Finance Stoch (2010) 14: 157–177
DOI 10.1007/s00780-009-0112-1

From implied to spot volatilities

Valdo Durrleman

Received: 2 February 2007 / Accepted: 19 December 2008 / Published online: 7 November 2009

Abstract This paper is concerned with the relation between spot and implied volatilities. The main result is the derivation of a new equation which gives the dynamics of the spot volatility in terms of the shape and the dynamics of the implied volatility surface. This equation is a consequence of no-arbitrage constraints on the implied volatility surface right before expiry. We first observe that the spot volatility can be recovered from the limit, as the expiry tends to zero, of at-the-money implied volatilities. Then, we derive the semimartingale decomposition of implied volatilities at any expiry and strike from the no-arbitrage condition. Finally the spot volatility dynamics is found by performing an asymptotic analysis of these dynamics as the expiry tends to zero. As a consequence of this equation, we give general formulas to compute the shape of the implied volatility surface around the at-the-money strike and for short expiries in general spot volatility models.

Keywords Option price · Implied volatility · Spot volatility · Martingale representation · Asymptotic analysis · Itô–Wentzell formula

Mathematics Subject Classification (2000) 60H10 · 91B28

JEL Classification C60 · G13

I am extremely grateful to my Ph.D. advisor René Carmona and to Nicole El Karoui for their help and support. I should also like to thank the referees and the editor for their constructive remarks which have greatly enhanced the clarity of the paper.

V. Durrleman (✉)
Centre de Mathématiques Appliquées, École Polytechnique–CNRS, Route de Saclay,
91128 Palaiseau, France
e-mail: valdo@cmap.polytechnique.fr

1 Introduction

The aim of this paper is to answer the following question: How much is known about an underlying security's volatility if we can observe sufficiently many European call and put option prices written on that security? More generally, we should like to better understand the relationship between spot and implied volatilities. The usual way to go from a spot volatility process σ to the corresponding implied volatilities is easy to state but rather convoluted in practice. One has to solve the stochastic differential equations for the stock price S together with those driving the other economic factors, then compute expectations under the law of the terminal stock price, and finally invert the Black–Scholes formula. Implied volatilities are the market-observable quantities, whereas σ is a mathematical construction which is the key concept when it comes to the risk management of exotic options.

This important practical question has been studied for quite some time. Let us first mention the very well-known result of [5]. If $C_t(T, K)$ denotes the price at time t of a call option with strike K and maturity T, then knowing $C_t(T, K)$ at time t for all K and for a given T is equivalent to knowing the risk-neutral distribution of the terminal value of the stock price S_T.

An important further step was achieved in [11, 12], and [23]. For instance, it is shown in [12] that under the assumption that the spot process is continuous and Markov under the pricing measure, the distribution of the process describing the future evolution is completely specified by implied volatilities. More precisely, let us suppose that the spot volatility is a deterministic function of the current spot price and possibly time, say, $\sigma(t, S)$. The above works show that we can recover this function if we observe the entire implied volatility surface (as a function of T and K) at a given date.

We tackle this problem from a different angle in the sense that we study dynamics instead of focusing on data at a given date. Loosely speaking, the previous works used Markov methods where (t, ω) are essentially fixed, whereas (T, K) vary. We keep (T, K) "fixed" and let (t, ω) vary. As explained by [17] and [8] among many others, such a dynamical approach is very useful from a practitioner's point of view. Implied volatility dynamics are crucial when it comes to risk management of portfolios of options. Having a model for the implied volatility dynamics allows one to calculate the different vega risks in a consistent manner. Dynamics are also crucial for hedging exotic options. This is particularly clear in the case of a barrier option: Its price is precisely a function of the implied volatility smile when the spot touches the barrier, as made apparent by the classical static hedging argument of [3].

To solve the problem, we take a point of view that is very close to that of [18], where a continuum of assets is modeled. We consider here a market where the primary securities are the spot and liquid options on it. We restrict ourselves to a continuous spot process. Apart from that, we try to make as few assumptions as possible on the spot dynamics in order to have the most general understanding. We shall see that there is a lot of information about the spot process that is contained in option prices. In fact, we shall see that under some regularity conditions, one can recover the spot volatility *dynamics* by observing implied volatilities' *dynamics*.

First, the observation is made that at any time t, the implied volatility of a very short maturity at-the-money option is equal to the spot volatility (corollary to Theorem 3.1). Therefore the dynamics of the spot volatility can be obtained from the limit of the implied volatility surface dynamics as the time to maturity tends to zero. The martingale representation theorem yields a predictable representation for the price of a vanilla option (and hence its implied volatility) in terms of the underlying Wiener process. It is then shown that subject to suitable conditions, the implied volatility surface at time t is a family of semimartingales parameterized in a continuous fashion by time to maturity and strike. The Doob–Meyer decomposition of this semimartingale is then obtained from the Itô–Wentzell formula (Proposition 3.4). The limit of this semimartingale as the time to maturity tends to zero in this decomposition is taken by using dominated convergence and continuity of quadratic variation, which, as originally observed, gives a semimartingale decomposition of the spot volatility process.

