From Market Data to Agent-Based Models and Stochastic Differential Equations

R. Vilela Mendes

Abstract A survey of results from 2008 to 2014 on the construction of a stochastic market model, from the empirical data to its modelling interpretation and proof of mathematical consistency (no-arbitrage and completeness).

Keywords Fractional volatility · Arbitrage · Completeness · Agent-based models

1 Introduction

After 2008, excessive reliance on mathematical models was blamed as the cause for the subprime collapse and the following global crisis. This criticism soon reached the popular level on newspaper articles and cartoons

Un krach des maths
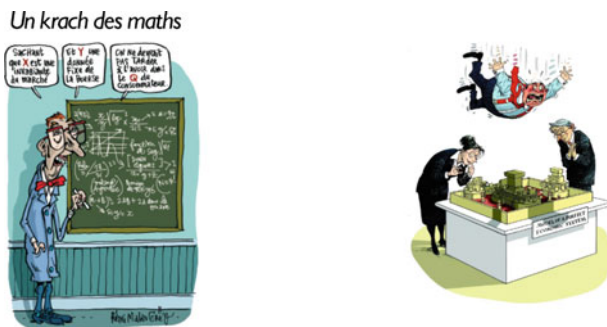
It is true that a closed mathematical model of the economy may be an impossible dream, as is also true that some models use unrealistic assumptions, trading empirical evidence for mathematical beauty or simplicity. On the other hand, there are also cases where the mathematics is there, providing sound results, but nobody pays attention,

R.V. Mendes (✉)
Centro de Matemática e Aplicações Fundamentais, University of Lisbon,
Lisbon, Portugal
e-mail: rvmendes@fc.ul.pt; rvilela.mendes@gmail.com

because sometimes not paying attention is more profitable in the short term. An example:

A theorem by Föllmer and Schied [1] states that if M is the set of all probability measures on a finite space of scenarios Ω, ρ will be a convex risk measure iff there is a penalty function $\alpha : M \to (-\infty, \infty]$ such that

$$\rho(X) = \sup_{Q \in P}(E_Q[-X] - \alpha(Q)) \tag{1}$$

with α convex, lower semicontinuous and $\alpha(Q) \geq -\rho(0)$.

In such a convex risk measure, the first term represents the **maximal expected loss in the scenario** Q and $\alpha(Q)$ accounts for the **probability of the scenario**. Whereas the calculation of $E_Q[-X]$ is a simple exercise in stochastic analysis, estimation of $\alpha(Q)$ involves many factors which are frequently not taken into account. *For example, if the historical data that is being used does not contain unfavorable events, it is tempting (or profitable) to say that meltdowns are improbable*.

Intrigued by why the well-qualified experts of the rating agencies had rated AAA the "toxic" products and had not predicted the 2008 crisis, 4 economists of the Federal Reserve of Atlanta, made an extensive analysis of their reports in the years before the crisis. The conclusion was that most experts reported that a small fall in the price of the houses would lead to disaster, but assigned a very small (*penalty*) probability to that event [2]. In the Atlanta paper, instead of the language of convex measures, the authors consider the probability of foreclosures decomposed into

$$\frac{df}{dt} = \frac{df}{dp}\frac{dp}{dt} \tag{2}$$

$\frac{df}{dp}$ being the sensitivity of foreclosures to price (HPA) and $\frac{dp}{dt}$ the time variation of the price of houses. Their conclusion is that the estimation of $\frac{df}{dp}$ was correct but not of $\frac{dp}{dt}$ (which is equivalent to the effect of the penalty term). However looking at the housing bubble it should have been clear after 2001 that the probability of a downturn in HPA was high. In addition, how could inflation-adjusted prices continue to rise when real incomes of most Americans, especially at the bottom, continued to fall?

Why was a small value assigned to the *penalty term?* Conflict of interests is a possible reason. The SEC recognition of the main agencies (Fitch, Moody's, S&P and Dominion) together with the recommendations of Basel II, put them in the center of the financial world. However their clients were exactly the creators of the securities. If one agency does not provide favorable reports, look elsewhere. A recent event lends credibility to this hypothesis: In February 2011, Redwood Trust put in the market the second (since the crisis) mortgage-based private-label security in the USA. Redwood asked both Fitch and Moody's for ratings, but only published Fitch's report. Later on, perhaps to assert its credibility, Moody's published its report which was quite negative. On the other hand if the client were the buyer of the securities, a

conflict of interest of opposite sign might also occur. A recent proposal is the creation of government-sponsored clearing agencies. Will it ever work?

Another example is the Bernard Madoff affair. Already in 1999 and later in 2005 Markopoulos' letters to the SEC had shown, using simple mathematics that Madoff's strategy could not generate the 12 % average annual return unless he was either using **insider-trading** or a **Ponzi scheme.** Although some large brokerage institutions (Goldman Sachs, for example) stayed away from any deals with Madoff, he continued to attract a lot of investment in Europe and the USA. Once again, the mathematics was there but very few were paying attention.

The work that is reported in this paper concerns another aspect of the relation of mathematics to economic life, namely the modelization of the price $S(t)$ fluctuations in stock exchanges. The basic (stylized) facts provided by the empirical data are:

1. The **returns** $(r(t, \Delta) = \frac{S(t+\Delta)-S(t)}{S(t)})$ have nearly no autocorrelation;
2. The **autocorrelations of** $|r(t, \Delta)|$ decline slowly with increasing lag Δ, a long memory effect;
3. **Leptokurtosis**: asset returns have distributions with fat tails and excess peakedness at the mean;
4. **Autocorrelations of sign** $r(t, \Delta)$ are insignificant;
5. **Volatility clustering**: there is a tendency of large changes to follow large changes and small changes to follow small changes. Volatility occurs in bursts;
6. **Volatility is mean-reversing** and the distribution is close to lognormal or inverse gamma;
7. **Leverage effect**: volatility tends to rise more following a large price fall than following a price rise.

