Mathematical Finance, Vol. 8, No. 4 (October 1998), 291–323

LONG MEMORY IN CONTINUOUS-TIME STOCHASTIC VOLATILITY MODELS

FABIENNE COMTE

URA 1321 and LSTA, University Paris 6 and CREST-ENSAE

ERIC RENAULT

GREMAQ-IDEI, University of Toulouse and Institut Universitaire de France

This paper studies a classical extension of the Black and Scholes model for option pricing, often known as the Hull and White model. Our specification is that the volatility process is assumed not only to be stochastic, but also to have long-memory features and properties. We study here the implications of this continuous-time long-memory model, both for the volatility process itself as well as for the global asset price process. We also compare our model with some discrete time approximations. Then the issue of option pricing is addressed by looking at theoretical formulas and properties of the implicit volatilities as well as statistical inference tractability. Lastly, we provide a few simulation experiments to illustrate our results.

KEY WORDS: continuous-time option pricing model, stochastic volatility, volatility smile, volatility persistence, long memory

1. INTRODUCTION

If option prices in the market were conformable with the Black–Scholes (1973) formula, all the Black–Scholes implied volatilities corresponding to various options written on the same asset would coincide with the volatility parameter σ of the underlying asset. In reality this is not the case, and the Black–Scholes (BS) implied volatility $\sigma_{t,T}^{imp}$ heavily depends on the calendar time t, the time to maturity $T - t$, and the moneyness of the option. This may produce various biases in option pricing or hedging when BS implied volatilities are used to evaluate new options or hedging ratios. These price distortions, well-known to practitioners, are usually documented in the empirical literature under the terminology of the smile effect, where the so-called “smile” refers to the U-shaped pattern of implied volatilities across different strike prices.

It is widely believed that volatility smiles can be explained to a great extent by a modeling of stochastic volatility, which could take into account not only the so-called volatility clustering (i.e., bunching of high and low volatility episodes) but also the volatility effects of exogenous arrivals of information. This is why Hull and White (1987), Scott (1987), and Melino and Turnbull (1990) have proposed an option pricing model in which the volatility of the underlying asset appears not only time-varying but also associated with a specific

A previous version of this paper has benefitted from helpful comments from S. Pliska, L. C. G. Rogers, M. Taqqu, and two anonymous referees. All remaining errors are ours.

Initial manuscript received December 1995; final revision received September 1997.

Address correspondence to F. Comte, ISUP Boite 157, Université Paris 6, 4 Place Jussieu, 75 252 Paris cedex 05 France; e-mail: comte@ccr.jussieu.fr.

risk according to the "stochastic volatility" (SV) paradigm

$$\text{(1.1)} \qquad \begin{cases} \frac{dS(t)}{S(t)} = \mu(t, S(t))dt + \sigma(t)dw^1(t) \\ d(\ln \sigma(t)) = k(\theta - \ln \sigma(t))dt + \gamma dw^2(t), \end{cases}$$

where $S(t)$ denotes the price of the underlying asset, $\sigma(t)$ is its instantaneous volatility, and $(w^1(t), w^2(t))$ is a nondegenerate bivariate Brownian process. The nondegenerate feature of (w^1, w^2) is characteristic of the SV paradigm, in contrast to continuous-time ARCH-type models where the volatility process is a deterministic function of past values of the underlying asset price.

The logarithm of the volatility is assumed to follow an Ornstein–Uhlenbeck process, which ensures that the instantaneous volatility process is stationary, a natural way to generalize the constant-volatility Black and Scholes model. Indeed, any positive-valued stationary process could be used as a model of the stochastic instantaneous volatility (see Ghysels, Harvey and Renault (1996) for a review). Of course, the choice of a given statistical model for the volatility process heavily influences the deduced option pricing formula. More precisely, Hull and White (1987) show that, under specific assumptions, the price at time t of a European option of exercise date T is the expectation of the Black and Scholes option pricing formula where the constant volatility σ is replaced by its quadratic average over the period:

$$\text{(1.2)} \qquad \sigma^2_{t,T} = \frac{1}{T-t}\int_t^T \sigma^2(u)\, du,$$

and where the expectation is computed with respect to the conditional probability distribution of $\sigma^2_{t,T}$ given $\sigma(t)$. In other words, the square of implied Black–Scholes volatility $\sigma^{imp}_{t,T}$ appears to be a forecast of the temporal aggregation $\sigma^2_{t,T}$ of the instantaneous volatility viewed as a flow variable.

It is now well known that such a model is able to reproduce some empirical stylized facts regarding derivative securities and implied volatilities. A symmetric smile is well explained by this option pricing model with the additional assumption of independence between w^1 and w^2 (see Renault and Touzi (1996)). Skewness may explain the correlation of the volatility process with the price process innovations, the so-called leverage effect (see Hull and White 1987). Moreover, a striking empirical regularity that emerges from numerous studies is the decreasing amplitude of the smile being a function of time to maturity; for short maturities the smile effect is very pronounced (BS implied volatilities for synchronous option prices may vary between 15% and 25%), but it almost completely disappears for longer maturities. This is conformable to a formula like (1.2) because it shows that, when time to maturity is increased, temporal aggregation of volatilities erases conditional heteroskedasticity, which decreases the smile phenomenon.

The main goal of the present paper is to extend the SV option pricing model in order to capture well-documented evidence of **volatility persistence** and particularly occurrence of fairly pronounced smile effects even for rather long maturity options. In practice, the decrease of the smile amplitude when time to maturity increases turns out to be much slower than it goes according to the standard SV option pricing model in the setting (1.1). This evidence is clearly related to the so-called volatility persistence, which implies that temporal aggregation (1.2) is not able to fully erase conditional heteroskedasticity.

Generally speaking, there is widespread evidence that volatility is highly persistent.

Particularly for high frequency data one finds evidence of near unit root behavior of the conditional variance process. In the ARCH literature, numerous estimates of GARCH models for stock market, commodities, foreign exchange, and other asset price series are consistent with an IGARCH specification. Likewise, estimation of stochastic volatility models show similar patterns of persistence (see, e.g., Jacquier, Polson and Rossi 1994). These findings have led to a debate regarding modeling persistence in the conditional variance process either via a unit root or a long memory-process. The latter approach has been suggested both for ARCH and SV models; see Baillie, Bollerslev, and Mikkelsen (1996), Breidt, Crato, and De Lima (1993), and Harvey (1993). This allows one to consider **mean-reverting** processes of stochastic volatility rather than the extreme behavior of the IGARCH process which, as noticed by Baillie et al. (1996), has low attractiveness for asset pricing since “the occurence of a shock to the IGARCH volatility process will persist for an infinite prediction horizon.”

The main contribution of the present paper is to introduce long-memory mean reverting volatility processes in the continuous time Hull and White setting. This is particularly attractive for option pricing and hedging through the so-called **term structure of BS implied volatilities** (see Heynen, Kemna, and Vorst 1994). More precisely, the long-memory feature allows one to capture the well-documented evidence of persistence of the stochastic feature of BS implied volatilities, when time to maturity increases. Since, according to (1.2), BS implied volatilities are seen as an average of expected instantaneous volatilities in the same way that long-term interest rates are seen as average of expected short rates, the type of phenomenon we study here is analogous to the studies by Backus and Zin (1993) and Comte and Renault (1996) who capture persistence of the stochastic feature of long-term interest rates by using long-memory models of short-term interest rates.

Indeed, we are able to extend Hull and White option pricing to a continuous-time long-memory model of stochastic volatility by replacing the Wiener process w^2 in (1.1) by a fractional Brownian motion w^2_α, with α restricted to $0 \leq \alpha < \frac{1}{2}$ (instead of $|\alpha| < \frac{1}{2}$ allowed by the general definition because long memory occurs on that range only). Note that the Wiener case corresponds to $\alpha = 0$. Of course, for nonzero α, w^2_α is no longer a semimartingale (see Rogers 1995), and thus usual stochastic integration theory is not available. But, following Comte and Renault (1996), we only need L^2 theory of integration for Gaussian processes and we obtain option prices that, although they are functions of the underlying volatility processes, do ensure the semimartingale property as a maintained hypothesis for asset price processes (including options).[1] This semimartingale property is all the more important for asset prices processes because stochastic processes that are not semimartingales do not admit equivalent martingale measures. Indeed we know from Delbaen and Schachermayer (1994) that an asset price process admits an equivalent martingale measure if and only if the NFLVR (no free lunch with vanishing risk) condition holds. As stressed by Rogers (1995), when this condition fails, “this does not of itself imply the existence of arbitrage, though in any meaningful economic sense it is just as bad as that.” In that event, Rogers (1995) provides a direct construction of arbitrage with fractional Brownian motion. As long as the volatility itself is not a traded asset, all asset price processes that we consider here (underlying asset and options written on it) are conformable

[1] We are very grateful to L. C. G. Rogers to have helped us, in a private communication, to check this point. The semimartingale property of an option price C_t comes from the fact that it is computed as a conditional expectation of a (nonlinear) function of $\int_t^T \sigma^2(u)\, du$. This integration reestablishes the semimartingale property that was lost by the volatility process itself.

to the NFLVR. Note that we have nevertheless the same usual problem as in all models of that kind: the nonuniqueness of the neutral-risk equivalent measure.

The paper is organized as follows. We study in Section 2 the probabilistic properties of our Fractional Stochastic Volatility (FSV) model in continuous time, obtained by replacing the Wiener process w^2 in (1.1) by the following process that may be seen as a truncated version of the general fractional Brownian motion[2]:

$$w_\alpha^2(t) = \int_0^t \frac{(t-s)^\alpha}{\Gamma(1+\alpha)}\, dw^2(s), \qquad 0 < \alpha < \frac{1}{2}.$$

We explain why a high degree of fractional differencing α allows one to take into account the apparent widespread finding of integrated volatility for high frequency data. Section 3 gives the basis for more empirical studies of our FSV model through discrete time approximations. We stress the advantages of continuous-time modeling of long memory with respect to the usual ARFIMA à la Geweke and Porter-Hudak (1983) or their FIGARCH analogue in the ARCH literature. The main point is that only a continuous-time definition of the parameters of interest allows one to clearly disentangle long-memory parameters from short-memory ones.

Section 4 is devoted to the issue of option pricing and the study of the properties and features of implied volatilities. Since the first equation of (1.1) has remained invariant by our long-memory generalization of the Hull and White (1987) option pricing model, their argument can be extended in order to set an option pricing formula. The only change is the law of motion of the instantaneous volatility, whose long-memory feature modifies the orders of conditional heteroskedasticity (forecasted, temporally aggregated, ...) and of kurtosis coefficients with respect to time to maturity. We derive some formulas about these orders which extend those of Drost and Werker (1996) and thus "close the FIGARCH gap."

The statistical inference issue is addressed in Section 5. Of course, if the instantaneous volatility $\sigma(t)$ were observed, Comte and Renault's (1996) work about the estimation of continuous-time long-memory models could be used. But instantaneous volatilities are not directly observed and can only be filtered, either by an extension to FIGARCH models of Nelson and Foster's (1994) methodology or by using option prices as Pastorello, Renault, and Touzi (1993) do in the Hull and White context. Note that for $\alpha \neq 0$ the volatility process is no longer Markovian, so this may make awkward the practical use of the Hull and White option pricing formula. Nevertheless, it is shown how one could extend the Pastorello et al. (1993) methodology to the present framework. The alternative methodology we suggest in the present paper is to use approximate discretizations of the $S(t)$ stock price process in order to obtain some proxies of instantaneous volatilities and work with approximate likelihoods.

The discretizations found in Section 3 are used in Section 6 to perform some simulation experiments about continuous-time FSV models. A descriptive study of the resulting paths can then be obtained. The estimation procedures are compared through these Monte Carlo experiments. The misspecification bias introduced by a FIGARCH approximation of our continuous-time models is documented.

[2]This process is a tool for easy L^2 definitions of integrals w.r.t. the Fractional Brownian Motion (FBM), but can be replaced by the true FBM $\int_{-\infty}^0 ((t-s)^\alpha - (-s)^\alpha)dw^2(s) + w_\alpha^2(t)$.

2. THE FRACTIONAL STOCHASTIC VOLATILITY MODEL

2.1. A Simple Fractional Long-Memory Process

Comte and Renault (1996) used fractional processes to generalize the notion of Stochastic Differential Equation (SDE) of order p. We consider here only the first-order fractional SDE:

$$dx(t) = -kx(t)dt + \sigma dw_\alpha(t), \qquad x(0) = 0, \quad k > 0, \quad 0 < \alpha < \frac{1}{2}. \tag{2.1}$$

The solution can be written (see Comte and Renault 1996) as $x(t) = \int_0^t e^{-k(t-s)}\sigma\, dw_\alpha(s)$. Integration with respect to w_α is defined only in the Wiener L^2 sense and for the integration of deterministic functions only. We thus obtain families of Gaussian processes. The process $x(t)$ also can be written as $\int_0^t a(t-s)\, dw(s)$ with

$$\begin{aligned} a(x) &= \frac{\sigma}{\Gamma(1+\alpha)} \frac{d}{dx} \int_0^x e^{-ku}(x-u)^\alpha\, du \\ &= \frac{\sigma}{\Gamma(1+\alpha)} \left(x^\alpha - ke^{-kx} \int_0^x e^{ku} u^\alpha\, du \right). \end{aligned} \tag{2.2}$$

We denote by $y(t)$ the "stationary version" of $x(t)$, $y(t) = \int_{-\infty}^t a(t-s)\, dw(s)$. Therefore, the solution x of the fractional SDE is given by

$$x(t) = \int_0^t \frac{(t-s)^\alpha}{\Gamma(1+\alpha)}\, dx^{(\alpha)}(s), \tag{2.3}$$

where its derivative of order α is the solution

$$x^{(\alpha)}(t) = \frac{d}{dt} \int_0^t \frac{(t-s)^{-\alpha}}{\Gamma(1-\alpha)} = \int_0^t e^{-k(t-s)}\sigma\, dw(s) \tag{2.4}$$

of the associated standard SDE.