A few papers ([4, 21, 24], and more recently, [25]) have already studied a financial market where the primary securities are the spot and liquid options on it. These papers look at the problem of modeling implied volatilities in an arbitrage-free way. We do not tackle this very interesting and challenging problem. Instead, we take the implied volatility dynamics as given and perform an asymptotic analysis to find new and interesting relations between the spot and implied volatilities.

For an excellent survey on the subject of implied volatilities, one should consult [22]. The book [16] provides an excellent account on the topic from the practical perspective.

Modeling implied volatilities is equivalent to modeling option prices, as seen in (2.3), and it is merely a parameterization of options' volatilities. The difficulties in modeling implied volatilities have led researchers to look for other and possibly more tractable parameterizations.

First, following a program started in [6, 10] model option prices by modeling Dupire's local volatility as a random field. They are able to find explicit drift conditions and some examples of such dynamics. The Dupire local volatility surface specifies the spot volatility in the short maturity limit just like implied volatilities do (corollary to Theorem 3.1). However, it does not have complicated static arbitrage restrictions in terms of butterfly, call, and calendar spreads like implied volatilities (see, for instance, [16]).

Another way of parameterizing option prices consists in modeling their intrinsic values, i.e., the difference between the option price and the payoff if the option was exercised today. This is the approach taken by [19] in a very general semimartingale framework. Exactly as with implied volatilities (corollary to Theorem 3.1), this approach yields a spot specification when options are close to maturity.

Finally, let us mention the recent work [26], where the authors introduce new quantities, the "local implied volatilities" and "price level", to parameterize option prices. These have nicer dynamics and naturally satisfy the static arbitrage conditions. They derive an existence result for the infinite system of equations driving these quantities.

The paper is organized as follows. Section 2 introduces the framework and notation for the derivation of the main results. In Sect. 3 we establish the main results of the paper, namely Theorems 3.1 and 3.2. In Sect. 4, we study the particular case of stochastic volatility models. Finally, Sect. 5 hints at potential applications of these results.

2 The implied volatility surface dynamics and its boundary behavior

2.1 Framework and notation

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space with an n-dimensional Wiener process $(\boldsymbol{W}_t)_{t\geq 0}$ on it. We shall use boldface letters for vectors. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are two vectors, then $\boldsymbol{x} \cdot \boldsymbol{y}$ denotes their usual scalar product and $|\boldsymbol{x}|$ the Euclidean norm of $\boldsymbol{x}$. The filtration generated by the Wiener process has been augmented as usual and is denoted by $(\mathcal{F}_t)_{t\geq 0}$. We assume that the probability measure $\mathbb{P}$ is risk neutral, that is, discounted price processes are local martingales if there is "no arbitrage" in the market. This assumption is based on the famous fundamental theorem of asset pricing (see, e.g., [9]).

We first define the risk-neutral stock process S. For simplicity, we assume that interest rates are zero, but it would make no difference here if they were deterministic functions of time. We are given a spot volatility process $\boldsymbol{\sigma}$ taking values in $\mathbb{R}^n$, which is jointly measurable, adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$, and satisfies the integrability condition

$$\forall t \geq 0 \qquad \int_0^t |\boldsymbol{\sigma}_s|^2 \,\mathrm{d}s < \infty \quad \text{a.s.} \tag{2.1}$$

Since $\boldsymbol{\sigma}$ satisfies (2.1), we can define S to be the stochastic exponential

$$S_t = S_0 \exp\left(\int_0^t \boldsymbol{\sigma}_s \cdot \mathrm{d}\boldsymbol{W}_s - \frac{1}{2}\int_0^t |\boldsymbol{\sigma}_s|^2 \,\mathrm{d}s\right).$$

In other words, S is a typical positive local martingale in a Brownian filtration. We shall assume that it is a martingale. Let us stress that $\boldsymbol{\sigma}$ is adapted to the entire filtration $(\mathcal{F}_t)_{t\geq 0}$ generated by $(\boldsymbol{W}_t)_{t\geq 0}$, which usually is larger than that generated by $(S_t)_{t\geq 0}$. In financial terms, we do not assume market completeness. We make the following basic nondegeneracy assumption:

Assumption 1 S is a martingale (relative to $(\mathcal{F}_t)_{t\geq 0}$), and $|\boldsymbol{\sigma}_t(\omega)| > 0$, $Leb \otimes \mathbb{P}$-a.e.

Our main and fundamental assumption is that options are liquid instruments, just as the stock itself. Each option is therefore an attainable claim, and in the absence of arbitrage opportunities, we can assume that option prices agree with the risk-neutral expectation of their terminal payoffs.

More precisely, let $t \geq 0$, $T > t$, and $K > 0$ be fixed; $C_t(T, K)$ denotes the price at time t of the call option struck at K with maturity T.

Assumption 2 For every $T > t$ and K,

$$C_t(T, K) = \mathbb{E}\left\{(S_T - K)^+ \middle| \mathcal{F}_t\right\} \quad \text{a.s.} \tag{2.2}$$

The fact that call prices are martingales will play a key role in deriving the implied volatility dynamics via the martingale representation theorem in Sect. 2.3.