Geometrical Brownian motion (GBM)

$$\frac{dS_t}{S_t} = \mu\,dt + \sigma\,dB(t) \tag{3}$$

is a basis for most of mathematical finance (Black-Scholes, etc.). Is it consistent with the empirical data? No. In GMB price changes would be log-normal. No leptokurtosis and scaling properties $E\left|\frac{S(t+\Delta)-S(t)}{S(t)}\right| \approx \Delta^{1/2}$ which is not born out by the data. In addition, the volatility σ being constant, there is no volatility clustering nor leverage effect. One of the most famous consequences of GBM is the Black-Scholes formula for option pricing. When the historical volatility σ is used in the formula, the resulting price is quite distinct from the one that is actually practiced in the market. Nevertheless the Black-Scholes continues to be used in the following way: The market price of liquid options is used to infer what would be the (implied) volatility σ_{imp} that would lead to that price. σ_{imp} is then used to compute the price of less traded options. Black-Scholes is used as an invertible mapping between price and σ_{imp} and not for its theoretical value.

In conclusion: the wide use of GBM is a case where oversimplification of the mathematics leads to important departures from the empirical evidence. This was the motivation that led to the present attempt to construct a market model, which although preserving some of the nice features of GBM, would be consistent with (and directly inspired by) the empirical data.

2   A Data-Reconstructed Market Model [3]

The basic hypothesis for the model construction were:

- (**H1**) *The log-price process* $\log S_t$ *belongs to a probability space* $\Omega \otimes \Omega'$, where the first one, Ω, is the Wiener space W and the second, Ω', is a probability space to be empirically reconstructed.
  Denote $\log S_t(\omega, \omega')$ with $\omega \in \Omega$, $\omega' \in \Omega'$, and $\mathscr{F}_t$, $\mathscr{F}'_t$ are the $\sigma-$algebras in Ω and Ω' generated by the processes up to time t.
- (**H2**) The second hypothesis is stronger: *Assume that for each fixed* ω', $\log S_t(\bullet, \omega')$ *is a square integrable random variable in* Ω

From (H2) it follows [4] that, for each fixed ω',

$$\tfrac{dS_t}{S_t}(\bullet, \omega') = \mu_t(\bullet, \omega')dt + \sigma_t(\bullet, \omega')dB(t) \tag{4}$$

with $\mu_t(\bullet, \omega')$ and $\sigma_t(\bullet, \omega')$ well-defined processes in Ω.
If $\{X_t, \mathscr{F}_t\}$ is a process such that

$$dX_t = \mu_t dt + \sigma_t dB(t) \tag{5}$$

with μ_t and σ_t being $\mathscr{F}_t-$adapted processes, then

$$\begin{aligned} \mu_t &= \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \left\{ E(X_{t+\varepsilon} - X_t) | \mathscr{F}_t \right\} \\ \sigma_t^2 &= \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \left\{ E(X_{t+\varepsilon} - X_t)^2 \big| \mathscr{F}_t \right\} \end{aligned} \tag{6}$$

The process associated to the probability space Ω' is now inferred from the data. For each fixed ω' realization in Ω' one has

$$\sigma_t^2(\bullet, \omega') = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left\{ E(\log S_{t+\varepsilon} - \log S_t)^2 \right\} \tag{7}$$

Because each set of market data corresponds to a particular realization ω', the σ_t^2 process may indeed be reconstructed from the data. The question is how to construct a mathematical model for this *induced volatility* process. For this purpose we looked for scaling properties of the data, namely
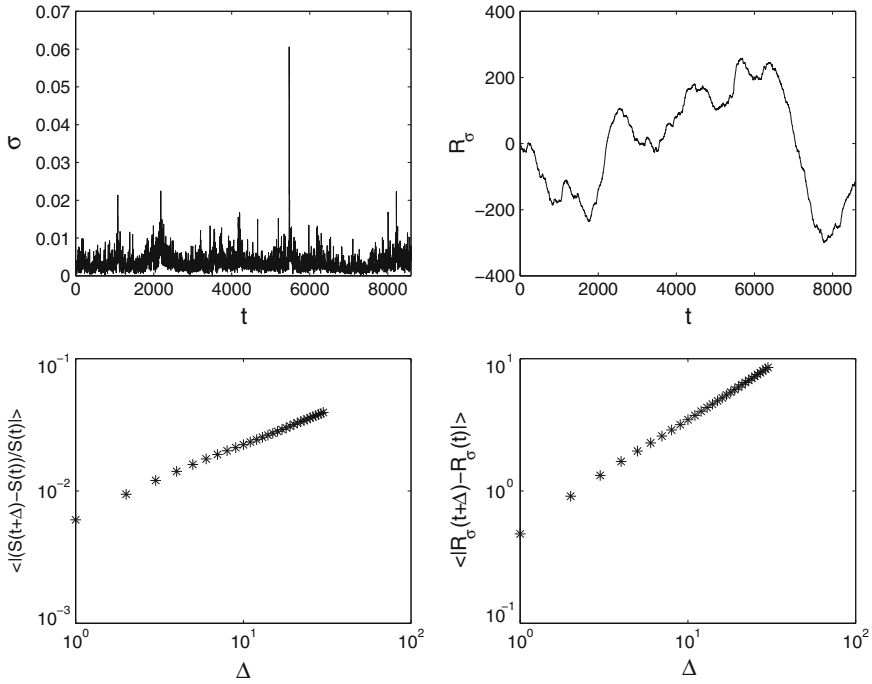
Fig. 1 The R_σ process and its scaling properties

$$E\left|\sigma(t+\Delta) - \sigma(t)\right| \sim \Delta^H \qquad E\left|\frac{\sigma(t+\Delta) - \sigma(t)}{\sigma(t)}\right| \sim \Delta^H \qquad (8)$$

but neither of these hold. By contrast, the empirical integrated log-volatility is well represented by a relation of the form $\sum_{n=0}^{t/\delta} \log \sigma(n\delta) = \beta t + R_\sigma(t)$ with the $R_\sigma(t)$ process displaying very accurate self-similar properties (Fig. 1).

