ann. of Math* (2) 54, 173-182.

[18] GARCIA, R., R. LUGER and E. RENAULT, E. (2001): "Asymmetric smiles, leverage effects and structural parameters." Working Paper CIRANO 2001s-01.

[19] GEMAN, H. and M. YOR (1993): "Bessel Processes, Asian options and perpetuity." *Mathematical Finance,* 3-4, 349-375.

[20] GEWEKE, J. and S. PORTER-HUDAK (1983): "The estimation and application of long memory time series models." *Journal of Time Series Analysis,* 4, 221-238.

[21] HESTON, S.L. (1993): "A closed-form solution for options with stochastic volatility with applications to bond and currency options." *The Review of Financial Studies,* 6, 327-343.

[22] HU, Y. and D. NUALART (2005): "Some processes associated with fractional Bessel processes." J. Theoret. Probab., 18 377-397.

[23] HULL, J. and A. WHITE (1987): "The pricing of options on assets with stochastic volatilities." *The Journal of Finance,* 3, 281-300.

[24] JACQUIER, E., N.G. POLSON and P.E. ROSSI (1994): "Bayesian analysis of stochastic volatility models" (with discussion). *J. Bus. & Econ. Statistics* 12, 371-417.

[25] LAMBERTON, D. and B. LAPEYRE (1996): *Introduction to Stochastic Calculus Applied to Finance*. London, Chapman&Hall.

[26] MOREAUX, F., P. NAVATTE and C. VILLA (1998): "Market volatility index and implicit maximum likelihood estimation of stochastic volatility models." Preprint of the CREREQ, University of Rennes, France.

[27] NELSON, D.B. (1991): "Conditional heteroskedasticity in asset returns: a new approach." *Econometrica,* 59, 347-370.

[28] PASTORELLO, S., E. RENAULT and N. TOUZI (2000): "Statistical Inference for Random Variance Option Pricing." *Journal of Business and Economic Statistics,* 18, 358-367.

[29] RENAULT, E. (1997): "Econometric models of option pricing errors in advances econometrics," Seventh World Congress, Vol.3. Eds. D.M. Kreps and K.F. Wallis), Cambridge University Press.

[30] RENAULT, E. and N. TOUZI (1996): "Option hedging and implicit volatilities in a stochastic volatility model." *Mathematical Finance,* 6, 279-302.

[31] ROBINSON, P.M. (1996): "Semiparametric analysis of long-memory time series." *Annals of Statistics,* 22, 515-539.

[32] ROGERS, L. C. G. (1995): "Which model for term structure of interest rates should one use?" in *IMA, Vol. 65: Mathematical Finance,* eds. MHA Davis et al. Springer, 96-116.

[33] ROMANO, M. and N. TOUZI (1997): "Contingent claims and market completeness in a stochastic volatilty model." *Mathematical Finance,* 7, 399-412.

[34] SAMKO, S.G., A.A. KILBAS and O.I. MARICHEV (1993): *Fractional Integrals and Derivatives, Theory and Applications.* Gordon and Breach Science Publishers, Amsterdam.

[35] SUNDARESAN, S. (2000): "Continuous-time methods in Finance: a review and an assessment." *Journal of Finance,* 4, 1569-1622.

[36] SUTTON, J. (2000): *Marshall tendencies.* Cambridge University Press.

[37] VELASCO, C. (2000): "Non-Gaussian log-periodogram regression." *Econometric Theory,* 16, 44-79.

E-mail address: `laure.coutin@math-info.univ-paris.fr`

E-mail address: `fabienne.comte@univ-paris5.fr`

E-mail address: `renault@email.unc.edu`