We wish to study $C_t(T,K)$ as a function of (T,K). We are dealing with an uncountable family of stochastic processes indexed by (T,K). We assume that all these processes have been constructed simultaneously on our probability space.

Assumption 3 For every $T>t$ and K,

$$(S_t-K)^+ < C_t(T,K) < S_t \quad \text{a.s.}$$

Given an option price $C_t(T,K)$, Assumption 3 ensures that $C_t(T,K)$ is in the range of the Black–Scholes formula. We can therefore define the Black–Scholes implied volatility $\Sigma_t(T,K)$ as the unique volatility parameter for which the Black–Scholes formula recovers the option price. In other words, $\Sigma_t(T,K)$ is the unique solution to the equation

$$K\mathcal{BS}\big(S_t/K, \Sigma_t(T,K)\sqrt{T-t}\big) = C_t(T,K), \tag{2.3}$$

where $\mathcal{BS}$ is the normalized Black–Scholes function

$$\mathcal{BS}(u,v) = u\Phi\left(\frac{\ln u}{v}+\frac{v}{2}\right) - \Phi\left(\frac{\ln u}{v}-\frac{v}{2}\right).$$

We use the notation $\varphi(x)$ and $\Phi(x)$ for the density and the cumulative distribution functions of the standard Gaussian distribution.

Assumption 3 ensures that for every $T>t$ and K, $\Sigma_t(T,K)>0$ a.s.

2.2 Boundary behavior

In this subsection, we study the behavior of implied volatilities just before maturity. This behavior is intimately related to that of the corresponding option prices.

In the absence of arbitrage, option prices are continuous in maturity. Indeed, we can rewrite (2.2) as

$$C_t(T,K) = \mathbb{E}^{\mathbb{P}_t}\big\{(S_T-K)^+\big\},$$

where we denote by $\mathbb{P}_t$ a fixed version of the regular conditional distribution of the process S under $\mathbb{P}$ with respect to $\mathcal{F}_t$. Since S is a martingale, call–put parity yields

$$C_t(T,K) = S_t - K - \mathbb{E}^{\mathbb{P}_t}\big\{(K-S_T)^+\big\},$$

and the dominated convergence theorem (puts are bounded by their strikes) gives that, for each $t \ge 0$, we have

$$\lim_{T\downarrow t} C_t(T,K) = (S_t-K)^+ \quad \text{a.s.} \tag{2.4}$$

This property translates into the following for the implied volatilities:

Proposition 2.1 *For t and $K > 0$ fixed,*

$$\lim_{T \downarrow t} \Sigma_t(T, K)\sqrt{T-t} = 0 \quad a.s. \tag{2.5}$$

Proof Suppose that $\limsup_{T\downarrow t} \Sigma_t(T, K)\sqrt{T-t} > 0$. This would then imply that $\limsup_{T\downarrow t} K\mathcal{BS}(S_t/K, \Sigma_t(T, K)\sqrt{T-t}) > (S_t - K)^+$. In view of (2.3), that would contradict (2.4). □

Proposition 2.2 *For every $K > 0$,*

$$\lim_{\theta\to 0} \Sigma_t(t+\theta, K)\sqrt{\theta} = 0$$

in the sense of uniform convergence on compact intervals in t and in probability (u.c.p. for short).

Proof Let $\theta_n \downarrow 0$. From Proposition 2.1 we know that for every t and K, $\Sigma_t(t+\theta_n, K)\sqrt{\theta_n}$ goes to 0 for a.e. ω as n goes to infinity. To show that the pointwise limit is in fact uniform in t on compact intervals for a.e. ω, we note that call option prices are increasing with maturity (recall that interest rates are zero) and therefore that $\theta \mapsto \Sigma_t(t+\theta, K)\sqrt{\theta}$ is increasing. Then, using Dini's theorem since the limit is obviously continuous, we get the uniform convergence in t for any decreasing sequence (θ_n). Now, let us take an arbitrary sequence $\theta_n \to 0$; we want to show that $\Sigma_t(t+\theta_n, K)\sqrt{\theta_n}$ converges to 0 uniformly in t on compacts and in probability. Take a subsequence (θ_{n_k}); it has a further subsequence that is decreasing to 0, and the preceding reasoning applies to that subsequence. The convergence therefore holds in the sense of uniform convergence on compacts in t and in probability. □

The proposition above is the first key ingredient of the paper. The second ingredient is the implied volatility dynamics (more precisely, the implied variance dynamics). This is the content of the next section.

2.3 Implied volatility dynamics

We are going to derive the dynamics for Σ from that of C. In view of (2.2), for each (T, K), $C(T, K)$ is a martingale adapted to a Brownian filtration. By the martingale representation theorem, we write it as a stochastic integral. More specifically, there exists an adapted process $\boldsymbol{H}(T, K)$ such that for $t \le T$,

$$C_t(T, K) = C_0(T, K) + \int_0^t \boldsymbol{H}_s(T, K) \cdot \mathrm{d}\boldsymbol{W}_s.$$

We assume that such a representation holds simultaneously for all $T > t$ and K a.s., and $\boldsymbol{H}_t(T, K)$ is a family of stochastic processes indexed by $T > t$ and K.