We can also give the general (continuous-time) spectral density of processes that are solutions of (2.1):

$$f^c(\lambda) = \frac{\sigma^2}{\Gamma(1+\alpha)^2 \lambda^{2\alpha}} \frac{1}{\lambda^2 + k^2}. \tag{2.5}$$

Lastly, it seems interesting to note that long-memory fractional processes as considered in Comte and Renault (1996) and solutions of (2.1) in particular have the following properties proved in Comte (1996):

i. The covariance function $\gamma = \gamma_y$ associated with x satisfies for $h \to 0$ and ψ constant:

$$\gamma(h) = \gamma(0) + \frac{1}{2}\psi.|h|^{2\alpha+1} + o(|h|^{2\alpha+1}). \tag{2.6}$$

ii. x is ergodic in the L^2 sense: $\frac{1}{T}\int_0^T x(s)\,ds \underset{T\to+\infty}{\overset{m.s.}{\longrightarrow}} 0$.

iii. There is a process $z(t)$ equivalent[3] to $x(t)$ and such that the sample function of z satisfies a Lipschitz condition of order β, $\forall\beta \in (0, \alpha + \frac{1}{2})$, a.s.

Thus the greater the value of α, the smoother the path of the process.

2.2. Properties of the Volatility in the FSV Model

The basic idea of our modeling strategy (see (1.1)) is to assume that the logarithm $x(t) = \ln \sigma(t)$ of the stochastic volatility is a solution of the first-order SDE (2.1). For the sake of simplicity, we assume $\theta = 0$ since it does not change the probabilistic properties of the process. Thus the volatility process $\sigma(t)$ is asymptotically equivalent (in quadratic mean) to the stationary process:

$$\tilde{\sigma}(t) = \exp\left(\int_{-\infty}^{t} e^{-k(t-s)}\gamma\, dw_\alpha^2(s)\right), \qquad k > 0,\quad 0 < \alpha < \frac{1}{2}. \tag{2.7}$$

As in usual diffusion models of stochastic volatility, the volatility process is assumed to be asymptotically stationary and nowhere differentiable. This is the reason we do not use an SDE (even fractional) of higher order. Nevertheless, the fractional exponent α provides some degree of freedom in the order of regularity. Indeed, it is possible to show for $\sigma(t)$ the same type of regularity property as for the fractional process $x(t) = \ln \sigma(t)$.

PROPOSITION 2.1. *Let $r_\sigma(h) = cov\ (\tilde{\sigma}(t+h), \tilde{\sigma}(t))$, where $\tilde{\sigma}$ is given by (2.7). Then, for $h \to 0$, $r_\sigma(h) = r_\sigma(0) + \eta.|h|^{2\alpha+1} + o(|h|^{2\alpha+1})$, where η is a given constant.*

(See Appendix A for all proofs.)

Roughly speaking, the autocorrelation function of the stationary process σ fulfills the regularity condition that ensures the Lipschitz feature of the sample paths. The greater α is, the smoother the path of the volatility process is. Therefore, a high degree of fractional differencing α allows one to take into account the apparent widespread finding of integrated volatility for high frequency data (see the simulation in Section 6.2). As a matter of fact, we can see that

$$\alpha > 0 \Rightarrow \frac{r_\sigma(h) - r_\sigma(0)}{h} \underset{h\to 0}{\longrightarrow} 0,$$

which could be interpreted as a near-integrated behavior

$$\frac{r_\sigma(h) - r_\sigma(0)}{h} = \frac{\rho^h - 1}{h} \underset{h\to 0}{\longrightarrow} \ln \rho \underset{\rho\to 1}{\longrightarrow} 0$$

if $\sigma(t)$ is considered as a continuous-time AR(1) process with a correlation coefficient ρ near 1.

[3] Two processes are called equivalent if they coincide almost surely.

This analogy between a unit root hypothesis and its fractional alternatives has already been used for unit root tests by Robinson (1993). Robinson's methodology could be a useful tool for testing integrated volatility against long memory in stochastic volatility behavior.

The concept of persistence that we advance thanks to the fractional framework is that of long memory instead of indefinite persistence of shocks as in the IGARCH framework. Indeed, we can prove the following result:

PROPOSITION 2.2. *In the context of Proposition 2.1, we have*

(i) $r_\sigma(h)$ *is of order* $O(|h|^{2\alpha-1})$ *for* $h \to +\infty$.
(ii) $\lim_{\lambda\to 0} \lambda^{2\alpha} f_\sigma(\lambda) = c \in \mathbb{R}^+$, *where* $f_\sigma(\lambda) = \int_{\mathbb{R}} r_\sigma(h) e^{i\lambda h}\, dh$ *is the spectral density of* $\tilde{\sigma}$.

Proposition 2.2 illustrates that the volatility process itself (and not only its logarithm) does entail the long-memory properties (generally summarized as in (i) and (ii) by the behavior of the covariance function near infinity and of the spectral density near zero) we could expect in the FSV model.

3. DISCRETE APPROXIMATIONS OF THE FSV MODEL

3.1. The Volatility Process

The volatility process dynamics are characterized by the fact that $x(t) = \ln \sigma(t)$ is a solution of the fractional SDE (2.1). So we know two integral expressions for $x(t)$ (with the notations of Section 2.1):

$$x(t) = \int_0^t \frac{(t-s)^\alpha}{\Gamma(1+\alpha)}\, dx^{(\alpha)}(s) = \int_0^t a(t-s)\, dw^2(s),$$

where $a(t-s)$ is given by (2.2).

A discrete time approximation of the volatility process is a formula to numerically evaluate these integrals using only the values of the involved processes $x^{(\alpha)}(s)$ and $w^2(s)$ on a discrete partition of $[0,t]$: j/n, $j = 0, 1, \ldots, [nt]$.[4] A natural way to obtain such approximations (see Comte 1996) is to approximate the integrands by step functions:

$$x_{n,1}(t) = \int_0^t \frac{\left(t - \frac{[ns]}{n}\right)^\alpha}{\Gamma(1+\alpha)}\, dx^{(\alpha)}(s) \quad \text{and} \quad x_{n,2}(t) = \int_0^t a\left(t - \frac{[ns]}{n}\right) dw^2(s), \tag{3.1}$$

which gives, neglecting the last terms for large values of n,

$$\hat{x}_n(t) = \sum_{j=1}^{[nt]} \frac{\left(t - \frac{j-1}{n}\right)^\alpha}{\Gamma(1+\alpha)} \Delta x^{(\alpha)}\left(\frac{j}{n}\right) \quad \text{and} \tag{3.2}$$
$$\tilde{x}_n(t) = \sum_{j=1}^{[nt]} a\left(t - \frac{j-1}{n}\right) \Delta w^2\left(\frac{j}{n}\right),$$

[4] $[z]$ is the integer k such that $k \le z < k+1$.

where we use the following notations: $\Delta x^{(\alpha)}(\frac{j}{n}) = x^{(\alpha)}(\frac{j}{n}) - x^{(\alpha)}(\frac{j-1}{n})$ and $\Delta w^2(\frac{j}{n}) = w^2(\frac{j}{n}) - w^2(\frac{j-1}{n})$.

Indeed, all these approximations converge toward the x process in distribution in the sense of convergence in distribution for stochastic processes as defined in Billingsley (1968); this convergence is denoted by $\overset{\mathcal{D}}{\Rightarrow}$. This result is proved in Comte (1996).

PROPOSITION 3.1. *$x_{n,1} \overset{\mathcal{D}}{\Rightarrow} x$, $x_{n,2} \overset{\mathcal{D}}{\Rightarrow} x$, $\hat{x}_n \overset{\mathcal{D}}{\Rightarrow} x$, and $\tilde{x}_n \overset{\mathcal{D}}{\Rightarrow} x$ when n goes to infinity.*

The proxy $\hat{x}_n$ is the most useful for comparing our FSV model with the standard discrete time models of conditional heteroskedasticity, whereas the most tractable for mathematical work is $\tilde{x}_n$.

3.2. FSV versus FIGARCH

Expression (3.2) provides a proxy $\hat{x}_n$ of x in function of the process $x^{(\alpha)}(\frac{j}{n})$, $j = 0, 1, \ldots, [nt]$, which is an AR(1) process associated with an innovation process $u(\frac{j}{n})$, $j = 0, 1, \ldots, [nt]$. Let us denote by

$$(1 - \rho_n L_n) x^{(\alpha)}\left(\frac{j}{n}\right) = u\left(\frac{j}{n}\right) \tag{3.3}$$

the representation of this process, where L_n is the lag operator corresponding to the sampling scheme $\frac{j}{n}$, $j = 0, 1, \ldots$, $L_n Y(\frac{j}{n}) = Y(\frac{j-1}{n})$, and $\rho_n = e^{-k/n}$ is the correlation coefficient for the time interval $\frac{1}{n}$.

Since the process $x^{(\alpha)}$ is asymptotically stationary, we can assume without loss of generality that its initial value is zero, $x^{(\alpha)}(\frac{j}{n}) = 0$ for $j \le 0$, which of course implies $u(\frac{j}{n}) = 0$ for $j \le 0$. Then we can write

$$\begin{aligned}\hat{x}_n\left(\frac{j}{n}\right) &= \sum_{i=1}^{j} \frac{(j-i+1)^\alpha}{n^\alpha \Gamma(1+\alpha)} \left[x^{(\alpha)}\left(\frac{i}{n}\right) - x^{(\alpha)}\left(\frac{i-1}{n}\right)\right] \\ &= \left[\sum_{i=0}^{j-1} \frac{(i+1)^\alpha - i^\alpha}{n^\alpha \Gamma(1+\alpha)} L_n^i\right] x^{(\alpha)}\left(\frac{j}{n}\right).\end{aligned}$$

Thus,

$$\hat{x}_n(\frac{j}{n}) = \left[\sum_{i=0}^{j-1} \frac{(i+1)^\alpha - i^\alpha}{n^\alpha \Gamma(1+\alpha)} L_n^i\right] (1 - \rho_n L_n)^{-1} u\left(\frac{j}{n}\right). \tag{3.4}$$

Expression (3.4) gives a parameterization of the volatility dynamics in two parts: a long-memory part that corresponds to the filter $\sum_{i=0}^{+\infty} a_i L_n^i / n^\alpha$ with $a_i = ((i+1)^\alpha - i^\alpha)/\Gamma(1+\alpha)$ and a short-memory part that is characterized by the AR(1) process: $(1 - \rho_n L_n)^{-1} u(\frac{j}{n})$.

We can show that the long-memory filter is "long-term equivalent" to the usual discrete time long-memory filter $(1-L)^{-\alpha} = \sum_{i=0}^{+\infty} b_i L^i$, where $b_i = \Gamma(i+\alpha)/(\Gamma(i+1)\Gamma(\alpha))$, in the sense that there is a long-term relationship (a cointegration relation) between the two types of processes. Indeed, we can show (see Comte 1996) that the two long-memory processes, $Y_t = \sum_{i=0}^{+\infty} a_i u_{t-i}$ and $Z_t = \sum_{i=0}^{+\infty} b_i u_{t-i}$, where a_i and b_i are defined previously and u_t is any short-term memory stationary process, are cointegrated: $Y_t - Z_t$ is short memory and $\sum_{i=0}^{+\infty} |a_i - b_i| < +\infty$, whereas $\sum_{i=0}^{+\infty} a_i = \sum_{i=0}^{+\infty} b_i = +\infty$.

But this long-term equivalence between our long-memory filter and the usual discrete time one $(1-L)^{-\alpha}$ does not imply that the standard parameterization ARFIMA$(1, \alpha, 0)$ is well-suited in our framework. Indeed, short-memory characteristics may be hidden by the short-term difference between the two filters. In other words, not only $(1 - \rho_n L_n)(n(1 - L_n))^\alpha \hat{x}_n(\frac{j}{n})$ is not in general a white noise,[5] but we are not even sure that $(n(1 - L_n))^\alpha \hat{x}_n\left(\frac{j}{n}\right)$ is an AR(1) process (even though we know that it is a short-memory stationary process). The usual discrete time filter $(1-L)^\alpha$ introduces some mixing between long- and short-term characteristics (see Comte 1996 and Section 6.3 for illustration).

This is the first reason why we believe that the FSV model is more relevant for high-frequency data than the FIGARCH model since the latter is based on an ARFIMA modeling of the squared innovations (see Baillie et al. 1996). The second reason is that the FSV model represents the log-volatility as an "AR(1) long-memory" process with a specific risk (in the particular case $\alpha = 0$, (3.4) corresponds to the stochastic variance model of Harvey, Ruiz, and Shephard 1994), but the GARCH type modeling does not introduce an exogenous risk of volatility and, by the way, does not explain why option markets are useful to hedge a specific risk.

3.3. The Global Filtering Model

In order to obtain a complete discrete time approximation of our FSV model, we have to discretize not only the volatility process, but also the associated asset price process $S(t)$ according to (1.1). Since it is not difficult to compute some discretizations of the trend part of an SDE, we can assume in this subsection, for the sake of notational simplicity, that $\ln S(t)$ is a martingale. Not only are we always able to perform a preliminary centering of the price process in order to be in this case, but also it is well known that the martingale hypothesis is often directly accepted, for exchange rates for example. So, with $Y(t) = \ln S(t)$ we are interested in the following dynamics:

$$\left\{ \begin{array}{l} dY(t) = \sigma(t) dw^1(t) \\ d(\ln \sigma(t)) = -k \ln \sigma(t) dt + \gamma dw_\alpha^2(t). \end{array} \right. \tag{3.5}$$

For a known process σ, a discretized approximation Y_n of the process Y can directly be obtained by a way similar to (3.1):

$$\begin{aligned} Y_n(t) &= \int_0^t \sigma\left(\frac{[ns]}{n}\right) dw^1(s) \\ &= \sum_{j=1}^{[nt]} \sigma\left(\frac{j-1}{n}\right) \Delta w^1\left(\frac{j}{n}\right) + \sigma\left(\frac{[nt]}{n}\right)\left(w^1(t) - w^1\left(\frac{[nt]}{n}\right)\right). \end{aligned}$$

[5] The fractional differencing operator $(1-L)^\alpha$ has to be modified into $(n(1-L_n))^\alpha$ in order to correctly normalize the unit root with respect to the unit period of time.