If a nondegenerate process X_t has finite variance, stationary increments and is self-similar

$$Law(X_{at}) = Law(a^H X_t)$$

then it necessarily has covariance

$$Cov(X_s, X_t) = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |s - t|^{2H})E(X_1^2) \qquad (9)$$

and the simplest process with these properties is a Gaussian process called *fractional Brownian motion*. Therefore the simplest model compatible with the data is

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma_t S_t dB(t) \\ \log \sigma_t &= \beta + \tfrac{k}{\delta}\{B_H(t) - B_H(t - \delta)\} \end{aligned} \qquad (10)$$

which has been called the *fractional volatility model* (FVM). δ is the observation time scale and, from the data, H is found to be in the range 0.8–0.9. From (10) it follows that the volatility (at resolution δ) is

$$\sigma(t) = \theta e^{\frac{k}{\delta}\{B_H(t) - B_H(t-\delta)\} - \frac{1}{2}(\frac{k}{\delta})^2 \delta^{2H}} \tag{11}$$

3 Mathematical Consistency of the Fractional Volatility Model: Arbitrage and Completeness [5]

The main consistency check for any market model is the no-arbitrage principle. In addition completeness or incompleteness of the model should also be checked.

The no-arbitrage principle: *A (perfect) market does not allow for risk-free profits with no initial investment or, equivalently, to profits without any risk.*

A self-financing portfolio $V(t) = \sum_{i=1}^{n} h_t^i S_t^i$ is an *arbitrage portfolio* if

$$V(0) = 0; \; P(V(T) > 0) > 0 \tag{12}$$

for $T > 0$, P being the probability measure on the market scenarios. *A market is arbitrage free if and only if there is an equivalent martingale measure Q for the discounted price processes* [1, 6].

On the other hand completeness is related to the possibility of hedging portfolios. H is said to be *an hedge* for the portfolio X (or to *replicate X*) if

$$H \text{ is self-financing}; \; V_H(T) = X(T), \; P - \text{almost surely} \tag{13}$$

A market is *complete* if all X can be hedged. Furthermore, *a market is complete if and only if the martingale measure Q is unique*.

3.1 No-Arbitrage

In the fractional volatility model (FVM) one has two probability spaces $(\Omega_1, \mathscr{F}_1, P_1)_W$ and $(\Omega_2, \mathscr{F}_2, P_2)_{B_H}$ and the product space $(\overline{\Omega}, \overline{\mathscr{F}}, \overline{P})$ with π_1 and π_2 the projections of $\overline{\Omega}$ onto Ω_1 and Ω_2.

Consider a risky asset with price S_t and a risk-free asset with dynamics

$$dA_t = rA_t \, dt \quad A_0 = 1 \tag{14}$$

The volatility σ_t of the risky asset is a measurable $\overline{\mathbb{F}}$–adapted process satisfying for $0 \leq t < \infty$

$$\mathbb{E}_{\overline{P}}\left[\int_0^t \sigma_s^2 ds\right] = \int_0^t \theta^2 e^{-(\frac{k}{\delta})^2 \delta^{2H}} \mathbb{E}_{\overline{P}}\left[e^{\frac{2k}{\delta}\{B_H(s) - B_H(s-\delta)\}}\right] ds$$

$$= \theta^2 \exp\left\{(\frac{k}{\delta})^2 \delta^{2H}\right\} t < \infty \tag{15}$$

by Fubini's theorem and the moment generating function of the Gaussian random variable $B_H(s) - B_H(s - \delta)$.

$\int_0^t |\mu_s|\, ds$ being finite $\overline{P}$–almost surely for $0 \leq t < \infty$, application of Itô's formula yields

$$S_t = S_0 \exp\left\{\int_0^t (\mu_s - \frac{1}{2}\sigma_s^2) ds + \int_0^t \sigma_s dB_s\right\} \tag{16}$$

Lemma 1 *Consider the measurable process*

$$\gamma_t = \frac{r - \mu_t}{\sigma_t}, \qquad 0 \leq t < \infty \tag{17}$$

with $\mu \in L^\infty([0, T] \times \overline{\Omega})$. Then, for a continuous version of B_H

$$\exp\left[\frac{1}{2}\int_0^T \gamma_s^2(\omega_2) ds\right] < A(\omega_2) < \infty \tag{18}$$

P_2–almost all $\omega_2 \in \Omega_2$.

The proof uses the fact that P_2–almost surely, a continuous version of fractional Brownian motion is Hölder continuous of any order $\alpha \geq 0$ less than H, that is, there is a random variable $C_\alpha > 0$ such that for P_2–almost all $\omega_2 \in \Omega_2$ $|B_H(t) - B_H(s)| \leq C_\alpha(\omega_2) |t - s|^\alpha$

Theorem 1 *The market (A_t, S_t, σ_t) is free of arbitrage*

Proof Restricting the process to a particular path ω_2 of the B_H–process, construct the stochastic exponential of $\int_0^t \gamma_s(\omega_2) dB_s$,

$$\eta_t(\omega_2) = \exp\left\{\int_0^t \gamma_s(\omega_2) dB_s - \frac{1}{2}\int_0^t \gamma_s^2(\omega_2) ds\right\} \tag{19}$$

The bound in Lemma 1 is the Kallianpur condition that insures

$$\mathbb{E}_{P_1}[\eta_t(\omega_2)] = 1 \qquad \omega_2 - a.s. \tag{20}$$

Hence, we are in the framework of Girsanov theorem and $\eta_t(\omega_2)$ is a true P_1-martingale. We can define for each $0 \leq T < \infty$ a new probability measure $Q_T(\omega_2)$ on $\mathscr{F}_1$ by

$$\frac{dQ_T(\omega_2)}{dP_1} = \eta_T(\omega_2), \qquad P_1 - a.s. \tag{21}$$

By the Cameron-Martin-Girsanov theorem, for each $T \in [0, \infty)$, the process

$$B_t^* = B_t - \int_0^t \frac{r - \mu_s}{\sigma_s(\omega_2)} \, ds \qquad 0 \leq t \leq T \tag{22}$$

is a Brownian motion on the probability space $(\Omega, \mathscr{F}_1, Q_T(\omega_2))$.

Under the new probability measure $Q_T(\omega_2)$ (equivalent to P_1 on $\mathscr{F}_1$) the discounted price process, $Z_t = \frac{S_t}{A_t}$ $0 \leq t \leq T$ with dynamical law

$$Z_t(\omega_2) = Z_0 + \int_0^t \sigma_s(\omega_2) Z_s(\omega_2) \, dB_s^* \tag{23}$$

is a martingale in the probability space $(\Omega_1, \mathscr{F}_1, Q_T(\omega_2))$. *By the fundamental theorem of asset pricing* [1, 6]*, the existence of an equivalent martingale measure for Z_t implies that there are no arbitrages*, that is, $\mathbb{E}_{Q_T(\omega_2)} \left[Z_t(\omega_2) | \mathscr{F}_{1,s} \right] = Z_s(\omega_2)$ for $0 \leq s < t \leq T$.