By Assumption 3, $\Sigma_t(T, K) > 0$ for $(T, K) \in (t, +\infty) \times (0, +\infty)$ a.s. We can therefore define the random variables $\boldsymbol{\xi}_t(T, K)$ for $(T, K) \in (t, +\infty) \times (0, +\infty)$ by

$$\boldsymbol{\xi}_t(T, K) = \frac{\boldsymbol{H}_t(T, K) - \Phi(d_1) S_t \boldsymbol{\sigma}_t}{S_t \Sigma_t(T, K)\sqrt{T-t}\,\varphi(d_1)}, \tag{2.6}$$

where, as usual, $d_1 = \frac{\ln(S_t/K)}{\Sigma_t(T,K)\sqrt{T-t}} + \frac{1}{2}\Sigma_t(T,K)\sqrt{T-t}$. Then, the martingale representation takes the form

$$C_t(T,K) = C_0(T,K) + \int_0^t S_s\big[\Phi(d_1)\boldsymbol{\sigma}_s + \Sigma_s\sqrt{T-s}\varphi(d_1)\boldsymbol{\xi}_s\big](T,K)\cdot \mathrm{d}\boldsymbol{W}_s.$$

Writing $\boldsymbol{H}_t(T,K)$ in this way will prove easier to handle when it comes to the study of implied volatilities.

The above martingale representation has the following financial interpretation. To manage the risk of a short option, one can buy the Black–Scholes quantity of $\Delta = \Phi(d_1)$ stocks, whose dynamics are given by the first term above. The second term is a correction term proportional to the vega that takes into account the fact that the spot volatility is not constant. In the Black–Scholes world, that last term would be zero, and the risk management of the short option could be done with a self-financing portfolio consisting solely of S and a risk-free bank account.

Itô's formula shows that for fixed (T,K), $(\Sigma_t(T,K))$ is a semimartingale on $t \le T$. The proposition below gives its decomposition. From it, we note that $\boldsymbol{\xi}_t(T,K)$ has the interpretation of the implied volatility's volatility vector.

Proposition 2.3 *For fixed* (T,K) *and* $t < T$,

$$\begin{aligned}\Sigma_t(T,K)\sqrt{T-t} = \Sigma_0(T,K)\sqrt{T} - \int_0^t \Bigg[\frac{|\boldsymbol{\sigma}_s - \ln(S_s/K)\boldsymbol{\xi}_s|^2}{2\Sigma_s\sqrt{T-s}} \\ + \frac{1}{2}\Sigma_s\sqrt{T-s}\boldsymbol{\sigma}_s\cdot\boldsymbol{\xi}_s - \frac{1}{8}\Sigma_s^3(T-s)^{3/2}|\boldsymbol{\xi}_s|^2\Bigg](T,K)\,\mathrm{d}s \\ + \int_0^t \Sigma_s(T,K)\sqrt{T-s}\boldsymbol{\xi}_s(T,K)\cdot\mathrm{d}\boldsymbol{W}_s.\end{aligned}$$

Proof Since $C_t(T,K) = K\mathcal{BS}(S_t/K, \Sigma_t(T,K)\sqrt{T-t})$, it suffices to show that we recover the correct martingale representation for $C_t(T,K)$ if we assume that $\Sigma_t(T,K)\sqrt{T-t}$ has the semimartingale decomposition stated in the proposition. This is a simple application of Itô's formula. $\square$

3 The spot volatility dynamics in terms of implied volatilities

In this section, we prove our main results, namely, Theorems 3.1 and 3.2.

3.1 Assumptions

In order to perform our asymptotic analysis of the implied volatility surface, we need to make a few assumptions on the implied volatility surface Σ_t and on its volatility vector $\boldsymbol{\xi}_t$. Throughout ∂_T and ∂_K will denote partial derivatives with respect to T and K.

Assumption 4 For every t and a.s.,

$$\begin{aligned}
&\forall K > 0, \quad T \mapsto \Sigma_t(T, K) \quad \text{is } \mathcal{C}^1 \text{ on } (t, +\infty),\\
&\forall T > t, \quad K \mapsto \partial_T \Sigma_t(T, K) \quad \text{is } \mathcal{C}^3 \text{ on } (0, +\infty),\\
&\forall T > t, \quad K \mapsto \Sigma_t(T, K) \quad \text{is } \mathcal{C}^4 \text{ on } (0, +\infty),\\
&\forall T > t, \quad K \mapsto \boldsymbol{\xi}_t(T, K) \quad \text{is } \mathcal{C}^4 \text{ on } (0, +\infty).
\end{aligned}$$

This last assumption is a mild one. It guarantees smoothness of the implied volatility surface in the open set $(t, +\infty) \times (0, +\infty)$. As we shall explain in Sect. 4, it holds in any stochastic volatility model with smooth coefficients. However it says nothing about the behavior of the surface as $T \to t$. For that, we need the following two assumptions.