And by a remark of the same type as (3.2), we can also consider $\hat{Y}_n(t) = \sum_{j=1}^{[nt]} \sigma(\frac{j-1}{n}) \Delta w^1(\frac{j}{n})$. It can be proved that:

LEMMA 3.1. *$Y_n \overset{\mathcal{D}}{\Rightarrow} Y$ and $\hat{Y}_n \overset{\mathcal{D}}{\Rightarrow} Y$, when n grows to infinity.*

But from a practical viewpoint, the discretizations Y_n and $\hat{Y}_n$ are not very useful because they are based on the values of the process σ, which cannot be computed without some other errors of discretization. Thus we are more interested in the following joint discretization:

$$\tilde{\sigma}_n(t) = \exp\left[\sum_{j=1}^{[nt]} a\left(t - \frac{j-1}{n}\right) \Delta w^2\left(\frac{j}{n}\right)\right], \tag{3.6}$$

$$\tilde{Y}_n(t) = \sum_{j=1}^{[nt]} \tilde{\sigma}_n\left(\frac{j-1}{n}\right) \Delta w^1\left(\frac{j}{n}\right).$$

We can then prove the following proposition.

PROPOSITION 3.2.

$$\begin{pmatrix} \tilde{Y}_n \\ \tilde{\sigma}_n \end{pmatrix} \overset{\mathcal{D}}{\Rightarrow} \begin{pmatrix} Y \\ \sigma \end{pmatrix} \quad \textit{and thus} \quad \begin{pmatrix} \tilde{S}_n = \ln \tilde{Y}_n \\ \tilde{\sigma}_n \end{pmatrix} \overset{\mathcal{D}}{\Rightarrow} \begin{pmatrix} S \\ \sigma \end{pmatrix} \quad \textit{when} \quad n \to \infty.$$

Another parameterization can be obtained by using $\hat{\sigma}_n(t) = \exp(\hat{x}_n(t))$ rather than $\tilde{\sigma}_n(t) = \exp(\tilde{x}_n(t))$; the previous section has shown how this parameterization is given by α and ρ_n.

We have something like a discrete time stochastic variance model à la Harvey et al. (1994) which converges toward our FSV model when the sampling interval $\frac{1}{n}$ converges toward zero. The only difference is that, when $\alpha \neq 0$, $\ln \tilde{\sigma}_n(t)$ is not an AR(1) process but a long-memory stationary process. Such a generalization has in fact been considered in discrete time by Harvey (1993) in a recent working paper. He works with $y_t = \sigma_t \varepsilon_t$, $\varepsilon_t \sim IID(0, 1)$, $t = 1, \dots, T$, $\sigma_t^2 = \sigma^2 \exp(h_t)$, $(1 - L)^d h_t = \eta_t$, $\eta_t \sim IID(0, \sigma_\eta^2)$, $0 \leq d \leq 1$. The analogy with (3.6) is then obvious, with the remaining problem being the choice of the right approximation of the fractional derivation studied in the previous subsection. Moreover, our case is a little different from the one studied by Harvey in that we have in mind a volatility process of the type ARFIMA $(1, \alpha, 0)$ where he has an ARFIMA$(0, d, 0)$. But such discrete time models may be also useful for statistical inference.

4. OPTION PRICING AND IMPLIED VOLATILITIES

4.1. Option Pricing

The maintained assumption of our option pricing model is characterized by the price model (1.2), where $(w^1(t), w^2(t))$ is a standard Brownian motion. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the fundamental probability space. $(\mathcal{F}_t)_{t \in [0,T]}$ denotes the $\mathbb{P}$-augmentation of the filtration generated by $(w^1(\tau), w^2(\tau))$, $\tau \leq t$. It coincides with the filtration generated by $(S(\tau), \sigma(\tau))$, $\tau \leq t$ or $(S(\tau), x^{(\alpha)}(\tau))$, $\tau \leq t$, with $x(t) = \ln \sigma(t)$.

We look here for the call option premium C_t, which is the price at time $t \leq T$ of a European call option on the financial asset of price S_t at t, with strike K and maturing at time T. The asset is assumed not to pay dividends, and there are no transaction costs.

Let us assume that the instantaneous interest rate at time t, $r(t)$, is deterministic, so that the price at time t of a zero coupon bond of maturity T is $B(t,T) = \exp(-\int_t^T r(u)\,du)$.

We know from Harrison and Kreps (1981) that the no free lunch assumption is equivalent to the existence of a probability distribution $\mathbf{Q}$ on $(\Omega, \mathcal{F})$, equivalent to $\mathbb{P}$, under which the discounted *price processes* are martingales. We emphasize that no change of probability of the Girsanov type could have transformed the volatility process into a martingale, but there is no such problem for the price process $S(t)$. This stresses the interest of such models where the nonstandard fractional properties are set on $\sigma(t)$ and not directly on $S(t)$. This avoids any of the possible problems of stochastic integration with respect to a fractional process, which does not admit any standard decomposition. Indeed, the σ process appears only as a predictible and even L^2 continuous integrand.

Then we can use the standard arguments. An equivalent measure $\mathbf{Q}$ is characterized by a continuous version of the density process of $\mathbf{Q}$ with respect to $\mathbb{P}$ (see Karatzas and Shreve 1991, p. 184):

$$M(t) = \exp\left(-\int_0^t \lambda(u)' dW(u) - \frac{1}{2}\int_0^t \lambda(u)'\lambda(u)\,du\right),$$

where $W = (w^1, w^2)'$ and $\lambda = (\lambda_1, \lambda_2)'$ is adapted to $\{\mathcal{F}_t\}$ and satisfies the integrability condition $\int_0^T \lambda(u)'\lambda(u)\,du < \infty$ a.s. The processes λ_1 and λ_2 can be considered as risk premia relative to the two sources of risk w^1 and w^2. Moreover, the martingale property under $\mathbf{Q}$ of the discounted asset prices implies that: $\lambda_1(t)\sigma(t) = \mu(t, S(t)) - r(t)$.

As the market is incomplete, as is usual in such a context (two sources of risk and only one risky asset traded), there is no such relation fixing the volatility risk premium λ_2 and, indeed, the martingale probability $\mathbf{Q}$ is not unique.

We need to restrict the set of equivalent martingale probabilities by assuming that the process $\lambda_2(t)$ is a deterministic function $\bar{\lambda}_2$ of the two arguments t and $\sigma(t)$:

$$(\mathbf{A}) \quad \lambda_2(t) = \bar{\lambda}_2(t, \sigma(t)), \qquad \forall t \in [0, T],$$

which is a common assumption.

Girsanov's theorem leads to a characterization of the distribution under $\mathbf{Q}$ of the underlying asset. Let:

$$\tilde{w}^1(t) = w^1(t) + \int_0^t \lambda_1(u)\,du \quad \text{and} \quad \tilde{w}^2(t) = w^2(t) + \int_0^t \lambda_2(u)\,du.$$

Then $(\tilde{w}^1, \tilde{w}^2) = \tilde{w}'$ is a two-dimensional standard $\mathbf{Q}$-Wiener process adapted to $\{\mathcal{F}_t\}$. In particular, $\tilde{w}^1$ and $\tilde{w}_2$ are independent under $\mathbf{Q}$ by construction. Moreover σ is the solution to an equation depending only on w^2 that can be rewritten as a stochastic differential equation in $\tilde{w}^2$ (depending also on λ_2). Thus the processes $\tilde{w}^1$ and σ are still independent under $\mathbf{Q}$. With $\mathbf{Q}$ defined as previously, the call option price is given by

$$C_t = B(t,T)\mathbb{E}^{\mathbf{Q}}[\mathrm{Max}(0, S_T - K) \mid \mathcal{F}_t], \tag{4.1}$$

where $\mathbb{E}^{\mathbf{Q}}(.\mid\mathcal{F}_t)$ is the conditional expectation operator, given $\mathcal{F}_t$, when the price dynamics is governed by $\mathbf{Q}$. Since $\tilde{w}^1$ and σ are independent under $\mathbf{Q}$, the $\mathbf{Q}$ distribution of $\ln(S_T/S_t)$ given by $d\ln S_t = (r(t)-(\sigma(t)^2/2)dt+\sigma(t)d\tilde{w}^1(t)$ conditionally on both $\mathcal{F}_t$ and the whole volatility path $(\sigma(t))_{t\in[0,T]}$ is Gaussian with mean $\int_t^T r(u)du - \frac{1}{2}\int_t^T \sigma(u)^2 du$ and variance $\int_t^T \sigma(u)^2 du$. Therefore, computing the expectation (4.1) conditionally on the volatility path gives:

$$C_t = S(t)\left\{\mathbb{E}_t^{\mathbf{Q}}\left[\Phi\left(\frac{m_t}{U_{t,T}}+\frac{U_{t,T}}{2}\right)\middle|\,\mathcal{F}_t\right] - e^{-m_t}\mathbb{E}_t^{\mathbf{Q}}\left[\Phi\left(\frac{m_t}{U_{t,T}}-\frac{U_{t,T}}{2}\right)\middle|\,\mathcal{F}_t\right]\right\}, \tag{4.2}$$

where $m_t = \ln\left(\frac{S(t)}{KB(t,T)}\right)$, $U_{t,T} = \sqrt{\int_t^T \sigma(u)^2\,du}$, and $\Phi(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^u e^{-t^2/2}\,dt$.

The dynamics of σ are now given by

$$\ln\frac{\sigma(t)}{\sigma(0)} = \left(-k\int_0^t \ln\sigma(u)\ du - \gamma\int_0^t \frac{(t-s)^\alpha}{\Gamma(1+\alpha)}\lambda_2(s)\,ds\right) + \gamma\tilde{w}_\alpha^2(t),$$

where

$$\tilde{w}_\alpha^2(t) = \int_0^t \frac{(t-s)^\alpha}{\Gamma(1+\alpha)}\,d\tilde{w}^2(s).$$

Then differentiating $x(t) = \ln\sigma(t)$ with fractional order α gives:

$$dx^{(\alpha)}(t) = (-kx^{(\alpha)}(t) + \gamma\lambda_2(t))dt + \gamma d\tilde{w}^2(t), \tag{4.3}$$

where

$$x^{(\alpha)}(t) = \frac{d}{dt}\int_0^t \frac{(t-s)^{-\alpha}}{\Gamma(1-\alpha)}x(s)\ ds$$

is the derivative of (fractional) order α of x.

We can give the general solution of (4.3):

$$x^{(\alpha)}(t) = \left(c + \int_0^t \gamma e^{ks}\lambda_2(s)ds + \int_0^t \gamma e^{ks}d\tilde{w}^2(s)\right)e^{-kt}$$

and deduce x by fractional integration.

As usual, when one wants to perform statistical inference using arbitrage pricing models, two approaches can be imagined: either specify a given parametric form of the risk premium or assume that the associated risk is not compensated. When trading of volatility is observed it might be relevant to assume a risk premium on it. But we choose here, for the sake of simplicity (see, e.g., Engle and Mustafa 1992 or Pastorello et al. 1993 for similar strategies in short-memory settings) to assume that the volatility risk is not compensated, i.e., that $\lambda_2 = 0$. Under this simplifying assumption, which has some microeconomics foundations (see Pham and Touzi 1996), the probability distributions of $U_{t,T}$ are the same under $\mathbb{P}$ and under $\mathbf{Q}$. In other words the expectation operator in the option pricing formula (4.2) can be considered with respect to $\mathbb{P}$.

4.2. Implied Volatilities

Practitioners are used to computing the so-called Black–Scholes implicit volatility by inversion of the Black–Scholes option pricing formula on the observed option prices. If we assume that these option prices are given by (4.2) and that the volatility risk is not compensated, the Black–Scholes implicit volatility appears to be a forecast of the average volatility $\sigma_{t,T}$ on the lifetime of the option ($\sigma_{t,T}^2 = (T-t)^{-1}U_{t,T}^2$). If we consider the proxy of the option price (4.2) deduced from a first-order Taylor expansion (around $(T-t)^{-1}\mathbb{E}U_{t,T}^2$) of the Black–Scholes formula considered as a function of $\sigma_{t,T}^2$, the Black–Scholes implicit volatility dynamics would be directly related to the dynamics of

$$\sigma_{imp,T}^2(t) = \frac{1}{T-t}\int_t^T \mathbb{E}\left(\sigma^2(u) \mid \mathcal{F}_t\right) du.$$

To describe the dynamics of this "implicit volatility" we start by analyzing the conditional laws and moments of σ:

$$\mathbb{E}(\sigma(t+h) \mid \mathcal{F}_t) = \exp\left(g(t+h) + \int_0^t a(t+h-s)\, dw^2(s) + \frac{1}{2}\int_0^h a^2(x)\, dx\right)$$

for $x(t) = \ln\sigma(t) = g(t) + \int_0^t a(t-s)\, dw^2(s)$, $g(t) = x(0) + (1-e^{-kt})\theta$, and $a(x)$ as usual. Or, if we work with the stationary version of σ:

$$\mathbb{E}(\sigma(t+h) \mid \mathcal{F}_t) = \exp\left(\int_{-\infty}^t a(t+h-s)\, dw^{(2)}(s) + \frac{1}{2}\int_0^h a^2(x)\, dx\right).$$

To have an idea of the behavior of the implicit volatility, we can prove:

PROPOSITION 4.1. *$y_t = \mathbb{E}(\sigma^2(t+1) \mid \mathcal{F}_t)$ is a long-memory process in the sense that* cov (y_t, y_{t+h}) *is of order $O(|h|^{2\alpha-1})$ for $h \to +\infty$ and $\alpha \in]0,\ 1/2[$.*

Var($\mathbb{E}(\sigma(t+h) \mid \mathcal{F}_t)$) *is of order $O(|h|^{2\alpha-1})$ for $h \to +\infty$ if $\alpha \in]0,\ 1/2[$ and of order $e^{-k|h|}$ if $\alpha = 0$.*

Proposition 4.1 shows that, thanks to the long-memory property of the instantaneous volatility process, the stochastic feature of forecasted volatility does not vanish at the very high exponential rate but at the lower hyperbolic rate. This rate of convergence explains the stochastic feature of implicit volatilities, even for fairly long maturity options.

Since $T > t$, we can set $T = t + \tau$. We take $\tau = 1$ for simplicity and study the long-memory properties of the stationary (if we work with the stationary version of σ) process which is now defined by

$$\sigma_{imp}^2(t) = \int_0^1 \mathbb{E}\left(\sigma^2(t+u) \mid \mathcal{F}_t\right) du.$$

PROPOSITION 4.2. *$z_t := \sigma_{imp}^2(t)$ is a long-memory process in the sense that* cov(z_t, z_{t+h}) *is of order $O(|h|^{2\alpha-1})$ for $h \to +\infty$ and $\alpha \in]0,\ 1/2[$.*

We have already documented (see Section 6.5) some empirical evidence to confirm the theoretical result of Proposition 4.2. Indeed, when we use daily data on CAC40 and option prices on CAC40 (of the Paris Stock Exchange) and we try to estimate a long-memory parameter by regression on the log-periodogram (see Robinson 1996), we find that the stock price process S is a short-memory process and the B.S. implicit volatility process is a long-memory one.