This proves that there are no arbitrages for P_2-almost all ω_2 trajectories of the B_H process. Because this process is independent from the B process it follows that the no-arbitrage result is also valid in the product space. $\qquad\qquad\square$

3.2   Incompleteness

In this financial model, trading takes place only in the risky asset and in the money market. As a consequence the volatility risk cannot be hedged. Having more sources of risk than tradable assets, suggests that the market is incomplete

Theorem 2 *The market defined by (A_t, S_t, σ_t) is incomplete*

Proof Use an integral representation for the fractional Brownian motion

$$B_H(t) = \int_0^t K_H(t, s) dW_s \tag{24}$$

W_t being a Brownian motion independent from B_t and K_H is the square integrable kernel

$$K_H(t, s) = C_H s^{\frac{1}{2} - H} \int_s^t (u - s)^{H - \frac{3}{2}} u^{H - \frac{1}{2}} \, du, \quad s < t \tag{25}$$

$(H > 1/2)$. Then the process

$$\eta'_t = \exp(W_t - \frac{1}{2}t) \tag{26}$$

is a square-integrable P_2–martingale. Now, define a standard bi-dimensional Brownian motion, $W^*_t = (B_t, W_t)$ and the process $\eta^*_t(\omega_2) = \eta_t \eta'_t(\omega_2)$

$$\eta^*_t(\omega_2) = \exp\left\{\int_0^t \Gamma_s(\omega_2) \bullet dW^*_t - \frac{1}{2}\int_0^t \|\Gamma_s(\omega_2)\|^2 \, ds\right\} \tag{27}$$

where, by the Lemma 1, $\Gamma(\omega_2) = (\gamma(\omega_2), 1)$ is also a P_1–martingale. Then, by the Cameron-Martin-Girsanov theorem, the process $\widetilde{W}^*_t = (\widetilde{W}^{*(1)}_t, \widetilde{W}^{*(2)}_t)$ defined by

$$\widetilde{W}^{*(1)}_t = B_t - \int_0^t \gamma_s(\omega_2)ds; \;\; \widetilde{W}^{*(2)}_t = W_t - t \tag{28}$$

is a bi-dimensional Brownian motion on the probability space $(\Omega_1, \mathscr{F}_1, Q^*_T(\omega_2))$, where $Q^*_T(\omega_2)$ is the probability measure $\frac{dQ^*_T(\omega_2)}{dP_1} = \eta^*_T(\omega_2)$. Moreover, the discounted price process Z remains a martingale with respect to the new measure $Q^*_T(\omega_2)$. $Q^*_T(\omega_2)$ being an equivalent martingale measure distinct from $Q_T(\omega_2)$, the market is incomplete. □

3.3 Leverage, a Modified Model and Completeness [5, 7]

The following nonlinear correlation of the returns

$$L(\tau) = \langle |r(t+\tau)|^2 r(t)\rangle - \langle |r(t+\tau)|^2\rangle \langle r(t)\rangle \tag{29}$$

is called *leverage* and the *leverage effect* is the fact that, for $\tau > 0$, $L(\tau)$ starts from a negative value whose modulus decays to zero whereas for $\tau < 0$ it has almost negligible values (see Fig. 2 which shows a typical behavior in the NYSE data).

As expressed in (10) the fractional volatility model has the volatility process σ_t acting on the log-price, but not conversely. Therefore, in its simplest form, the *fractional volatility model contains no leverage effect*.

However, leverage may be implemented by a small modification of the model. Use a (truncated) representation for fractional Brownian motion as a stochastic integral over Brownian motion and *identify the random generator of the log-price process with the stochastic integrator of the volatility*. A leverage effect is then obtained.

$$\mathscr{H}(t) = \Pi^{(M)}\left[C_H\left\{\begin{array}{l}\int_{-\infty}^0 ((t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}})dW_u \\ + \int_0^t (t-u)^{H-\frac{1}{2}}dW_u\end{array}\right\}\right] \tag{30}$$
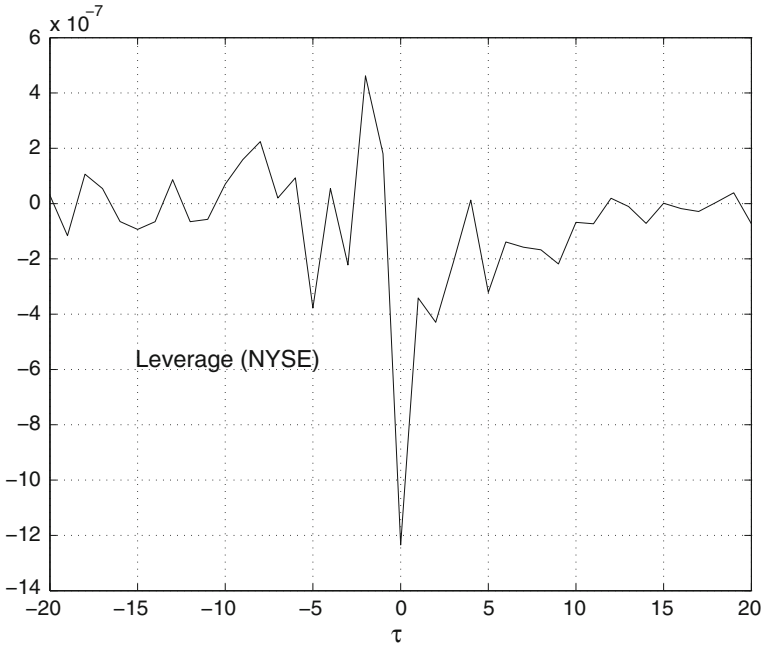
Fig. 2 Example of the leverage effect in NYSE data

$\Pi^{(M)}$ meaning the truncation of the representation to an interval $[-M, M]$ with M arbitrarily large.