Assumption 5 For every K and t, the families of random variables (indexed by θ)

$$\Sigma_t(t+\theta, K), \qquad \partial_T \Sigma_t(t+\theta, K), \qquad \partial_K \Sigma_t(t+\theta, K), \quad \text{and} \quad \partial_K^2 \Sigma_t(t+\theta, K)$$

have a limit in probability as θ decreases to 0. The limits are denoted by

$$\Sigma_t(t, K), \qquad \partial_T \Sigma_t(t, K), \qquad \partial_K \Sigma_t(t, K), \quad \text{and} \quad \partial_K^2 \Sigma_t(t, K).$$

Similarly, $\boldsymbol{\xi}_t(t+\theta, K)$, $\partial_K \boldsymbol{\xi}_t(t+\theta, K)$, and $\partial_K^2 \boldsymbol{\xi}_t(t+\theta, K)$ have a limit in probability as θ decreases to 0, with limits denoted by $\boldsymbol{\xi}_t(t, K)$, $\partial_K \boldsymbol{\xi}_t(t, K)$, and $\partial_K^2 \boldsymbol{\xi}_t(t, K)$.

This last assumption gives us existence of limits for the implied volatility surface and its volatility vector.

Assumption 6 For every compact set $\mathcal{K} \subset (0, +\infty)$ and t, there exists a $\bar{\theta} > 0$ such that for every $p \geq 0$,

$$\int_0^t \sup_{K \in \mathcal{K}} \sup_{0 < \theta \leq \bar{\theta}} \left[\sum_{i=0}^4 \left| \partial_K^i \Sigma_s \right|^p + \sum_{i=0}^3 \left| \partial_T \partial_K^i \Sigma_s \right|^p \right] (s+\theta, K) \, \mathrm{d}s < +\infty \quad \text{a.s.}$$

and

$$\int_0^t \sup_{K \in \mathcal{K}} \sup_{0 < \theta \leq \bar{\theta}} \left[\sum_{i=0}^4 \left| \partial_K^i \boldsymbol{\xi}_s(s+\theta, K) \right|^p \right] \mathrm{d}s < +\infty \quad \text{a.s.}$$

This assumption gives us convenient integrability conditions, so that we shall be able to easily interchange limits and Lebesgue or Itô integrals. We actually only need these integrability conditions for $p \leq 10$. We shall show in Sect. 4 that all the above assumptions hold for a wide class of stochastic volatility models.

3.2 The relation between implied volatilities and their volatility vectors at short maturities

The goal of this subsection is to prove the following result relating implied volatilities and their volatility vectors at short maturities.

Theorem 3.1 *For every K and almost every t,*

$$\Sigma_t(t,K) = \left|\boldsymbol{\sigma}_t - \ln\left(\frac{S_t}{K}\right)\boldsymbol{\xi}_t(t,K)\right| \quad a.s. \tag{3.1}$$

Proof The first step towards this theorem is to use Itô's formula to get the semimartingale decomposition of the processes $\Sigma_t(t+\theta,K)$. This is possible since $\Sigma_t(T,K)$ is $\mathcal{C}^1$ in T (Assumption 4). We obtain that for each $\theta>0$,

$$\begin{aligned}\Sigma_t(t+\theta,K)^2\theta = \Sigma_0(\theta,K)^2\theta + \int_0^t \Big[& 2\theta\,\Sigma_s\partial_T\Sigma_s + \Sigma_s^2 - \theta\,\Sigma_s^2\boldsymbol{\sigma}_s\cdot\boldsymbol{\xi}_s \\ & - \left|\boldsymbol{\sigma}_s - \ln(S_s/K)\boldsymbol{\xi}_s\right|^2 + \frac{\theta^2}{4}\Sigma_s^4|\boldsymbol{\xi}_s|^2 + \theta\,\Sigma_s^2|\boldsymbol{\xi}_s|^2\Big](s+\theta,K)\,\mathrm{d}s \\ & + \int_0^t 2\theta\,\Sigma_s^2\boldsymbol{\xi}_s(s+\theta,K)\cdot\mathrm{d}\boldsymbol{W}_s.\end{aligned}$$

The main idea of the proof is to take the limit $\theta\to 0$ in the previous equation. We are going to see that most terms will have limit 0 in probability. Let us fix K. From Proposition 2.2 we have

$$\mathbb{P}\text{-}\lim_{\theta\to 0}\Sigma_t(t+\theta,K)^2\theta = \mathbb{P}\text{-}\lim_{\theta\to 0}\Sigma_0(\theta,K)^2\theta = 0$$

uniformly on compact time intervals. Here and throughout the rest of the paper, the notation $\mathbb{P}$-lim will denote convergence in probability. On the other hand, by Assumption 6 and the Cauchy–Schwarz inequality,

$$\mathbb{P}\text{-}\lim_{\theta\to 0}\int_0^t \left(2\theta\,\Sigma_s^2\boldsymbol{\xi}_s(s+\theta,K)\right)^2\mathrm{d}s = 0,$$

which yields, using Theorem 2.2.15 of [20], that

$$\mathbb{P}\text{-}\lim_{\theta\to 0}\int_0^t 2\theta\,\Sigma_s^2\boldsymbol{\xi}_s(s+\theta,K)\cdot\mathrm{d}\boldsymbol{W}_s = 0$$

uniformly on compacts in time.