Finally, the dynamics of conditional heteroskedasticity of the stock price process S can be described through the marginal kurtosis. We are not only able to prove a convergence property like Corollary 3.2 of Drost and Werker (1996) but also to measure the effect of the long-memory parameter on the speed of convergence:

PROPOSITION 4.3. *Let $\varphi(h) = \mathbb{E}|Y(h) - \mathbb{E}Y(h)|^4 = \mathbb{E}Z(h)^4$ denote the fourth centered moment of the rate of return $Y(h) = \ln \frac{S(h)}{S(0)}$ on $[0, h]$, with $Z(t) = \int_0^t \sigma(u)\, dw^1(u)$. Then $\varphi(h)/h^2$ is bounded on $\mathbb{R}$.*

Moreover, let $kurt_Y(h) = \varphi(h)/(Var\,Y(h))^2$ denote the kurtosis coefficient of $Y(h)$. Then

$$\lim_{h\to 0} kurt_Y(h) = 3\frac{\mathbb{E}(\sigma^4)}{(\mathbb{E}(\sigma^2))^2} > 3, \quad \textit{for } \alpha \in \left[0, \frac{1}{2}\right[$$

at rate $h^{2\alpha+1}$ (continuity in $\alpha = 0$),[6] *and $\lim_{h\to+\infty} kurt_Y(h) = 3$ for $\alpha \in [0, \frac{1}{2}[$ at rate $h^{2\alpha-1}$ if $\alpha \in]0, \frac{1}{2}[$,*[7] *and at rate $e^{-(k/2)h}$ if $\alpha = 0$.*

The discontinuity in 0 of the speed of convergence of $\lim_{h\to+\infty} kurt_Y(h)$ with respect to α is additional evidence of the persistence in volatility introduced by the α parameter. When there is long memory ($\alpha > 0$) the leptokurtic feature due to conditional heteroskedasticity vanishes with temporal aggregation at a slow hyperbolic rate, while with a usual short-memory volatility process it vanishes at an exponential rate.

Note that the limit for h going to 0 of $kurt_Y(h)$ is close to 3 (and thus the log-return Y is close to Gaussian) if and only if Var σ^2 is close to 0, that is, if σ is close to deterministic (small value of the diffusion coefficient γ); this leads us back to the standard Black–Scholes world.

5. STATISTICAL INFERENCE IN THE FSV MODEL

5.1. Statistical Inference from Stock Prices

Several methods are provided in Comte and Renault (1996) and Comte (1996) to estimate the parameters of an "Ornstein–Uhlenbeck long-memory" process, which here is the set of parameters $(\alpha, k, \theta, \gamma)$ implied by the first-order equation fulfilled by the log-volatility process. Those methods of course are all based on a discrete time sample of observations of one path of $\ln\sigma$. Such a path is not available here.

The idea then is to find approximations of the path deduced from the observed $S(t_i)$ and to replace the true observations usually used by their approximations in the estimation procedure. Let us recall briefly that those procedures are as follows:

[6] That is, $kurt_Y(h) - 3[\mathbb{E}(\sigma^4)/(\mathbb{E}(\sigma^2))^2]$ is of order $h^{2\alpha+1}$ for $h \to 0$.
[7] That is, $kurt_Y(h) - 3$ is of order $h^{2\alpha-1}$ for $h \to +\infty$.

- either we find α by log-periodogram regression using the semiparametric results of Robinson (1996) and (k, θ, γ) by estimating an AR(1) process after fractional differentiation at the estimated order,
- or all parameters are estimated by minimizing the Whittle-type criterium approximating the likelihood in the frequency domain, as studied by Fox and Taqqu (1986) and Dahlhaus (1989).

The natural idea for approximating σ is then based on the quadratic covariation of $Y(t) = \ln(S(t))$. Indeed, $\langle Y \rangle_t = \int_0^t \sigma^2(s)\, ds$ and, if $\{t_1, \ldots, t_m\}$ is a partition of $[0, t]$ and $t_0 = 0$, then

$$\lim_{step \to 0} \sum_{k=1}^{m} (Y_{t_k} - Y_{t_{k-1}})^2 = \langle Y \rangle_t \text{ in probability, where step} = \operatorname*{Max}_{1 \le i \le m} \{|t_i - t_{i-1}|\}.$$

Then as $(\langle Y \rangle_t - \langle Y \rangle_{t-h})/h \underset{h \to 0}{\longrightarrow} \sigma^2(t)$ a.s. and provided that high-frequency data are available, we can think of cumulating the two limits by considering a partition of the partition to obtain estimates of the derivative of the quadratic variation.

Let $[0, T]$ be the interval of observation, let $t_k = kT/N$, $N = np$, be the dates of observations, and let Y_{t_k}, $k = 0, \ldots, N$, be the sample of observations of the log-prices. Then we have n blocks of length p and we set: $\widehat{\langle Y \rangle}_t^{(N)} = \sum_{k=0}^{[(tN)/T]=[nt]} (Y_{t_k} - Y_{t_{k-1}})^2$ so that $(\widehat{\langle Y \rangle}_t^{(N)} - \widehat{\langle Y \rangle}_{t-h}^{(N)})/h$ is computed from the underlying blocks with $h = T/n$. In other words,

$$\hat{\sigma}_{n,p}^2(t) = \frac{n}{T} \sum_{k=[\frac{tN}{T}]-p+1}^{[\frac{tN}{T}]} (Y_{t_k} - Y_{t_{k-1}})^2$$

because $[((t - (T/n)N)/T] = [tN/T - p]$. Then we have:

PROPOSITION 5.1. *Let* $Y(t) = \int_0^t \sigma(s)\, dw^1(s)$ *and* $\sigma(t) = \tilde{\sigma}(t)$ *with* $\tilde{\sigma}$ *given by formula (2.7). Then* $\forall \varepsilon > 0$,

$$\lim_{\substack{n \to +\infty \\ p \to +\infty}} \sup_{t \in [0,T]} p^{1-\varepsilon} \mathbb{E} \left(\hat{\sigma}_{n,p}^2(t) - \sigma^2(t) \right)^2 = 0.$$

Thus p must be as large as possible for the rate of convergence to be optimal. On the other hand we are interested in large sizes n of the sample of deduced volatilities. This is the reason there is a trade-off between n and p, taking into account the constraint $N = np$. A possible choice could be to choose n and p of order $\sqrt{N}$.

Then we have to estimate μ, supposed to be a constant, and we notice that the finite variation terms that have been omitted in Y are known to have no weight in the quadratic covariation. The estimate of μ can be chosen here as usual (see Renault and Touzi 1996): $e^{\hat{\mu}_{np} \Delta t} = \frac{1}{np} \sum_{k=1}^{np} \frac{S_{t_k}}{S_{t_{k-1}}}$, $\Delta t = \frac{T}{np}$, $t_k = \frac{kT}{np}$, or sometimes: $\tilde{\mu}_{np} = \frac{np}{T} \frac{1}{np} \sum_{k=1}^{np} \frac{S_{t_k} - S_{t_{k-1}}}{S_{t_{k-1}}}$, which completes the estimation procedure.

5.2. Statistical Inference from Option Prices

Another way to estimate the volatility parameters could be the use of the informational content of option prices and associated implied volatilities in the spirit of Engle and Mustafa (1992) or Pastorello et al. (1993) (assuming that the volatility risk is not compensated). Unfortunately, the non-Markovian feature of the long-memory process implies that the Hull and White option pricing formula is not so simple to invert to recover latent instantaneous volatilities as in the usual case. Nevertheless, if sufficiently high frequency data are available to approximate integrals by finite sums, we are able to generalize the Pastorello et al. (1993) procedure thanks to a first-stage estimate of the long-memory parameter α. To see this point, let us assume for instance that we observe at times t_i, $i = 0, 1, \ldots, n$, option prices C_{t_i} for options of exercise dates $t_i + \Delta$ (for a fixed Δ), that are at the money in the generalized sense: $S_{t_i} = K_{t_i} B(t_i, t_i + \Delta)$, where K_{t_i} is the exercise price of an option traded at date t_i. In this case, we know from (4.2) that:

$$C_{t_i} = S_{t_i}\left(2\mathbb{E}^{\mathbb{P}}\left(\Phi\left(\frac{U_{t_i,t_i+\Delta}}{2}\right)\right) - 1 \,\middle|\, \mathcal{F}_{t_i}\right). \tag{5.1}$$

The information set $\mathcal{F}_{t_i}$ in the above expectation is defined as the sigma-field generated by $(w^1(\tau), \sigma(\tau), \tau \le t_i)$. But since the two processes w^1 and σ are independent and $U_{t_i,t_i+\Delta}$ is depending on σ only, the information provided by $w^1(\tau), \tau \le t_i$ is irrelevant in the expectation (5.1). Moreover, thanks to (2.4) and (2.3), we know that the sigma-field generated by $x(\tau) = \ln \sigma(\tau)$, $\tau \le t_i$, coincides with the sigma-field generated by the short-memory process $x^{(\alpha)}(\tau)$, $\tau \le t_i$. On the other hand, thanks to (2.3), $U_{t_i,t_i+\Delta}$ appears like a complicated function (see Appendix B) of $x^{(\alpha)}(\tau)$, $\tau \le t_i + \Delta$.

In other words, (5.1) gives the option price as a function of:

- first, the past values $x^{(\alpha)}(\tau)$, $\tau \le t_i$, which define the deterministic part of $U_{t_i,t_i+\Delta}$,
- second, the Ornstein–Uhlenbeck parameters (k, θ, γ), which characterize the conditional probability distribution of $x^{(\alpha)}(\tau)$, $\tau > t_i$, given the available information $\mathcal{F}_{t_i}$ summarized by $x^{(\alpha)}(t_i)$,
- third, the long-memory parameter α, which defines the functional relationship between $U_{t_i,t_i+\Delta}$ and the process $x^{(\alpha)}$.

The Black–Scholes implicit volatility $\sigma^{BS}_{imp}(t_i)$ is by definition related to the option price C_{t_i} in a one-to-one fashion by

$$C_{t_i} = S_{t_i}\left[2\Phi\left(\frac{\sqrt{\Delta}\sigma^{BS}_{imp}(t_i)}{2}\right) - 1\right]. \tag{5.2}$$

The comparison of (5.1) and (5.2) shows that the dynamics of $\sigma^{BS}_{imp}(t_i)$ are determined not only by the dynamics of the Ornstein–Uhlenbeck process $x^{(\alpha)}$ but also by the complicated functional relationship between $U_{t_i,t_i+\Delta}$ and the past values of $x^{(\alpha)}$. This is why the BS implicit volatility is itself a long-memory process whose dynamics cannot analytically be related to the dynamics of the instantaneous latent volatility. Nevertheless, the relationship between $U_{t_i,t_i+\Delta}$ and $x^{(\alpha)}$ can be approximated by (see Appendix B):

$$(5.3)\ U^2_{t_i,t_i+\Delta} = \int_{t_i}^{t_i+\Delta} \exp\left(2\sum_{\tau<t_i} \frac{(u-\tau)^\alpha}{\Gamma(1+\alpha)} \Delta(x^{(\alpha)}(\tau))\right) \times \exp(f(x^{(\alpha)}(t_i);\, Z(u, t_i, \alpha))\, du,$$

where f is a deterministic function and $Z(u, t_i, \alpha)$ is a process independent of $\mathcal{F}_{t_i}$.

Thanks to Proposition 4.2, we can estimate α in a first step by a log-periodogram regression on the implicit volatilities. In a second stage, we shall assume that α is known (the lack of accuracy due to estimated α will not be considered here) and we propose the following scheme for an indirect inference procedure, in the spirit of Pastorello et al. (1993):

$$\begin{array}{ccccccccc} \theta & \xrightarrow{\text{simulation of}} & x^{(\alpha)}(t;\theta) & \longrightarrow & \tilde{C}_t(\theta) & \xrightarrow{BS^{-1}+\alpha\text{-filter}} & (\ln\tilde{\sigma})^{(\alpha)}_{imp}(\theta) & \longrightarrow & \tilde{\beta}(\theta) \\ & & & & C_t & \xrightarrow{BS^{-1}+\alpha\text{-filter}} & (\ln\sigma)^{(\alpha)}_{imp} & \longrightarrow & \hat{\beta}. \end{array}$$

The meaning of this scheme is the following: for a given value θ of the parameters of the Ornstein–Uhlenbeck process $x^{(\alpha)}$, we are able to simulate a sample path $x^{(\alpha)}(t;\theta)$; then thanks to (5.3) and (5.1), we can get simulated values $\tilde{C}_t(\theta)$ conformable to the Hull–White pricing. Of course, this procedure is computer intensive since the expectation (5.1) itself has to be computed by Monte Carlo simulations. Nevertheless, as soon as option prices $\tilde{C}_t(\theta)$ are available, the associated Black–Scholes implicit volatilities $\tilde{\sigma}_{imp}(\theta)$ are easy to compute, and finally, through the fractional differential operator, we obtain a process $(\ln\tilde{\sigma}_{imp})^{(\alpha)}(t;\theta)$ whose dynamics should mimic the dynamics of the Ornstein–Uhlenbeck process $x^{(\alpha)}$.

This proxy of the instantaneous volatility dynamics provides the basis of our indirect inference procedure. More precisely, $\tilde{\beta}(\theta)$ (respectively, $\hat{\beta}$) denotes the pseudomaximum likelihood estimator of the parameters of the simulated process $(\ln\tilde{\sigma}_{imp})^{(\alpha)}(t;\theta)$ (respectively, the observed process $(\ln\sigma_{imp})^{(\alpha)}(t)$) when the pseudolikelihood is defined by an Ornstein–Uhlenbeck modeling of these processes.

The basic idea of the indirect inference procedure is to compute a consistent estimator of the structural parameters θ by solving in θ the equations $\tilde{\beta}(\theta) = \hat{\beta}$.

It is clear that the consistency proof of Gourieroux, Monfort, and Renault (1993) can be easily extended to this setting thanks to the ergodicity property of the processes; on the other hand, the asymptotic probability distributions have to be reconsidered to take into account the long-memory features.