Now, instead of two, there is only one source of risk and the new fractional volatility model would be

$$dS'_t = \mu_t S'_t dt + \sigma_t S'_t dW_t$$

$$\log \sigma'_t = \beta + \frac{k'}{\delta} \{ \mathscr{H}(t) - \mathscr{H}(t - \delta) \} \tag{31}$$

Theorem 3 *The market (A_t, S'_t, σ'_t) is free of arbitrage and complete.*

Proof Because the two processes are not independent we cannot use the same argument as before to obtain the Kallianpur condition. However, with the truncation, the Hölder condition is trivially verified for all the truncated paths of σ'_t and the construction of an equivalent martingale measure follows the same steps as in Theorem 2. Hence we have a P_1-martingale with respect to $(\mathscr{F}_{1,t})_{0 \leq t < T}$

$$\eta_t = \exp \left\{ \int_0^t \frac{r - \mu_s}{\sigma_s} dW_s - \frac{1}{2} \int_o^t (\frac{r - \mu_s}{\sigma_s})^2 ds \right\} \tag{32}$$

and Q_T, defined by $\frac{dQ_T}{dP_1} = \eta_T$ is an equivalent martingale measure.

The set of equivalent local martingale measures being non-empty, let Q^* be an element in this set. By the Girsanov converse there is a $\mathbb{R}$-valued process ϕ such that the Radon-Nikodym density of Q^* is

$$\frac{dQ_T^*}{dP_1} = \exp\left\{\int_0^T \phi_s dW_s - \frac{1}{2}\int_0^T \phi_s^2 ds\right\} \tag{33}$$

Moreover the process W_t^* given by

$$W_t^* = W_t - \int_0^t \phi_s ds \tag{34}$$

is a standard Q^*–Brownian motion and the discounted price process Z' satisfies the following stochastic differential equation

$$dZ_t' = (\mu_t - r + \sigma_t'\phi_t)Z_t'dt + \sigma_t Z_t'dW_t^* \tag{35}$$

Because Z_t' is a Q^*–martingale, then it must be $\mu(t,\omega) - r + \sigma'(t,\omega)\phi(t,\omega) = 0$ almost everywhere w.r.t. $dt \times P$ in $[0,T] \times \Omega$. It implies

$$\phi(t,\omega) = \frac{r - \mu(t,\omega)}{\sigma'(t,\omega)} \tag{36}$$

a. e. $(t,\omega) \in [0,T] \times \Omega_1$. Hence $Q_T^* = Q_T$, that is, Q_T is the unique equivalent martingale measure. This market model is complete. □

Figure 3 compares the leverage effect in the two models, that is, the original FVM (10) and the modified one (31).

3.4 A Remark on Long Memory and Fractional Brownian Motion

In the past, several authors had already tried to describe long memory effects in the market data by replacing in the price process Brownian motion by fractional Brownian motion with $H > 1/2$. However it was soon realized [8–11] that this replacement implied the existence of arbitrage. These results might be avoided either by restricting the class of trading strategies [12], introducing transaction costs [13] or replacing pathwise integration by a different type of integration [14, 15]. However this is not free of problems because the Skorohod integral approach requires the use of a Wick product either on the portfolio or on the self-financing condition, leading to unreasonable situations from the economic point of view (for example positive portfolio with negative Wick value, etc.) [16].
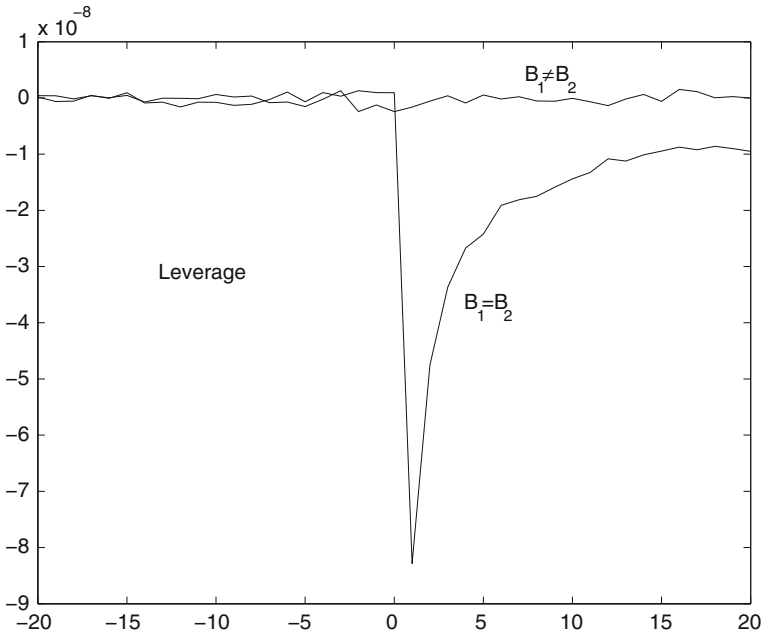
Fig. 3 Comparison of leverage in the original and the modified model

The fractional volatility model in Eq. (10) is not affected by these considerations, because it is the volatility process that is driven by fractional noise, not the price process and, as shown, a no-arbitrage result may be proven. This is no surprise because the requirement (H2) that, for each sample path $\omega_2 \in \Omega_2$, $\log S_t(\cdot, \omega_2)$ is a square integrable random variable in Ω_1 already implies that $\int \sigma_t dB_t$ is a martingale. The square integrability is also essential to guarantee the possibility of reconstruction of the σ process from the data.

4 Agent-Based Interpretation of the Fractional Volatility Model

In [17] two agent-based models were considered:

- In the first the traders strategies play a determinant role.
- In the second the determinant effect is the limit-order book dynamics, the agents having a random nature.

4.1 A Market Model with Self-adapted or Fixed Strategies

A set of investors is playing *against* the market. In addition to the impact of this group of investors, other factors are represented by a stochastic process η_t

$$z_{t+1} = f(z_t, \omega_t) + \eta_t \tag{37}$$

$(z_t = \log S_t)$ and ω_t is the total investment made by the group of traders. After r time steps, s agents copy the strategy of the s best performers and, at the same time, have some probability to mutate that strategy.