Similarly, repeated use of Assumption 6 and the Cauchy–Schwarz inequality yields that

$$\mathbb{P}\text{-}\lim_{\theta\to 0}\int_0^t\left[2\theta\,\Sigma_s\partial_T\Sigma_s + \frac{\theta^2}{4}\Sigma_s^4|\boldsymbol{\xi}_s|^2 + \theta\,\Sigma_s^2|\boldsymbol{\xi}_s|^2\right](s+\theta,K)\,\mathrm{d}s = 0$$

uniformly on compacts in time. We finally get

$$\mathbb{P}\text{-}\lim_{\theta\to 0}\int_0^t \left[\Sigma_s(s+\theta,K)-\left|\boldsymbol{\sigma}_s-\ln(S_s/K)\boldsymbol{\xi}_s(s+\theta,K)\right|^2\right]\mathrm{d}s=0.$$

Using Assumption 5, Assumption 6, and the dominated convergence theorem, we get

$$\int_0^t \left[\Sigma_s(s,K)-\left|\boldsymbol{\sigma}_s-\ln(S_s/K)\boldsymbol{\xi}_s(s,K)\right|^2\right]\mathrm{d}s=0.$$

The result follows since t is arbitrary. □

The statement of Theorem 3.1 is similar to the "no-bubble restriction" of [24]. The main difference is that in [24], only one option with maturity T is considered, and the limit is taken as t approaches T. Our point of view is different: today's date t is fixed, and we look at a continuum of option prices and at the limit as T approaches t.

The case where we let $K=S_t$ in Theorem 3.1 is particularly important. It says that

$$\Sigma_t(t,S_t)=|\boldsymbol{\sigma}_t|. \tag{3.2}$$

In other words, the current value of the spot volatility can be exactly recovered from the implied volatility smile. It is given by the short maturity limit of at-the-money implied volatilities. This relation is called "feedback condition" in [4] and was known to hold for some time (see, for instance, [7]).

Completely independently of the present paper, [14] shows that (3.2) holds in great generality even when jumps in the spot and/or the volatility are present.

Let us now draw parallels between our result and the important results obtained in [2]. In that paper, the authors show that in a time-homogeneous stochastic volatility model, the implied volatility in the short maturity limit can be expressed using the geodesic distance associated with the generator of the bivariate diffusion (x_t,y_t), where x_t is the log-moneyness, and y_t is the spot volatility ($|\boldsymbol{\sigma}_t|$ in our notation). Keeping their notation, we denote by $d(x,y)$ the geodesic distance from (x,y) to $(0,y)$, that is,

$$d(x,y)=\inf_{\substack{\gamma(0)=(x,y)\\ \gamma(1)=(0,y)}}\int_0^1\sqrt{\sum_{i,j=1}^{2} g_{ij}\dot{\gamma}_i(t)\dot{\gamma}_j(t)}\,\mathrm{d}t,$$

where g_{ij} is the ijth element of the inverse of the diffusion matrix of the bivariate diffusion (x_t,y_t), and γ are continuously differentiable paths starting from (x,y) and ending at $(0,y)$. $\dot{\gamma}$ denotes their derivatives with respect to their parameter. The result of [2] is that

$$\Sigma_t(t,K)=\frac{|\ln(S_t/K)|}{d(\ln(S_t/K),|\boldsymbol{\sigma}_t|)}. \tag{3.3}$$

It is now very tempting to compare (3.1) and (3.3) to relate the geodesic distance to the implied volatility surface's volatility vector $\boldsymbol{\xi}_t(T,K)$. The relation actually is

$$d\big(x,|\boldsymbol{\sigma}_t|\big)=\frac{|x|}{|\boldsymbol{\sigma}_t-x\boldsymbol{\xi}_t(t,S_te^{-x})|}.$$

As (3.3) indicates, $\Sigma_t(t,K)$ has the interpretation of the ratio between the usual Euclidean and geodesic distances between $(\ln(S_t/K),\sigma_t)$ and $(0,\sigma_t)$. Our theorem gives another interpretation of $\Sigma_t(t,K)$, namely

$$\Sigma_t(t,K)=\left|\boldsymbol{\sigma}_t-\ln\left(\frac{S_t}{K}\right)\boldsymbol{\xi}_t(t,K)\right|.$$

This is the Euclidean norm of the difference of the spot volatility vector and the implied volatility's volatility vector multiplied by log-moneyness.

3.3 The spot volatility's semimartingale decomposition in terms of implied volatilities

In this subsection we prove the following theorem, which is the main result of the paper. It gives a semimartingale decomposition of the spot volatility in terms of implied volatilities.