6. SIMULATION AND EXPERIMENTS

6.1. Simulation of the Path of a Fractional Stochastic Volatility Price

First, we illustrate in Figure 6.1 the general behavior of the sample path of $Y = \ln S$ generated by small step discretization, $h(= 1/n) = 0.02$ and $\mu = \theta = 0$, $(\alpha, k, \gamma) = (0.3, 1, 0.01)$. We can see that we obtain paths that are very similar to what is observed for exchange rates. Compare this for instance with the graph given by Baillie et al. (1996) for DM–US dollar exchange rate. In both cases, there seems to be a “long memory” of the main pick that seems to appear again after its occurrence, even if attenuated.

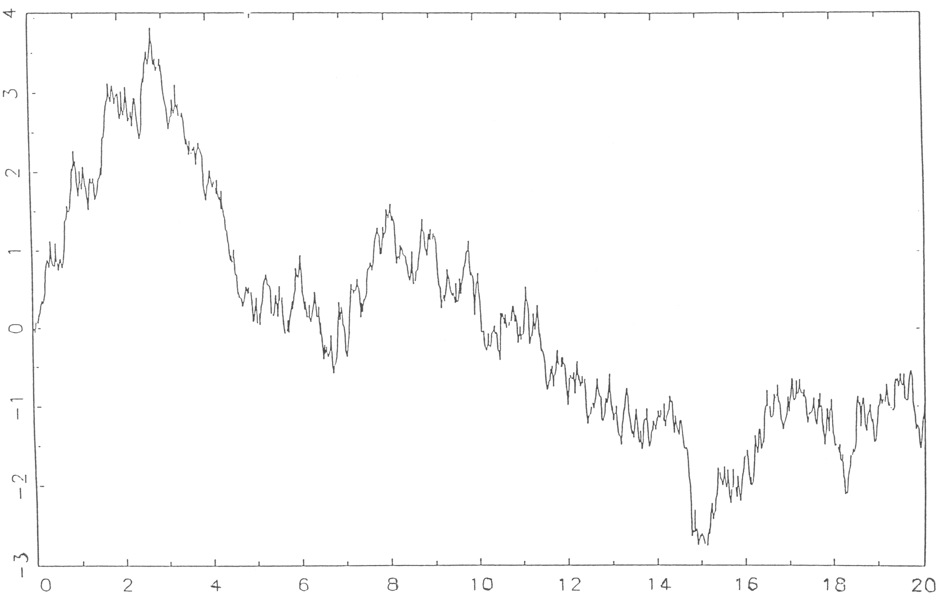

FIGURE 6.1. Simulated path of log-stock price in the long-memory FSV model; $N = 1000$, $h = 0.02$, $(\alpha, k, \gamma) = (0.3, 1, 0.01)$.

6.2. An Apparent Unit Root

Another comparison can be made with Baillie et al.'s (1996) work. Indeed, they argue that their discrete time fractional model gives another representation of persistence that can remain stationary, contrary to usual unit roots models.

Here, we want to show that our model may exhibit an apparent unit root if a wrong parameterization is assumed for estimation. For that purpose, we look at what is obtained if the model is estimated as if it were a GARCH(1,1) process:

$$\begin{cases} \varepsilon_t = \ln S_t = \sigma_t z_t, \qquad \mathbb{E}_{t-1} z_t = 0, \mathrm{Var}_{t-1} z_t = 1 \\ \sigma_t^2 = \omega + a\varepsilon_{t-1}^2 + b\sigma_{t-1}^2. \end{cases}$$

In other words,$(1 - \varphi L)\varepsilon_t^2 = \omega + (1 - bL)\nu_t$ where $\varphi = a + b$ and ν is a white noise.

We estimate the parameter (ω, φ, b) through minimizing $l(\theta, \varepsilon_1, \ldots, \varepsilon_T) = \sum_{t=1}^{T}(\ln \sigma_t^2 + \varepsilon_t^2 \sigma_t^{-2})$. The results are reported in Table 6.1 and Figure 6.2. One hundred forty samples have been generated, starting with 5000 points (with a step $hh = 0.1$) and with one point out of ten (i.e., 500 points), kept for the estimation procedure with a step $h = 1$ and $(\alpha, k, \gamma) = (0.3, 2, 1)$. We find also an apparent unit root for φ, and the empirical distribution of φ clearly appears to be centered at 1. We can see that φ is the more stable of the estimated coefficients and is always very near 1. Other tries have been made with other parameters for the continuous time model with the same kinds of results. Thus, continuous-time fractional models are good representations of apparent persistence.

TABLE 6.1
An Apparent Unit Root for φ. A GARCH(1,1) Is Estimated Instead of the True FSV Simulated. 140 Samples Generated

	Empirical mean	Empirical std. dev.
ω	1.5843	1.3145
φ	1.0453	0.1109
b	-0.1935	0.3821

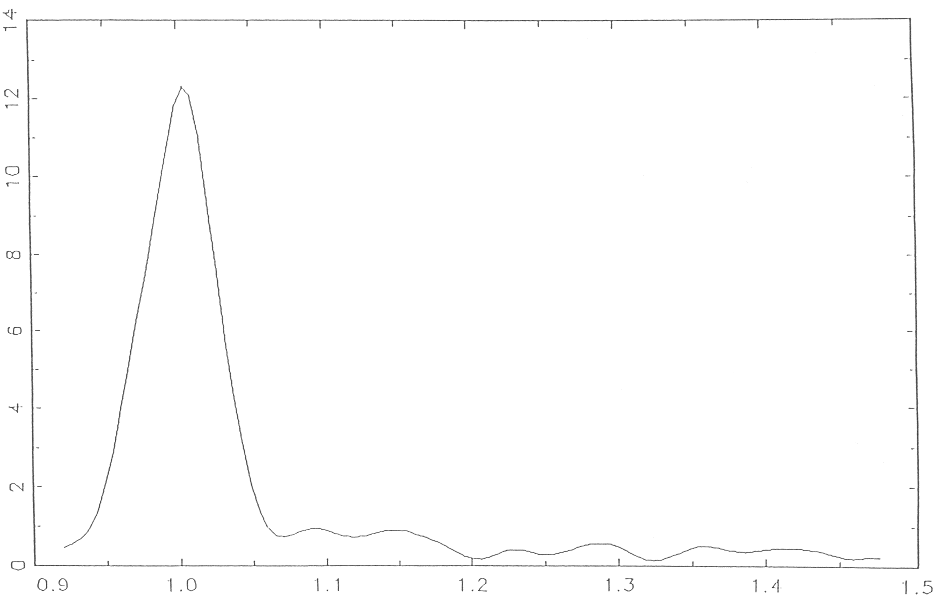

FIGURE 6.2. Empirical distribution of φ when the FSV model is estimated as a GARCH (1,1) process; $\epsilon_t = \sigma_t z_t$, $\sigma_t^2 = \omega + a\epsilon_{t-1}^2 + b\omega_{t-1}^2$, $\varphi = a + b$; 140 samples generated.

6.3. Comparison of the Filters

Now we give an illustration of the quality of the continuous-time filter defined by $a_i = ((i+1)^\alpha - i^\alpha)$ (see Section 3.2) as compared with the usual discrete time one $(1-L)^{-\alpha}$.

We generated N observations at step $hh = 0.01$ of the AR(1) process $x^{(\alpha)}$ as given in formula (3.3) with $(\alpha,\ k,\ \gamma) = (0.3, 3, 1)$, which gave an N-sample of x as given by (3.2). Then we kept $n = N/10$ observations of the true AR(1), $x^{(\alpha)}$, and of the x process. We applied the continuous-time filter at step $h = 10hh = 0.1$ to the n-sample x, which gave observations of a process $x_1^{(\alpha)}$; we applied the discrete time filter at step $h = 10hh$ to the same n-sample x, which gave observations of a process $x_2^{(\alpha)}$. The paths of $x^{(\alpha)}$, $x_1^{(\alpha)}$, and $x_2^{(\alpha)}$ can be compared. It appears that the continuous-time filter is better than the discrete time one.

We generated 100 such samples of $x^{(\alpha)}$, $x_1^{(\alpha)}$, and $x_2^{(\alpha)}$ and computed L^1 and L^2 distances

TABLE 6.2
L^1 and L^2 Distances between the Original Paths and the Filtered Paths with the Two Filters. Filter 1 Is the Continuous-Time Filter, Filter 2 the Discrete Time One

	$x^{(\alpha)}, x_1^{(\alpha)}$	$x^{(\alpha)}, x_2^{(\alpha)}$
d_{L^1}	0.1048	0.2135
d_{L^2}	0.1359	0.3607

between $x^{(\alpha)}$ and $x_i^{(\alpha)}$, $i = 1, 2$; that is,

$$d_{L^1}(x^{(\alpha)}, x_i^{(\alpha)}) = \frac{1}{n}\sum_{j=1}^{n} |x^{(\alpha)}(j) - x_i^{(\alpha)}(j)|,$$

$$d_{L^2}(x^{(\alpha)}, x_i^{(\alpha)}) = \frac{1}{n}\sum_{j=1}^{n}(x^{(\alpha)}(j) - x_i^{(\alpha)}(j))^2, \quad i = 1, 2.$$

The results are reported in Table 6.2. Even if the numbers do not have any meaning in themselves, the comparison leads clearly to the conclusion that the first filter is significantly better. For a convincing comparison of the twelve first partial autocorrelations of the three samples, see Comte (1996).

6.4. Estimation of α by Log-Periodogram Regression in Three Models

Lastly, we compared the estimations of α obtained by regression of $\ln I(\lambda)$ on $\ln \lambda$, where $I(\lambda)$ is the periodogram (see Geweke and Porter-Hudak 1983 for the idea, Robinson 1996 for the proof of the convergence and asymptotic normality of the estimator, and Comte 1996 to check the assumptions given by Robinson).

We used 100 samples with length 400, where 4000 points were generated for the continuous-time models with a step 0.1 and one point out of ten was kept for the estimation. We had $(\alpha,\ k,\ \gamma) = (0.3, 3, 1)$, in particular $\alpha = 0.3$ in all cases.

But we compared two ways of estimating α: either working directly on the log-periodogram of the process $x(t) = ln\sigma(t)$ (which exactly corresponds to our fractional Ornstein–Uhlenbeck model) or working on $\sigma(t) = exp(x(t))$, since it fulfills the same long-memory properties (see Proposition 2.2).

As a benchmark for this estimation of α, we considered a third estimation through the following procedure. Assuming that the observed path would be associated with $\tilde{x}(t) = (1-L)^{-\alpha}x^{(\alpha)}(t)$ (with a sampling frequency $h = 1$) instead of $x(t)$, we could then estimate α by a log-periodogram regression on the path $\tilde{x}(t)$, which is referred to below as the ARFIMA method. The three methods should provide consistent estimators of the same value for α. The results are reported in Table 6.3: they are better with x than with an ARFIMA model or with $\exp\ x$, and the recommendation is to work with x instead of $\exp x$. Let us nevertheless notice that the bad result for the ARFIMA model could be explained by the fact that the discrete time filter $(1-L)^{-\alpha}$ had been applied to the low frequency path $x^{(\alpha)}(t)$ with step $h = 1$.

TABLE 6.3
Estimation of α with Log-Periodogram Regression for Three Models. 100 Samples, Length 400. True Value: $\alpha = 0.3$

	Empirical mean	Empirical std. dev.
x	0.2877	0.0629
$\exp x$	0.2568	0.0823
ARFIMA	0.3447	0.0864

6.5. Estimation with Real Data

We carried out log-periodogram regressions on real stock prices and implicit volatilities associated with options on CAC40 of the Paris Stock Exchange. First, for the log-prices, we found from our 775 daily data that $\alpha_{\text{price}} = 0.0035$. This is near $\alpha = 0$ and confirms the short-memory feature of prices. On the contrary, we found for a sequence $|\sigma_{t+1} - \sigma_t|$ that $\alpha_{\text{volat}} = 0.2505$, which illustrates the long-memory feature of the volatilities. We emphasize that we have to take absolute values of the increments of the implicit volatilities (see Ding, Granger, and Engle 1993), since we otherwise have:

	$\ln(\sigma_{t+1}/\sigma_t)$	$\sigma_{t+1} - \sigma_t$	σ_t	$(\sigma_{t+1} - \sigma_t)^2$
α	-0.30493	-0.37589	0.67034	0.21490

so that squaring is also possible. Missing values are replaced by the global mean.

Two preliminary conclusions can be derived from the previous empirical evidence. First, it appears that the volatility is not stationary and must be differenced. Secondly, the long-memory phenomenom is stronger when we consider absolute values (or squared values) of the differenced volatility. This seems to indicate some asymmetric feature in the volatility dynamics, as observed in asset prices by Ding et al. (1993).

These two points lead us to modify our long-memory diffusion equation on $\sigma(t)$. This work is still in progress. Nevertheless, the previous empirical evidence has to be interpreted cautiously, because if we take into account small sample biases, it is clear that an autoregressive operator $(1 - \rho L)$ with ρ close to one is empirically difficult to identify against a fractional differentiation $(1 - L)^\alpha$.