The model was run with different initial conditions and with or without evolution of the strategies. When the model is run with evolution the asymptotic steady-state behavior depends on the initial conditions. Different types of return statistics corresponded to the relative importance of either "value investors" or "technical traders". The occurrence of market bubbles and fat tails corresponds to situations where technical trader strategies were well represented. A situation where there are 50 % of fundamental (value-investing) strategies and 50 % trend-following, was chosen to compare its statistical properties with those of the FVM. The squared volatility $\sigma_t^2 = \frac{1}{|T_0 - T_1|} var(\log p_t)$ and the parameters in $\sum_{n=0}^{t/\delta} \log \sigma(n\delta) = \beta t + R_\sigma(t)$ and $|R_\sigma(t + \Delta) - R_\sigma(t)|$ were estimated from the simulations. Figure 4 shows the results.

Notice the lack of scaling behavior of $R_\sigma(t)$ with an asymptotic exponent 0.55, denoting the lack of memory of the volatility process. This might already be evident from the time behavior of $R_\sigma(t)$ in the lower left plot. Also, although the returns have fat tails in this case, they are of different shape from those observed in the market data. Similar conclusions are obtained with other combinations of agent strategies.

In conclusion: *It seems that the features of the fractional volatility model (which are also those of the bulk market data) are not easily captured by a choice of strategies in an agent-based model*.

Agents' reactions and strategies are very probably determinant during market crisis and market bubbles but not in business-as-usual days.

4.2 A Limit-Order Book Market Model

In this model, *asks* and *bids* arrive at random on a window $[S_t - w, S_t + w]$ around the current price S_t. Every time a *buy* order arrives it is fulfilled by the closest non-empty ask slot, the new current price being determined by the value of the ask that fulfills it. If no ask exists when a buy order arrives it goes to a cumulative register to wait to be fulfilled. The symmetric process occurs when a *sell* order arrives, the new price being the bid that buys it. Sell and buy orders, asks and bids all arrive at random. Because the window around the current price moves up and down, asks and bids that are too far away from the current price are automatically eliminated.
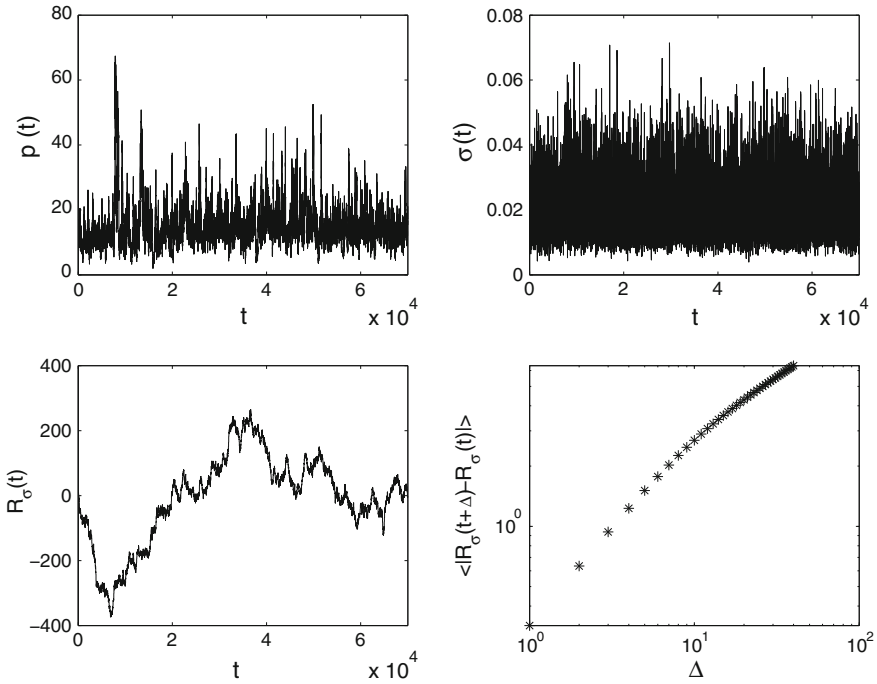
Fig. 4 Statistical properties of the model with agent strategies

The only parameters of the model are the width w of the limit-order book and the size n of the asks and bids, the sell and buy orders being normalized to one.

The model was run for different widths w and liquidities n. Although the exact values of the statistical parameters depend on w and n, the statistical nature of the results is essentially the same. In the Fig. 5 ($n = 2$) the limit-order book is divided into $2w + 1 = 21$ discrete price slots with $\Delta p = 0.1$. The scaling properties of $R_\sigma(t)$ are quite evident from the lower right plot in the figure, the Hurst coefficient being 0.96.

Conclusion: *the main statistical properties of the market data (fast decay of the linear correlation of the returns, non-Gaussianity and volatility memory) are already generated by the dynamics of the limit-order book with random behavior of the agents. A large part of the market statistical properties (in normal business-as-usual days) depends more on the nature of the price fixing financial institutions than on particular investor strategies.*
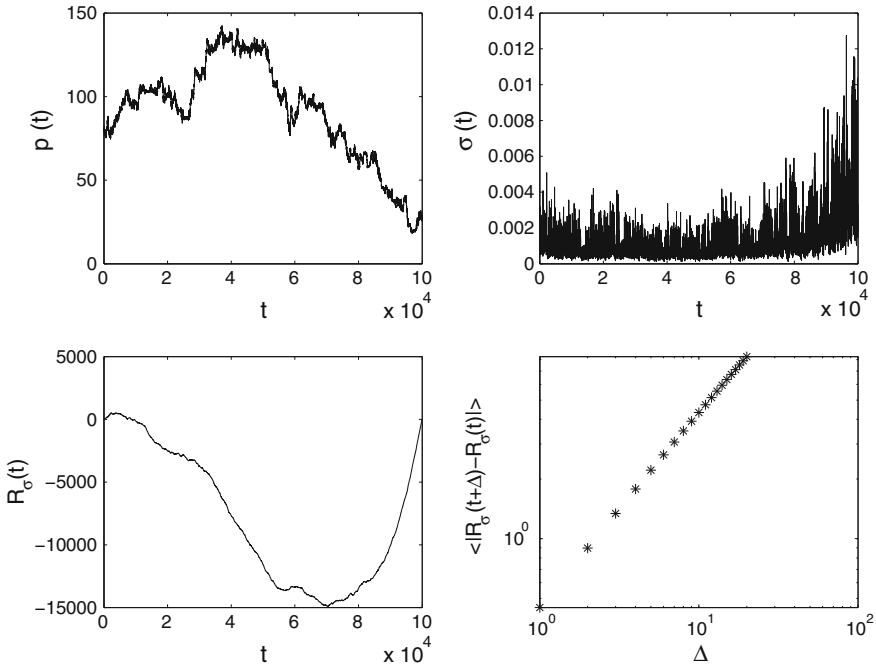
Fig. 5 Statistical properties of the limit order book model