Theorem 3.2 *There exists a scalar Wiener process $W^\perp$ on $(\Omega,\mathcal{H},\mathbb{P})$ adapted to $(\mathcal{F}_t)_{t\geq 0}$ such that for every t,*

$$\begin{aligned}|\boldsymbol{\sigma}_t|^2&=|\boldsymbol{\sigma}_0|^2\\&\quad+\int_0^t\left[4|\boldsymbol{\sigma}_s|\frac{\partial\Sigma_s}{\partial T}(s,S_s)+6|\boldsymbol{\sigma}_s|^2\left(S_s\frac{\partial\Sigma_s}{\partial K}(s,S_s)\right)^2\right.\\&\quad\left.+2|\boldsymbol{\sigma}_s|^3S_s^2\frac{\partial^2\Sigma_s}{\partial K^2}(s,S_s)\right]\mathrm{d}s\\&\quad+\int_0^t4|\boldsymbol{\sigma}_s|\frac{\partial\Sigma_s}{\partial K}(s,S_s)\,\mathrm{d}S_s+\int_0^t2|\boldsymbol{\sigma}_s|^2\xi_s^\perp\,\mathrm{d}W_s^\perp,\end{aligned}$$

where

$$\begin{aligned}\left(\xi_t^\perp\right)^2&=-\frac{2}{|\boldsymbol{\sigma}_t|}\frac{\mathrm{d}}{\mathrm{d}t}\left\langle S_t,\frac{\partial\Sigma_t}{\partial K}(t,S_t)\right\rangle+2\left(S_t\frac{\partial\Sigma_t}{\partial K}(t,S_t)\right)^2-|\boldsymbol{\sigma}_t|S_t\frac{\partial\Sigma_t}{\partial K}(t,S_t)\\&\quad-3|\boldsymbol{\sigma}_t|S_t^2\frac{\partial^2\Sigma_t}{\partial K^2}(t,S_t).\end{aligned}$$

Moreover, the two local martingales appearing in the decomposition are orthogonal in the sense that

$$\left\langle\int_0^t4|\boldsymbol{\sigma}_s|\frac{\partial\Sigma_s}{\partial K}(s,S_s)\,\mathrm{d}S_s,\int_0^t2|\boldsymbol{\sigma}_s|^2\xi_s^\perp\,\mathrm{d}W_s^\perp\right\rangle=0.$$

The proof is done in several steps.

Proposition 3.3 *For each $\theta>0$, the family of processes $\Sigma_t(t+\theta,K)$ is a $\mathcal{C}^3$ semimartingale in the sense of* [20].

Proof To check that the family of processes $\Sigma_t(t+\theta, K)$ is a $\mathcal{C}^3$ semimartingale (see the definition on pp. 79 and 84 in [20]), we introduce the local characteristics (see the definition on p. 84 in [20])

$$b(s,K) = \Big[2\theta\,\Sigma_s\partial_T\Sigma_s + \Sigma_s^2 - \theta\,\Sigma_s^2\boldsymbol{\sigma}_s\cdot\boldsymbol{\xi}_s - \big|\boldsymbol{\sigma}_s - \ln(S_s/K)\boldsymbol{\xi}_s\big|^2 \\ + \frac{\theta^2}{4}\Sigma_s^4|\boldsymbol{\xi}_s|^2 + \theta\,\Sigma_s^2|\boldsymbol{\xi}_s|^2\Big](s+\theta, K)$$

and

$$a(s,K,L) = 4\theta^2\Sigma_s^2(s+\theta,K)\Sigma_s^2(s+\theta,L)\boldsymbol{\xi}_s(s+\theta,K)\cdot\boldsymbol{\xi}_s(s+\theta,L).$$

In order for the bounded variation process $(\int_0^t b(s,K)\,\mathrm{d}s)$ to be a $\mathcal{C}^3$ process of bounded variation, it is enough to check that, for every t and compact set $\mathcal{K}\subset(0,+\infty)$,

$$\int_0^t \sup_{K\in\mathcal{K}}\sum_{i=0}^3\big|\partial_K^i b(s,K)\big|\,\mathrm{d}s < +\infty.$$

Our Assumption 6 guarantees that this integral is indeed finite.

In order for the local martingale part to be a $\mathcal{C}^3$ local martingale, it is enough to check that, for every t and compact set $\mathcal{K}\subset(0,+\infty)$,

$$\int_0^t \sup_{K,L\in\mathcal{K}}\sum_{i=0}^4\big|\partial_K^i\partial_L^i a(t,K,L)\big|\,\mathrm{d}s < +\infty.$$

Indeed, the joint quadratic variation $\int_0^t a(s,K,L)\,\mathrm{d}s$ will have a modification as a continuous $\mathcal{C}^4$ process, and the local martingale part will have a modification as a continuous $\mathcal{C}^3$ local martingale thanks to Theorem 3.1.2 in [20]. Now, the integrability condition above follows right away from our Assumption 6. □

Proposition 3.4 *For each $\theta > 0$, $\Sigma_t(t+\theta, S_t)^2$ is a semimartingale whose decomposition is given by*

$$\Sigma_t(t+\theta,S_t)^2 = \Sigma_0(\theta,S_0)^2 + \int_0^t\Big[2\Sigma_s\partial_T\Sigma_s + \frac{1}{\theta}\big(\Sigma_s^2 - |\boldsymbol{\sigma}_s|^2\big) - \Sigma_s^2\boldsymbol{\sigma}_s\cdot\boldsymbol{\xi}_s \\ + \frac{\theta}{4}\Sigma_s^4|\boldsymbol{\xi}_s|^2 + \Sigma_s^2|\boldsymbol{\xi}_s|^2 + S_s^2|\boldsymbol{\sigma}_s|^2\big(\Sigma_s\partial_K^2\Sigma_s + (\partial_K\Sigma_s)^2\big) \\ + 4S_s\Sigma_s\partial_K\Sigma_s\boldsymbol{\sigma}_s\cdot\boldsymbol{\xi}_s + 2S_s\Sigma_s^2\boldsymbol{\sigma}_s\cdot\partial_K\boldsymbol{\xi}_s\Big](s+\theta,S_s)\,\mathrm{d}s \\ + \int_0^t\big[2\Sigma_s^2\boldsymbol{\xi}_s + 2S_s\Sigma_s\partial_K\Sigma_s\boldsymbol{\sigma}_s\big](s+\theta,S_s)\cdot\mathrm{d}\boldsymbol{W}_s.$$