APPENDIX A: PROOFS

Proof of Proposition 2.1. We are going to use the result given in equation (2.6) (see Comte 1996) which gives for $\tilde{x} = \ln(\tilde{\sigma})$ and $h \to 0$: $r_{\tilde{x}}(h) = r_{\tilde{x}}(0) + \psi.|h|^{2\alpha+1} + o(|h|^{2\alpha+1})$ with, for $h \geq 0$:

$$V(\tilde{x}(t+h) - \tilde{x}(t)) = 2(r_{\tilde{x}}(0) - r_{\tilde{x}}(h)) = 2\left(\int_0^{+\infty} a^2(x)\,dx - \int_0^{\infty} a(x)a(x+h)\,dx\right),$$

and the fact that if $X \rightsquigarrow \mathcal{N}(0, s^2)$ then $\mathbb{E}(\exp X) = e^{s^2/2}$. Then, still for $h \geq 0$ (and $r_\sigma(-h) = r_\sigma(h)$):

$$r_\sigma(h) = \mathbb{E}(\exp(\tilde{x}(t+h) \times \exp(\tilde{x}(t)) - \mathbb{E}(\exp(\tilde{x}(t+h)) \times \mathbb{E}(\exp(\tilde{x}(t))$$

$$\begin{aligned} r_\sigma(h) \; = \; & \mathbb{E}\left(\exp\left(\int_{-\infty}^{t} (a(t+h-s) + a(t-s))\, dw^2(s)\right)\right. \\ & \quad \left. \times \; \exp\left(\int_{t}^{t+h} a(t+h-s)\, dw^2(s)\right)\right) \\ & - \mathbb{E}\left(\exp\left(\int_{-\infty}^{t+h} a(t+h-s)\, dw^2(s)\right)\right) \times \mathbb{E}\left(\exp\left(\int_{-\infty}^{t} a(t-s)\; dw^2(s)\right)\right) \end{aligned}$$

This yields, for $h \geq 0$, with the second point, to

$$r_\sigma(h) = \exp\left(\int_0^{+\infty} a^2(x)\, dx\right)\left(\exp\left(\int_0^{+\infty} a(x+h)a(x)\, dx\right) - 1\right).$$

Then, for $h \geq 0$:

$$\begin{aligned} r_\sigma(h) - r_\sigma(0) \; = \; & \exp\left(\int_0^{+\infty} a^2(x)\, dx\right) \\ & \times \; \left(\exp\left(\int_0^{+\infty} a(x+h)a(x)\, dx\right) - \exp\left(\int_0^{+\infty} a^2(x)\, dx\right)\right), \end{aligned}$$

and factorizing the first right-hand term again:

$$r_\sigma(h) - r_\sigma(0) = \exp\left(2\int_0^{+\infty} a^2(x)\, dx\right)(\exp\left(r_{\tilde{x}}(h) - r_{\tilde{x}}(0)\right) - 1).$$

Then for $h \to 0$, $K = \exp(\int_0^{+\infty} a^2(x)\, dx)$, we have:

$$r_\sigma(h) - r_\sigma(0) = K^2\left(\exp(\psi.|h|^{2\alpha+1} + o(|h|^{2\alpha+1})) - 1\right) = K^2\psi|h|^{2\alpha+1} + o(|h|^{2\alpha+1})$$

which gives the announced result. □

Proof of Proposition 2.2. (i) The previous computations give, with the same K as above: $r_\sigma(h) = K\,(\exp(r_{\tilde{x}}(h)) - 1)$ and it has been proved in Comte (1996) that $r_{\tilde{x}}(h) = \mu|h|^{2\alpha-1} + o(|h|^{2\alpha-1})$ for $h \to +\infty$, where μ is a constant. This implies straightforwardly that $r_\sigma(h) = K\mu|h|^{2\alpha-1} + o(|h|^{2\alpha-1})$, which gives (i). □

(ii) $\int_0^{+\infty} r_\sigma(h)\cos(\lambda h)dh = \int_0^A r_\sigma(h)\cos(\lambda h)dh + \frac{1}{\lambda}\int_{\lambda A}^{+\infty} r_\sigma(\frac{u}{\lambda})\cos u du$. Now for A chosen great enough, the development of r_σ near $+\infty$ implies $\lambda^{2\alpha} f_\sigma(\lambda) = \lambda^{2\alpha}\int_0^A r_\sigma(h)\cos(\lambda h)dh + \int_{\lambda A}^{+\infty} u^{2\alpha-1}\cos u du + o(1)$, and consequently $\lim_{\lambda\to 0}\lambda^{2\alpha} f_\sigma(\lambda) = \int_0^{+\infty} u^{2\alpha-1}\cos u du$ where the integral is convergent near 0 because $2\alpha > 0$ and near $+\infty$ because: $\int_1^{+\infty} u^{2\alpha-1}\cos u du = [u^{2\alpha-1}\sin u]_1^{+\infty} - (2\alpha-1)\int_1^{+\infty} u^{2\alpha-2}\sin u du$ where all terms are obviously finite. □

Proof of Lemma 3.1. For the proof of the first convergence: $Y_n \overset{\mathcal{D}}{\Rightarrow} Y$ on a compact set $[0,T]$, we check the L^2 pointwise convergence of $Y_n(t)$ toward $Y(t)$, and then a tightness criterion as given by Billingsley (1968, Th. 12.3): $\mathbb{E}|Y_n(t_2) - Y_n(t_1)|^p \le C.|t_2 - t_1|^q$ with $p > 0$, $q > 1$, and C a constant.

The L^2 convergence is ensured by computing:

$$\begin{aligned}\mathbb{E}(Y_n(t) - Y(t))^2 &= \mathbb{E}\left(\int_0^t \left(\sigma\left(\frac{[ns]}{n}\right) - \sigma(s)\right) dw^1(s)\right)^2\\ &= \mathbb{E}\left(\int_0^t \left(\sigma\left(\frac{[ns]}{n}\right) - \sigma(s)\right)^2 ds\right)\\ &= \int_0^t \mathbb{E}\left(\sigma\left(\frac{[ns]}{n}\right) - \sigma(s)\right)^2 ds \quad \text{with Fubini.}\end{aligned}$$

Then the L^2 convergence is obviously given by an inequality: $\mathbb{E}(\sigma(t_2) - \sigma(t_1))^2 \le \tilde{C}.|t_2 - t_1|^\gamma$ for a positive γ and a constant $\tilde{C}$.

As usual, let $x(t) = \int_0^t a(t-s)\,dw^2(s)$ and let $t_1 \le t_2$.

$$\begin{aligned}\mathbb{E}(\sigma(t_2) - \sigma(t_1))^2 &= \mathbb{E}(\exp(x((t_2)) - \exp(x(t_1)))^2 = \mathbb{E}\left(e^{2x(t_2)} + e^{2x(t_1)} - 2e^{x(t_1)+x(t_2)}\right)\\ &= e^{2\int_0^{t_1} a^2(x)dx} + e^{2\int_0^{t_2} a^2(x)dx}\\ &\quad - 2e^{\frac{1}{2}\int_0^{t_1} a^2(x)dx + \frac{1}{2}\int_0^{t_2} a^2(x)dx + \int_0^{t_1} a(x)a(t_2-t_1+x)dx}\\ &= e^{2\int_0^{t_2} a^2(x)dx}\\ &\quad\times \left(1 + e^{-2\int_{t_1}^{t_2} a^2(x)dx} - 2e^{-\frac{3}{2}\int_{t_1}^{t_2} a^2(x)dx - \int_0^{t_1} a(x)(a(x)-a(t_2-t_1+x))dx}\right)\\ &\le 2e^{2\int_0^{t_2} a^2(x)dx}\left(1 - e^{-\frac{3}{2}\int_{t_1}^{t_2} a^2(x)dx - \int_0^{t_1} a(x)(a(x)-a(t_2-t_1+x))dx}\right).\end{aligned}$$

The term inside the last parentheses being necessarily nonnegative, the term in the last great exponential is nonpositive. Moreover $|\int_{t_1}^{t_2} a^2(x)\,dx| \le M_1^2|t_2 - t_1|$ with $M_1 = \sup_{x\in[0,T]}|a(x)|$, and since a is α-Hölder,

$$\left|\int_0^{t_1} a(x)(a(x) - a(t_2 - t_1 + x))\,dx\right| \le C_\alpha|t_2 - t_1|^\alpha \int_0^{t_1} |a(x)|\,dx \le C_\alpha|t_2 - t_1|^\alpha M_1 T,$$

which implies $|\int_{t_1}^{t_2} a^2(x)\,dx + \int_0^{t_1} a(x)(a(x) - a(t_2 - t_1 + x))\,dx| \le M_2|t_2 - t_1|^\alpha$, with

$M_2 = M_2(T)$. Then using that $\forall u \leq 0,\quad 0 \leq 1 - e^u \leq |u|$, we have $\mathbb{E}(\sigma(t_2) - \sigma(t_1))^2 \leq 2K^2 M_2 |t_2 - t_1|^\alpha, \quad \alpha \in]0, \frac{1}{2}[$ where $K = \exp(\int_0^{+\infty} a^2(x)dx)$ as previously. Then, $\mathbb{E}(\sigma([ns]/n) - \sigma(s))^2 \leq 2K^2 M_2 (\frac{1}{n})^\alpha$ gives:

$$\mathbb{E}(Y_n(t) - Y(t))^2 \leq \frac{2K^2 M_2 T}{n^\alpha} \qquad \forall t \in [0, T],$$

which ensures the L^2 convergence.

We use a straightforward version of Burkholder's inequality (see Protter 1992, p. 174), $\mathbb{E}|M_t|^p \leq C_p \mathbb{E}\langle M\rangle_t^{p/2}$, where C_p is a constant and M_t a continuous local martingale, $M_0 = 0$, to write (with an immediate adaptation of the proof on $[t_1, t_2]$ instead of $[0, t]$):

$$\mathbb{E}|Y_n(t_2) - Y_n(t_1)|^p = \mathbb{E}\left|\int_{t_1}^{t_2} \sigma\left(\frac{[ns]}{n}\right) dw^1(s)\right|^p \leq C_p \mathbb{E}\left|\int_{t_1}^{t_2} \sigma^2\left(\frac{[ns]}{n}\right) ds\right|^{p/2}.$$

Let us choose $p = 4$:

$$\begin{aligned}
\mathbb{E}|Y_n(t_2) - Y_n(t_1)|^4 &\leq C_4 \mathbb{E}\left|\int_{t_1}^{t_2} \sigma^2\left(\frac{[ns]}{n}\right) ds\right|^2 \\
&= C_4 \iint_{[t_1,t_2]^2} \mathbb{E}\left(\sigma^2\left(\frac{[nu]}{n}\right)\sigma^2\left(\frac{[nv]}{n}\right)\right) du\, dv \\
&\leq C_4 \iint_{[t_1,t_2]^2} \sqrt{\mathbb{E}\tilde{\sigma}^4 \times \mathbb{E}\tilde{\sigma}^4}\, du\, dv \quad (\tilde{\sigma} \text{ given by (2.7)}) \\
&= C_4 \mathbb{E}\tilde{\sigma}^4 (t_2 - t_1)^2, \qquad \mathbb{E}\tilde{\sigma}^4 = \exp\left(8\int_0^{+\infty} a^2(x)ds\right).
\end{aligned}$$

This gives the tightness and thus the convergence.

The second convergence is deduced from the first one, the decomposition: $Y_n(t) = \hat{Y}_n(t) + u_n(t)$, with $u_n(t) = \sigma(\frac{[nt]}{n})(w^1(t) - w^1(\frac{[nt]}{n}))$, and Theorem 4.1 of Billingsley (1968): $(X_n \overset{\mathcal{D}}{\Rightarrow} X$ and $\rho(X_n, Z_n) \xrightarrow{P} 0) \Rightarrow (Z_n \overset{\mathcal{D}}{\Rightarrow} X)$, where $\rho(x, y) = \sup_{t\in[0,T]} |x(t) - y(t)|$. Here $\rho(Y_n, \hat{Y}_n) = \sup |u_n(t)|$ and $u_n(t) = M([nt]/n)$ is a martingale so that Doob's inequality (see Protter 1992, p. 12, Th. 20) gives:

$$\mathbb{E}\left(\sup_{t\in[0,T]} |u_n(t)|\right)^2 \leq 4 \sup_{t\in[0,T]} \mathbb{E}(u_n(t)^2).$$

Then,

$$\begin{aligned}
\mathbb{E}u_n(t)^2 &= \mathbb{E}\left(\sigma^2\left(\frac{[nt]}{n}\right)\left(w^1(t) - w^1\left(\frac{[nt]}{n}\right)\right)^2\right) \\
&= \mathbb{E}\left(\sigma^2\left(\frac{[nt]}{n}\right)\right)\mathbb{E}\left(w^1(t) - w^1\left(\frac{[nt]}{n}\right)\right)^2
\end{aligned}$$

$$= \left(t - \frac{[nt]}{n}\right) \mathbb{E}\left(\sigma^2\left(\frac{[nt]}{n}\right)\right) \leq \frac{1}{n}\mathbb{E}(\tilde{\sigma}^2),$$

which achieves the proof. □

Proof of Proposition 3.2.

- First we prove the following implication:

$$\begin{cases} \tilde{Y}_n \overset{\mathcal{D}}{\Rightarrow} Y & \text{or tight} \\ \tilde{\sigma}_n \overset{\mathcal{D}}{\Rightarrow} \sigma & \text{or tight} \end{cases} \quad \text{and} \quad \begin{pmatrix} \tilde{Y}_n(t) \\ \tilde{\sigma}_n(t) \end{pmatrix} \xrightarrow{L^2} \begin{pmatrix} Y(t) \\ \sigma(t) \end{pmatrix} \quad \text{imply} \quad \begin{pmatrix} \tilde{Y}_n \\ \tilde{\sigma}_n \end{pmatrix} \overset{\mathcal{D}}{\Rightarrow} \begin{pmatrix} Y \\ \sigma \end{pmatrix}.$$

  Indeed, the functional convergences of both sequences imply their tightness and thus the tightness of the joint process. This can be seen from the very definition of tightness as given in Billingsley (1968), that can be written for $\tilde{Y}_n$: $\forall \varepsilon,\ \exists K_n$ (compact set) so that $\mathbb{P}(\tilde{Y}_n \in K_n) > 1 - (\varepsilon/2)$ and then, for this ε, we have for $\tilde{\sigma}_n$: $\exists K'_n$ (compact set) so that $\mathbb{P}(\tilde{\sigma}_n \in K'_n) > 1 - \frac{\varepsilon}{2}$. Then:

$$\begin{aligned} \mathbb{P}((\tilde{Y}_n, \tilde{\sigma}_n) \in K_n \times K'_n) &= 1 - \mathbb{P}(\tilde{Y}_n \notin K_n \text{ or } \tilde{\sigma}_n \notin K'_n) \\ &\geq 1 - \mathbb{P}(\tilde{Y}_n \notin K_n) - \mathbb{P}(\tilde{\sigma}_n \notin K'_n) \\ &= \mathbb{P}(\tilde{Y}_n \in K_n) + \mathbb{P}(\tilde{\sigma}_n \in K'_n) - 1 \\ &\geq 1 - \varepsilon. \end{aligned}$$

  Now, the tightness and the pointwise L^2 convergence of the couple imply the convergence of the joint process.
- Let us check the pointwise L^2 convergence of $\tilde{Y}_n$:

$$\begin{aligned} \mathbb{E}(\tilde{Y}_n(t) - Y(t))^2 &= \mathbb{E}\left(\int_0^{[nt]/n}\left(\tilde{\sigma}_n\left(\frac{[ns]}{n}\right) - \sigma(s)\right) dw^1(s) + \int_{\frac{[nt]}{n}}^t \sigma(s)\, dw^1(s)\right)^2 \\ &= \int_0^{[nt]/n} \mathbb{E}\left(\left(\tilde{\sigma}_n\left(\frac{[ns]}{n}\right) - \sigma(s)\right)^2 ds + \int_{[nt]/n}^t \mathbb{E}\left(\sigma^2(s)\right) ds. \end{aligned}$$

  The last right-hand term is less than $\frac{1}{n}\mathbb{E}(\tilde{\sigma}^2)$ and goes to zero when n grows to infinity and the first right-hand term can be written, for the part under the integral, as

$$\begin{aligned} \mathbb{E}\left(\tilde{\sigma}_n\left(\frac{[ns]}{n}\right) - \sigma(s))^2\right) &= \mathbb{E}\left(\exp\left(\int_0^{[nt]/n} a\left(t - \frac{[ns]}{n}\right) dw^2(s)\right)\right. \\ &\qquad \left. - \exp\left(\int_0^t a(t-s)\, dw^2(s)\right)\right)^2. \end{aligned}$$

  Now, for X_n and X following $\mathcal{N}(0, \mathbb{E}X_n^2)$ and $\mathcal{N}(0, \mathbb{E}X^2)$ respectively, $\mathbb{E}(e^{X_n} - e^X)^2 =$

$e^{2\mathbb{E}X_n^2}+e^{2\mathbb{E}X^2}-2e^{\mathbb{E}(X_n+X)^2/2}$, which goes to zero when n grows to infinity if $\mathbb{E}X_n^2 \underset{n\to+\infty}{\longrightarrow} \mathbb{E}X^2$ and $\mathbb{E}(X_n+X)^2 \underset{n\to+\infty}{\longrightarrow} 4\mathbb{E}X^2$.