5   Further Properties of the Fractional Volatility Model

In the FVM the statistics of returns is obtained in closed form. From

$$P_\delta(r(\Delta)) = \frac{1}{4\pi\theta k\delta^{H-1}\sqrt{\Delta}} \int_0^\infty dx x^{-\frac{1}{2}} e^{-\frac{1}{C}(\log x)^2} e^{-\lambda x} \tag{38}$$

$$r(\Delta) = \log S_{t+\Delta} - \log S_t \ , \ \theta = e^\beta, \ \lambda = \frac{(r(\Delta) - r_0)^2}{2\Delta\theta^2} \tag{39}$$

$$r_0 = \left(\mu - \frac{\sigma^2}{2}\right)\Delta \ , \ C = 8k^2\delta^{2H-2} \tag{40}$$

one obtains

$$P_\delta(r(\Delta)) = \frac{1}{4\pi\theta k\delta^{H-1}\sqrt{\Delta}} \frac{1}{\sqrt{\lambda}} \left. (e^{-\frac{1}{C}(\log\lambda - \frac{d}{dz})^2} \Gamma(z)) \right|_{z=\frac{1}{2}} \tag{41}$$
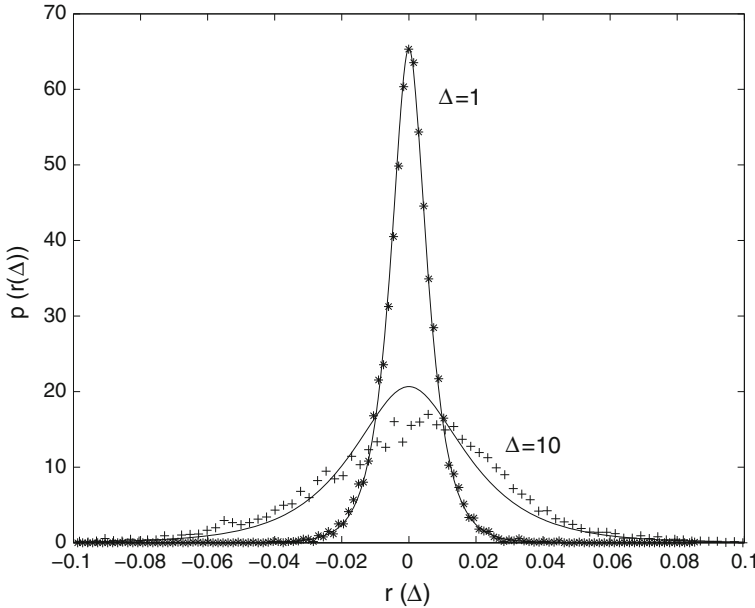
with asymptotic behavior, for large returns

Fig. 6 Statistics of returns in the FVM compared with NYSE data

$$\boxed{P_\delta(r(\Delta)) \sim \frac{1}{\sqrt{\Delta\lambda}} e^{-\frac{1}{C}\log^2\lambda}} \tag{42}$$

This form provides a good fit of the empirical data. Figure 6 compares NYSE data with (41) for $H = 0.83$, $k = 0.59$, $\beta = -5$, $\delta = 1$, $\Delta = 1$ and $\Delta = 10$.

That NYSE returns are well described by (41) is no wonder because the FVM itself was reconstructed from that data. What seemed stranger, at first sight, was the fact that, once the parameters of the model are fixed, a simple change of Δ would predict the returns in a quite different market. This is shown in Fig. 7 where the same parameters as above were used, simply changing to $\Delta = \frac{1}{440}$ (1 min). The prediction of the model is compared with 1-min data of USDollar-Euro market for a couple of months in 2001. The result may be surprising, because one would not expect the volatility parametrization to carry over to such a different time scale and also because one is dealing with a different market. However, if the conclusion from the agent-based models is correct, that in business-as-usual days the statistics of the data depends more on the price fixation process than on agent strategies or other market features, then this result is no longer a surprise.

Using a simple risk neutrality argument, a new option pricing was also obtained, namely