This can be checked quite straightforwardly here with $X = \int_0^t a(t-s)\,dw^2(s)$ and $X_n = \int_0^{[nt]/n} a(t-\frac{[ns]}{n})\,dw^2(s)$, so that:

$$\mathbb{E}X_n^2 = \int_0^{[nt]/n} a^2\left(t-\frac{[ns]}{n}\right) ds \underset{n\to+\infty}{\longrightarrow} \int_0^t a^2(t-s)\,ds = \mathbb{E}X^2,$$

$$\begin{aligned}\mathbb{E}(X_n+X)^2 \;&=\; \int_0^{[nt]/n} a^2\left(t-\frac{[ns]}{n}\right) ds + \int_0^t a^2(t-s)\,ds \\ &\quad + 2\int_0^{[nt]/n} a\left(t-\frac{[ns]}{n}\right) a(t-s)\,ds \underset{n\to+\infty}{\longrightarrow} 4\mathbb{E}X^2.\end{aligned}$$

This result gives in fact both L^2 convergences of $\tilde{Y}_n(t)$ and of $\tilde{\sigma}_n(t)$.

- The tightness of $\tilde{\sigma}_n$ is then known from Comte (1996) and the tightness of $\tilde{Y}_n$ can be deduced from the proof of Lemma 3.1 with $\mathbb{E}\tilde{\sigma}_n^4(t)$ instead of $\mathbb{E}\tilde{\sigma}^4$, which is still bounded. □

Proof of Proposition 4.1. We work with $\sigma(t) = \exp(\int_{-\infty}^t a(t-s)dw^2(s))$, but the results would obviously still be valid with the only asymptotically stationary version of σ. We use here and for the proof of Proposition (4.2), the following result:

$$\forall\eta \geq 0, \qquad \lim_{h\to+\infty} h^{1-2\alpha}\left(\int_\eta^{+\infty} a(x)a(x+h)\,dx\right) = C, \tag{A.1}$$

where C is a constant. This result can be straighforwardly deduced from Comte and Renault (1996) through rewriting the proof of the result about the long-memory property of the autocovariance function (extended here to the case $\eta \neq 0$).

- We know that $y_t = \mathbb{E}(\sigma^2(t+1) \mid \mathcal{F}_t) = \exp(2\int_0^1 a^2(x)\,dx)\exp(2\int_{-\infty}^t a(t+1-s)\,dw^2(s))$. Then:

$$\begin{aligned} cov(y_{t+h}, y_t) \;&=\; \mathbb{E}(y_{t+h}y_t) - \mathbb{E}(y_{t+h})\mathbb{E}(y_t) \\ &=\; \exp\left(4\int_0^1 a^2(x)\,dx\right) \\ &\quad \times \mathbb{E}\left(\exp\left(2\int_{-\infty}^{t+h} a(t+h+1-s)\,dw^2(s)\right.\right. \\ &\qquad\qquad \left.\left. + 2\int_{-\infty}^{t} a(t+h-s)\,dw^2(s)\right)\right) \\ &\quad - \exp\left(4\int_0^1 a^2(x)\,dx\right) \times \exp\left(4\int_0^{+\infty} a^2(x+1)\,dx\right)\end{aligned}$$

$$= \exp\left(4\int_0^{+\infty} a^2(x)\,dx\right)\left(\exp\left(4\int_1^{+\infty} a(x+h)a(x)\,dx\right)-1\right).$$

This proves the stationarity of the y process, and, with (A.1), which gives the order of the term inside the exponential, implies the announced order $h^{2\alpha-1}$.

- $\mathbb{E}(\sigma(t+h) \mid \mathcal{F}_t)$, still with the stationary version of σ, is given by

$$\exp\left(\int_{-\infty}^{t} a(t+h-s)\,dw^2(s)\right)\times\exp\left(\frac{1}{2}\int_t^{t+h} a^2(t+h-s)\,ds\right).$$

Then, as $(\mathbb{E}(\mathbb{E}(\sigma(t+h \mid \mathcal{F}_t)))^2 = (\mathbb{E}\sigma(t+h))^2$, we have:

$$\begin{aligned}\operatorname{Var}(\mathbb{E}(\sigma(t+h)\mid\mathcal{F}_t)) &= \exp\left(\int_0^{h} a^2(x)\,dx\right)\times\exp\left(2\int_0^{+\infty} a^2(x+h)\,dx\right)\\ &\quad-\exp\left(\int_0^{+\infty} a^2(x)\,dx\right)\\ &= \exp\left(\int_0^{+\infty} a^2(x)\,dx\right)\left(\exp\left(\int_h^{+\infty} a^2(x)\,dx\right)-1\right).\end{aligned}$$

As $a(x) = x^\alpha\tilde{a}(x) = x^{\alpha-1}.x\tilde{a}(x) = O(x^{\alpha-1})$ for $x \to +\infty$ since we know that $\lim_{x\to+\infty} x\tilde{a}(x) = a_\infty$. Then, for $h \to +\infty$,

$$\int_h^{+\infty}(x^{\alpha-1})^2(x\tilde{a}(x))^2\,dx = O\left(a_\infty^2\int_h^{+\infty} x^{2\alpha-2}\,dx\right) = O(h^{2\alpha-1}).$$

Developing again the exponential of this term for great h gives the order $h^{2\alpha-1}$, and even the limit of the variance divided by $h^{2\alpha-1}$ for $h \to +\infty$, which is $K(a_\infty^2/1-2\alpha)$ with $K = \exp(\int_0^{+\infty} a^2(x)\,dx)$.

For $\alpha = 0$, $a(x) = e^{-kx}$ gives obviously for the variance an order Ce^{-kh}. □

Proof of Proposition 4.2. We have to compute $\operatorname{cov}(z_t, z_{t+h})$.

$$\begin{aligned}\mathbb{E}(z_{t+h}z_t) &= \mathbb{E}\left(\int_0^1 \mathbb{E}(\sigma^2(t+h+u)\mid\mathcal{F}_{t+h})\,du\times\int_0^1\mathbb{E}(\sigma^2(t+v)\mid\mathcal{F}_t)\,dv\right)\\ &= \int_0^1\int_0^1\left[\mathbb{E}\left(e^{(2\int_0^u a^2(x)\,dx)}e^{(2\int_{-\infty}^{t+h} a(t+h+u-s)\,dw^2(s))}e^{(2\int_0^v a^2(x)\,dx)}e^{(2\int_{-\infty}^{t} a(t+v-s)\,dw^2(s))}\right)\right]du\,dv\\ &= \int_0^1\int_0^1\left[e^{(2\int_0^u a^2(x)\,dx+2\int_0^v a^2(x)\,dx)}e^{(2\int_{-\infty}^{t}(a(t+h+u-s)+a(t+v-s))^2\,ds)}e^{(2\int_t^{t+h} a^2(t+h+u-s)\,ds)}\right]du\,dv\\ &= \int_0^1\int_0^1 e^{(4\int_0^{+\infty} a^2(x)\,dx+4\int_0^{+\infty} a(x+h+u)a(x)\,dx)}\,du\,dv.\end{aligned}$$

Moreover, with the same kind of computations we have $\mathbb{E}(z_{t+h})\mathbb{E}(z_t) = \exp(4\int_0^{+\infty} a^2(x)\,dx)$ so that:

$$
\begin{aligned}
cov(z_t, z_{t+h}) &= \exp\left(4\int_0^{+\infty} a^2(x)\,dx\right) \\
&\quad \times \left[\int_0^1\int_0^1 \left(\exp\left(4\int_v^{+\infty} a(x+h+u-v)a(x)\,dx\right) - 1\right) du\,dv\right].
\end{aligned}
$$

Then z is stationary and another use of (A.1) gives the order $h^{2\alpha-1}$ for $h \to +\infty$. □

Proof of Proposition 4.3. Let $Z(t) = \int_0^t \sigma(u)dw^1(u)$. Then we know (see Protter 1992 p. 174, for $p = 4$) that $\mathbb{E}Z(t)^4 = \frac{4(4-1)}{2}\mathbb{E}\int_0^t Z^2(s)d\langle Z\rangle_s$. As $\langle Z\rangle_s = \int_0^s \sigma^2(u)\,du$, we find: $\mathbb{E}Z(t)^4 = 6\mathbb{E}(\int_0^t Z^2(s)\sigma^2(s)\,ds) = 6\int_0^t \mathbb{E}(Z^2(s)\sigma^2(s))\,ds$, with Fubini's theorem. Then $\mathbb{E}(Z^2(t)\sigma^2(t)) = \mathbb{E}(\int_0^t(\sigma(t)\sigma(u))\,dw^1(u))^2 = \mathbb{E}\int_0^t(\sigma^2(t)\sigma^2(u))\,du$ (σ and w^1 are independent). This yields

$$
\mathbb{E}Z(t)^4 = 6\int_0^t\int_0^s (r_{\sigma^2}(|s-u|) + (\mathbb{E}\sigma^2)^2)\,du\,ds,
$$

where r is the autocovariance function, and, lastly, $\varphi(h) = 3h^2(\mathbb{E}\sigma^2)^2 + 3\iint_{[0,h]^2} r_{\sigma^2}(|s-u|)\,du\,ds$.

- Near zero, the autocovariance function of σ^2 is of the same kind as the one of σ, with a replaced by $2a$, since $\sigma = \exp(x)$. Then we know from Proposition 2.1 that, for $h \to 0$: $r_{\sigma^2}(h) = r_{\sigma^2}(0) + Ch^{2\alpha+1} + o(h^{2\alpha+1})$, where C is a constant and $\alpha \in]0, \frac{1}{2}[$. Then replacing in $\varphi(h)$ gives

$$
\varphi(h) = 3h^2((\mathbb{E}\sigma^2)^2 + r_{\sigma^2}(0)) + \frac{3C}{(2\alpha+2)(2\alpha+3)}h^{2\alpha+3} + o(h^{2\alpha+3}).
$$

  For $\alpha = 0$, $a(x) = \exp(-kx)$ gives $r_{\sigma^2}(h) = e^{2/k}\left(\exp(\frac{2}{k}e^{kh}) - 1\right)$ which leads to: $r_{\sigma^2}(h) = r_{\sigma^2}(0) - 2e^{4/k}h + o(h)$, for $h \to 0$. This implies the continuity for $\alpha = 0$. Now, $\mathbb{E}(Y(h) - \mathbb{E}Y(h))^2 = \mathbb{E}(\int_0^h \sigma(u)\,dw^1(u))^2 = \mathbb{E}\int_0^h \sigma^2(u)\,du = h\mathbb{E}\sigma^2$ implies that

$$
\lim_{h\to 0} kurt_Y(h) = 3\frac{\mathbb{E}\sigma^4}{(\mathbb{E}\sigma^2)^2} > 3.
$$

- From Proposition 2.2, we know that, for $\alpha \in]0, \frac{1}{2}[$ and $h \to +\infty$,

$$
\iint_{[0,h]^2} r_{\sigma^2}(|s-u|)\,du\,ds = O\left(\iint_{[1,h]^2} u^{2\alpha-1}\,du\,ds\right) = O(h^{2\alpha+1}),
$$

  but for $\alpha = 0$, $r_{\sigma^2}(h) = e^{2/k}\left(\frac{2}{k}e^{-kh} + o(e^{-kh})\right)$. This gives the result and the exponential rate for $\alpha = 0$. □

Proof of Proposition 5.1. Let $m = [Nt/T]$, then

$$\begin{aligned}\mathbb{E}\hat{\sigma}^2_{n,p}(t) &= \frac{n}{T}\sum_{k=m-p+1}^{m}\mathbb{E}\left(\int_{t_{k-1}}^{t_k}\sigma(s)\,dw^1(s)\right)^2 = \frac{n}{T}\sum_{k=m-p+1}^{m}\mathbb{E}\left(\int_{t_{k-1}}^{t_k}\sigma^2(s)\,ds\right)\\ &= \frac{n}{T}\mathbb{E}\left(\int_{t_{m-p}}^{t_m}\sigma^2(s)\,ds\right) = \frac{n}{T}\mathbb{E}(\sigma^2)(t_m - t_{m-p}) = \frac{n}{T}p\times\frac{T}{np}\mathbb{E}\sigma^2, = \mathbb{E}\sigma^2,\end{aligned}$$

where $\mathbb{E}\sigma^2 = \exp(\frac{1}{2}\int_0^{+\infty} a^2(x)\,dx)$. This ensures the L^1 convergence of the sequence, uniformly in t.

Before computing the mean square, let:

$$\begin{aligned} f(z) &= f(|z|) = \mathbb{E}\sigma^2(u)\sigma^2(u+|z|)\\ &= \exp\left(4\int_0^{+\infty} a^2(x)dx + 4\int_0^{+\infty} a(x)a(x+|z|)\,dx\right).\end{aligned}$$

Then

$$\begin{aligned}\mathbb{E}[\hat{\sigma}^2_{n,p}(t) - \sigma^2(t)]^2 &= \mathbb{E}\left[\frac{n}{T}\sum_{k=m-p+1}^{m}(Y_{t_k} - Y_{t_{k-1}})^2 - \sigma^2(t)\right]^2\\ &= \frac{n^2}{T^2}\mathbb{E}\left[\sum_{k=m-p+1}^{m}(Y_{t_k} - Y_{t_{k-1}})^2\right]^2\\ &\quad + \mathbb{E}\sigma^4(t) - 2\frac{n}{T}\mathbb{E}\left[\sum_{k=m-p+1}^{m}\sigma^2(t)(Y_{t_k} - Y_{t_{k-1}})^2\right]^2.\end{aligned}$$

We consider separately the different terms.

$$\begin{aligned}\mathbb{E}\left[\sigma^2(t)(Y_{t_k} - Y_{t_{k-1}})^2\right]^2 &= \mathbb{E}\left[\int_{t_{k-1}}^{t_k}\sigma(t)\sigma(s)\,dw^1(s)\right]^2 = \mathbb{E}\left[\int_{t_{k-1}}^{t_k}\sigma^2(t)\sigma^2(s)\,ds\right]\\ &= \int_{t_{k-1}}^{t_k} f(t-s)\,ds\end{aligned}$$

as σ and w^1 are independent. As in a previous proof, we have:

$$\mathbb{E}\left[\int_{t_{k-1}}^{t_k}\sigma(s)\,dw^1(s)\right]^4 = 3\iint_{[t_{k-1},t_k]^2} f(u-v)\,du\,dv,$$

and for $j \neq k$: $\mathbb{E}[(Y_{t_k} - Y_{t_{k-1}})^2(Y_{t_j} - Y_{t_{j-1}})^2] = \mathbb{E}(\int_{t_{k-1}}^{t_k}\sigma^2(s)\,ds \times \int_{t_{j-1}}^{t_j}\sigma^2(s)\,ds)$. This gives:

$$\mathbb{E}\left[\sum_{k=m-p+1}^{m}(Y_{t_k}-Y_{t_{k-1}})^2\right]^2 = 2\sum_{k=m-p+1}^{m}\iint_{[t_{k-1},t_k]^2} f(u-v)\,du\,dv + \iint_{[t_{m-p},t_m]^2} f(u-v)\,du\,dv.$$

Now with all the terms:

$$\begin{aligned}\mathbb{E}[\hat{\sigma}^2_{n,p}(t)-\sigma^2(t)] &= \frac{n^2}{T^2}\left(2\sum_{k=m-p+1}^{m}\iint_{[t_{k-1},t_k]^2} f(u-v)\,du\,dv \right.\\ &\qquad \left. + \iint_{[t_{m-p},t_m]^2} f(u-v)\,du\,dv\right)\\ &\quad + \mathbb{E}\sigma^4(t) - \frac{2n}{T}\int_{[t_{m-p},t_m]} f(t-s)\,ds\\ &= \frac{2n^2}{T^2}\sum_{k=m-p+1}^{m}\iint_{[t_{k-1},t_k]^2}(f(u-v)-f(0))\,du\,dv\\ &\quad + \frac{n^2}{T^2}\iint_{[t_{m-p},t_m]^2}(f(u-v)-f(0))\,du\,dv\\ &\quad - \frac{2n}{T}\int_{[t_{m-p},t_m]}(f(t-s)-f(0))\,ds + 2\frac{\mathbb{E}\sigma^4}{p}.\end{aligned}$$

Let $K_1 = \exp(8\int_0^{+\infty} a^2(s)\,ds)$. Then $f(h) = K_1 r_\sigma(h)$ where r_σ is as in Proposition 2.1. Proposition 2.1 then implies $|\ f(h) - f(0)\ | \leq K_1 C\ |\ h\ |^{2\alpha+1}$, where C is a positive constant.

This implies that: $\forall\varepsilon > 0, \exists\eta > 0,$ so that $|v-u| < \eta \Rightarrow |f(v-u) - f(0)| < \varepsilon$. Let then $\varepsilon = 1/p$, then $\eta = \eta(\varepsilon)$ is fixed and

$$\mathbb{E}[\hat{\sigma}^2_{n,p}(t)-\sigma^2(t)]^2 \leq \frac{2n^2}{T^2}\times p\times\left(\frac{T}{np}\right)^2\times\frac{1}{p}+\frac{n^2}{T^2}\times\left(\frac{T}{n}\right)^2\times\frac{1}{p}+\frac{2n}{T}\times p\frac{T}{np}\times\frac{1}{p}+2\frac{\mathbb{E}\sigma^2}{p}$$

if $|t_m - t_{m-p}| = \frac{T}{n} < \eta$, which implies $|t_k - t_{k-1}| = \frac{T}{np} < \eta$. Then

$$\mathbb{E}[\hat{\sigma}^2_{n,p}(t)-\sigma^2(t)]^2 \leq \left(\frac{2}{p}+1+2+2\mathbb{E}\sigma^2\right)\times\frac{1}{p} = \left(\frac{2}{p}+3+2\mathbb{E}\sigma^2\right)\times\frac{1}{p}.$$

Then $\forall a > 0,\ n > T/\eta \Rightarrow p^{1-a}.\mathbb{E}\left[\hat{\sigma}^2_{n,p}(t)-\sigma^2(t)\right]^2 \leq \frac{C}{pa}$, where $C = 5 + 2\mathbb{E}\sigma^2$.

The stationarity implies that the result is uniform in t, so that $\lim\limits_{n,p\longrightarrow+\infty}\sup\limits_{t\in[0,T]} p^{1-a}\mathbb{E}[\hat{\sigma}^2_{n,p}(t)$ $-\sigma^2(t)]^2 = 0.$ □

APPENDIX B

We suppose here $r = 0$ and $\lambda_2 = 0$, so that $x^{(\alpha)}(t) = (\ln\sigma)^{(\alpha)}(t)$ can be written:

$$x^{(\alpha)}(t) = e^{-kt}\left(x^{(\alpha)}(0) + \int_0^t e^{ks}\gamma\, d\tilde{w}^2(s)\right).$$

Then $U_{t,T}^2 = \int_t^T \sigma^2(u)du$ can be written:

$$U_{t,T}^2 = \int_t^T \exp\left(2\int_0^t \frac{(u-s)^\alpha}{\Gamma(1+\alpha)}\, dx^{(\alpha)}(s)\right) \times \exp\left(2\int_t^u \frac{(u-s)^\alpha}{\Gamma(1+\alpha)}\, dx^{(\alpha)}(s)\right) du.$$

Then the first part, $\exp(2\int_0^t \frac{(u-s)^\alpha}{\Gamma(1+\alpha)}\, dx^{(\alpha)}(s))$, is “deterministic” knowing $\mathcal{F}_t$. For the second part, since we have for $s > t$ that $x^{(\alpha)}(s) = e^{-k(s-t)}x^\alpha(t) + \int_t^s e^{-k(s-x)}\gamma\, d\tilde{w}^2(x)$, we find that

$$x^{(\alpha)}(s) = e^{-k(t-s)}x^{(\alpha)}(t) + \int_s^t e^{k(s-x)}\gamma\, d\tilde{w}^2(x),$$

$$\begin{aligned}\int_t^u \frac{(u-s)^\alpha}{\Gamma(1+\alpha)}\, dx^{(\alpha)}(s) \;=\; & x^{(\alpha)}(t) \times \left(-k\int_t^u \frac{(u-s)^\alpha}{\Gamma(1+\alpha)} e^{-k(s-t)}\, ds\right)\\ & - k\int_t^u \frac{(u-s)^\alpha}{\Gamma(1+\alpha)} \int_t^s e^{k(s-x)}\gamma\, d\tilde{w}_2(x)\\ & +\gamma\int_t^u \frac{(u-s)^\alpha}{\Gamma(1+\alpha)}\, d\tilde{w}^2(s).\end{aligned}$$

This term depends then only on $(x^{(\alpha)}(t))$ and on **future** increments of the Brownian motion $\tilde{w}^2$; those increments are independent of $\mathcal{F}_t$. This is the reason we can write

$$\exp\left(2\int_t^u \frac{(u-s)^\alpha}{\Gamma(1+\alpha)}\, dx^{(\alpha)}(s)\right) = f(x^{(\alpha)}(t); Z(u,t,\alpha)).$$

At time $t = t_i$, this gives the announced formula, with $\ln[f(x^{(\alpha)}(t); Z(u,t,\alpha))] = x^{(\alpha)}(t)\varphi(t,u) + Z(u,t,\alpha)$; $\varphi(t,u)$ is a deterministic function, $Z(u,t,\alpha)$ is a process independent of $\mathcal{F}_t$:

$$Z(t,u,\alpha) = -k\int_t^u \frac{(u-s)^\alpha}{\Gamma(1+\alpha)} \int_t^s e^{k(s-x)}\gamma\, d\tilde{w}^2(x) + \gamma\int_t^u \frac{(u-s)^\alpha}{\Gamma(1+\alpha)}\, d\tilde{w}^2(s).$$

REFERENCES

BACKUS, D. K. and S. E. ZIN (1993): Long-Memory Inflation Uncertainty: Evidence from the Term Structure of Interest Rates *J. Money, Credit, Banking* 3, 681–700.

BAILLIE, R. T., T. BOLLERSLEV, and H. O. MIKKELSEN (1996): Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *J. Econometrics* 74, 1, 3–30.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

BLACK, F. and M. SCHOLES (1973): The Pricing of Options and Corporate Liabilities, *J. Political Econ.* 3, 637–654.

BREIDT, F. J., N. CRATO, and P. DE LIMA (1993): Modeling Long-Memory Stochastic Volatility, discussion paper, Iowa State University.

COMTE, F. (1996): Simulation and Estimation of Long Memory Continuous Time Models, *J. Time Series Anal.* 17, 1, 19–36.

COMTE, F. and E. RENAULT (1996): Long Memory Continuous Time Models, *J. Econometrics* 73, 101–149.

DALHUAS, R. (1989): Efficient Parameter Estimation for Self-Similar Processes, *Ann. Statistics* 17, 4, 1749–1766.

DELBAEN, F. and W. SCHACHERMAYER (1994): A General Version of the Fundamental Theorem of Asset Pricing, *Math. Ann.* 300, 3, 463–520.

DING, Z., C., W. J. GRANGER, and R. F. ENGLE (1993): A Long Memory Property of Stock Market Returns and a New Model, *J. Empirical Finance* 1, 1.

DROST, F. C. and B. J. M. WERKER (1996): Closing the GARCH Gap: Continuous Time GARCH Modeling, *J. Econometrics* 74, 31–58.

ENGLE, R. F. and C. MUSTAFA (1992): Implied ARCH Models from Options Prices, *J. Econometrics* 52, 289–331.

FOX, R. and M. S. TAQQU (1986): Large Sample Properties of Parameter Estimates for Strongly Dependent Time Series, *Ann. Statistics* 14, 517–532.

GEWEKE, J. and S. PORTER-HUDAK (1983): The Estimation and Application of Long Memory Time Series Models, *J. Time Series Analysis* 4, 221–238.

GHYSELS, E., A. HARVEY, and E. RENAULT (1996): Stochastic Volatility; In *Handbook of Statistics*, Vol. 14, *Statistical Methods in Finance*, G. S. Maddala, ed. Amsterdam: North Holland, Ch. 5, 119–191.

GOURIEROUX, C., A. MONFORT, and E. RENAULT (1993): Indirect Inference, *J. Appl. Econometrics* 8, S85–S118.

HARRISON, J. M. and D. M. KREPS (1981): Martingale and Arbitrage in Multiperiods Securities Markets, *J. Econ. Theory* 20, 381–408.

HARVEY, A. C. (1993): Long Memory in Stochastic Volatility, London School of Economics, working paper.

HARVEY, A., E. RUIZ, and N. SHEPHARD (1994): Multivariate Stochastic Variance Models, *Rev. Econ. Stud.* 61, 2, 247–264.

HEYNEN, R., A. KEMNA, and T. VORST (1994): Analysis of the Term Structure of Implied Volatilities, *J. Financial Quant. Anal.* 29, 31–56.

HULL, J. and A. WHITE (1987): The Pricing of Options on Assets with Stochastic Volatilities, *J. Finance* 3, 281–300.

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.

KARATZAS, I. and S. E. SHREVE (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. New York: Springer-Verlag.

MANDELBROT, B. B. and VAN NESS (1968): Fractional Brownian Motions, Fractional Noises and Applications, *SIAM Rev.* 10, 422–437.

MELINO, A. and TURNBULL, S. M. (1990): Pricing Foreign Currency Options with Stochastic Volatility, *J. Econometrics* 45, 239–265.

MERTON, R. (1973): The Theory of Rational Option Pricing, *Bell J. Econ. Mgmt. Sci.* 4, 141–183.

NELSON, D. B. and D. P. FOSTER (1994): Asymptotic Filtering Theory for ARCH Models, *Econometrica* 1, 1–41.

PASTORELLO, S., E. RENAULT, and N. TOUZI (1993): Statistical Inference for Random Variance Option Pricing, *Southern European Economics Discussion Series*.

PHAM, H. and N. TOUZI (1996): Intertemporal Equilibrium Risk Premia in a Stochastic Volatility Model, *Math. Finance* 6, 215–236.

PROTTER, P. (1992): *Stochastic Integration and Differential Equations*. New York: Springer-Verlag.

RENAULT, E. and N. TOUZI (1996): Option Hedging and Implicit Volatilities in a Stochastic Volatility Model, *Math. Finance* 6, 279–302.

ROBINSON, P. M. (1993): Efficient Tests for Nonstationary Hypotheses, Working paper, London School of Economics.

ROBINSON, P. M. (1996): Semiparametric Analysis of Long-Memory Time Series, *Ann. Statistics* 22, 515–539.

ROGERS, L. C. G. (1995): Arbitrage with Fractional Brownian Motion, University of Bath, discussion paper.

ROZANOV, YU. A. (1968): *Stationary Random Processes*. San Francisco: Holden-Day.

SCOTT, L. (1987): Option Pricing when the Variance Changes Randomly: Estimation and an Application, *J. Financial and Quant. Anal.* 22, 419–438.