$$\boxed{\begin{aligned} V(S_t, \sigma_t, t) = \{&S_t \left[aM\left(\alpha, a, b\right) + bM\left(\alpha, b, a\right)\right] \\ &-Ke^{-r(T-t)}\left[aM(\alpha, a, -b) - bM(\alpha, -b, a)\right]\}\end{aligned}} \tag{43}$$
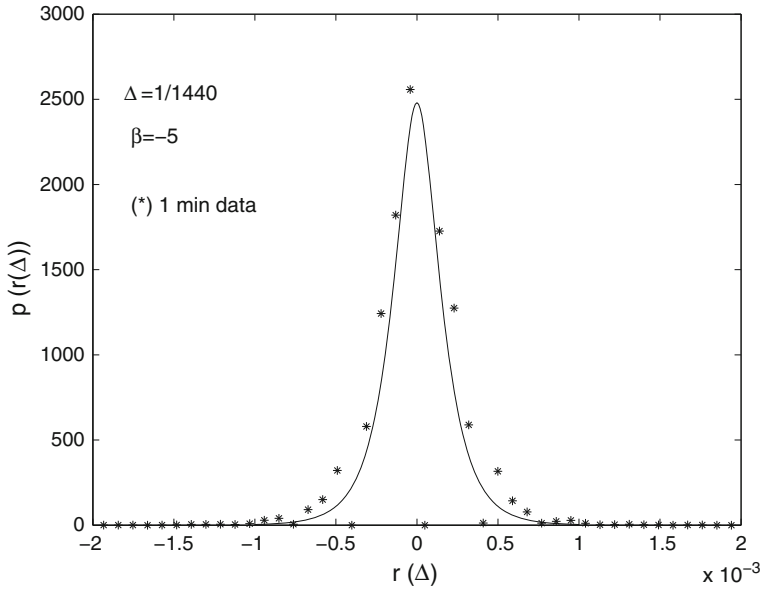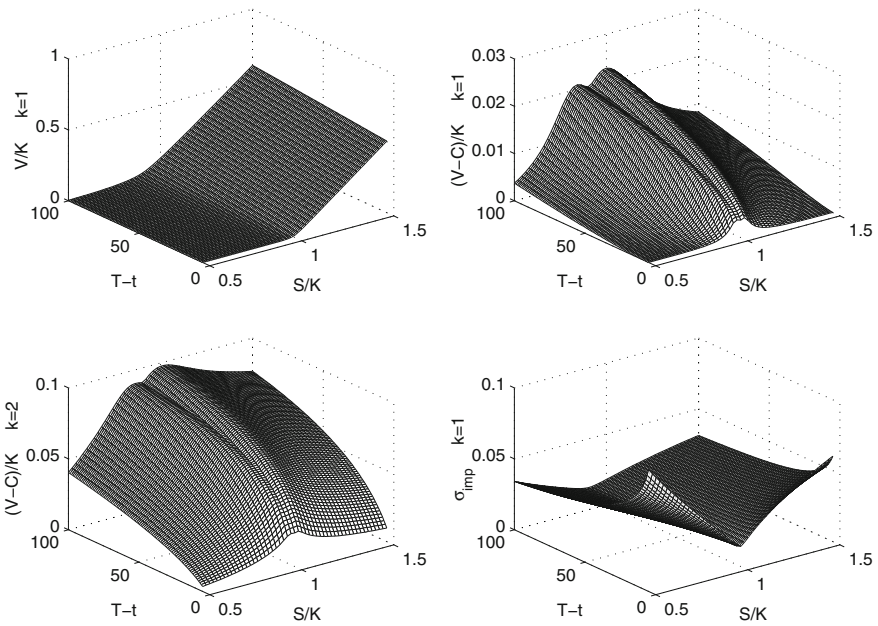
Fig. 7 Statistics of returns for 1 min exchange data

Fig. 8 Option price results V/K, comparison with Black-Scholes $(V - C)/K$ and the corresponding implied volatility

with

$$M(\alpha, a, b) = \frac{1}{2\pi\alpha} \int_{-1}^{\infty} dy \int_{0}^{\infty} dx e^{-\frac{\log^2 x}{2\alpha^2}} e^{-\frac{y^2}{2}(ax+\frac{b}{x})^2}$$

$$= \frac{1}{4\alpha} \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} dx \frac{e^{-\frac{\log^2 x}{2\alpha^2}}}{ax + \frac{b}{x}} erfc\left(-\frac{ax}{\sqrt{2}} - \frac{b}{\sqrt{2}x}\right) \qquad (44)$$

K is the strike price and T the maturity time. In Fig. 8 is plotted $V(S_t, \sigma_t, t)$ in the range $T - t \in [5, 100]$ with $S/K \in [0.5, 1.5]$ as well as $(V(S_t, \sigma_t, t) - C(S_t, \sigma_t, t))$ /K for $k = 1$ and $k = 2$. $C(S_t, \sigma_t, t)$ is the Black-Scholes result. Other parameters are fixed at $\sigma = 0.01$, $r = 0.001$, $\delta = 1$, $H = 0.8$. To compare with Black-Scholes (BS), the implied volatility, that would reproduce the same results, was also computed. The implied volatility surface corresponding to $V(S_t, \sigma_t, t)$ is shown for for $k = 1$. It predicts a smile effect with the smile increasing as maturity approaches.

References

1. Föllmer, H., Schied, A.: Stochastic Finance: An Introduction in Discrete Time. Walter de Gruyter, New York (2004)
2. Gerardi, G.S., Lehnert, S., Sherland, S.M., Willen, P.S.: Making sense of the subprime crisis, Working paper 2009-2, Federal Bank of Atlanta
3. Mendes, R.V., Oliveira, M.J.: A data-reconstructed fractional volatility model, E-Economics/discussion paper/2008-22
4. Nualart, D.: The Malliavin Calculus and Related Topics. Springer, Berlin (2006)
5. Mendes, R.V., Oliveira, M.J., Rodrigues, A.M.: No-arbitrage, leverage and completeness in a fractional volatility model. Phys. A: Stat. Mech. Appl. **419**, 470–478 (2015)
6. Björk, T.: Arbitrage Theory in Continuous Time. Oxford University Press, New York (2009)
7. Mendes, R.V.: A fractional calculus interpretation of the fractional volatility model. Nonlinear Dyn. **55**, 395–399 (2009)
8. Rogers, L.C.G.: Arbitrage with fractional Brownian motion. Math. Finance **7**, 95–105 (1997)
9. Shiryaev, A.N.: On arbitrage and replication for fractal models, Research Report 20. Department of Mathematical Sciences, University of Aarhus, Denmark, MaPhySto (1998)
10. Salopek, D.M.: Tolerance to arbitrage. Stochast. Process. Appl. **76**, 217–230 (1998)
11. Sottinen, T.: Fractional Brownian motion, random walks and binary market models. Finance Stochast. **5**, 343–355 (2001)
12. Cheridito, P.: Arbitrage in fractional Brownian motion models. Finance Stochast. **7**, 533–553 (2003)
13. Guasoni, P.: No arbitrage under transaction costs, with fractional Brownian motion and beyond. Math. Finance **16**, 569–582 (2006)
14. Hu, Y., Øksendal, B.: Fractional white noise calculus and applications to finance. Infinite Dimensional Anal. Quantum Probab. Relat. Top. **6**, 1–32 (2003)
15. Elliott, R.J., van der Hoek, J.: A general fractional white noise theory and applications to finance. Math. Finance **13**, 301–330 (2003)
16. Björk, T., Hult, H.: A note on Wick products and the fractional Black-Scholes model. Finance Stochast. **9**, 197–209 (2005)
17. Mendes, R.V.: The fractional volatility model: an agent-based interpretation. Phys. A: Stat. Mech. Appl. **387**, 3987–3994 (2008)