Proof The proposition is simply an application of the Itô–Wentzell formula (Theorem 3.3.1 in [20]) to the family of processes $\Sigma_t(t+\theta, K)^2$. The assumptions of that theorem are satisfied: $\Sigma_t(t+\theta, K)^2$ is both a $\mathcal{C}^2$ process and a $\mathcal{C}^1$ semimartingale thanks to Proposition 3.3. The local characteristics have the required integrability as can be seen from the proof of Proposition 3.3. □

Proposition 3.5 *$|\boldsymbol{\sigma}_t|^2$ is a semimartingale whose decomposition is given by*

$$|\boldsymbol{\sigma}_t|^2 = |\boldsymbol{\sigma}_0|^2 + \int_0^t \big[4\Sigma_s \partial_T \Sigma_s - \Sigma_s^2 \boldsymbol{\sigma}_s \cdot \boldsymbol{\xi}_s + \Sigma_s^2 |\boldsymbol{\xi}_s|^2 + S_s^2 |\boldsymbol{\sigma}_s|^2 \Sigma_s \partial_K^2 \Sigma_s$$
$$+ \big(|\boldsymbol{\sigma}_s| S_s \partial_K \Sigma_s\big)^2 + 4 S_s \Sigma_s \partial_K \Sigma_s \boldsymbol{\sigma}_s \cdot \boldsymbol{\xi}_s + 2 S_s \Sigma_s^2 \boldsymbol{\sigma}_s \cdot \partial_K \boldsymbol{\xi}_s\big](s, S_s)\,\mathrm{d}s$$
$$+ \int_0^t \big[2\Sigma_s^2 \boldsymbol{\xi}_s + 2 S_s \Sigma_s \partial_K \Sigma_s \boldsymbol{\sigma}_s\big](s, S_s) \cdot \mathrm{d}\boldsymbol{W}_s.$$

Proof We now simply take the limit as θ goes to 0 in the semimartingale decomposition of Proposition 3.4. Fix t and ω. First of all, Assumption 5 and (3.2) imply that

$$\lim_{\theta \to 0} \Sigma_t(t+\theta, S_t)^2 = |\boldsymbol{\sigma}_t|^2,$$

and similarly with $t = 0$.

In the bounded-variation part,

$$\int_0^t \frac{\Sigma_s^2(s+\theta, S_s) - |\boldsymbol{\sigma}_s|^2}{\theta}\,\mathrm{d}s = \int_0^t 2[\Sigma_s \partial_T \Sigma_s](s+\theta_s', S_s)\,\mathrm{d}s$$

for some $0 < \theta_s' < \theta$ by the mean value theorem. Each term of the bounded variation part converges by dominated convergence thanks to Assumptions 5 and 6. Note that S is a continuous process and that, thanks to Assumptions 5 and 6,

$$\int_0^t |\boldsymbol{\sigma}_s|^p\,\mathrm{d}s < +\infty$$

for every $p \geq 0$ and $t \geq 0$.

For the local martingale part, we have, thanks to Assumptions 5 and 6, that

$$\mathbb{P}\text{-}\lim_{\theta \to 0} \int_0^t \big|\big[2\Sigma_s^2 \boldsymbol{\xi}_s + 2 S_s \Sigma_s \partial_K \Sigma_s \boldsymbol{\sigma}_s\big](s+\theta, S_s)$$
$$- \big[2\Sigma_s^2 \boldsymbol{\xi}_s + 2 S_s \Sigma_s \partial_K \Sigma_s \boldsymbol{\sigma}_s\big](s, S_s)\big|^2\,\mathrm{d}s = 0.$$

Thanks to Theorem 2.2.15 in [20], this implies that

$$\mathbb{P}\text{-}\lim_{\theta \to 0} \int_0^t \big[2\Sigma_s^2 \boldsymbol{\xi}_s + 2 S_s \Sigma_s \partial_K \Sigma_s \boldsymbol{\sigma}_s\big](s+\theta, S_s) \cdot \mathrm{d}\boldsymbol{W}_s$$
$$= \int_0^t \big[2\Sigma_s^2 \boldsymbol{\xi}_s + 2 S_s \Sigma_s \partial_K \Sigma_s \boldsymbol{\sigma}_s\big](s, S_s) \cdot \mathrm{d}\boldsymbol{W}_s.$$

We finally obtain the semimartingale decomposition stated in the proposition. □

Proposition 3.5 gives the semimartingale decomposition of $|\boldsymbol{\sigma}_t|$. The remaining task is now to compute $\boldsymbol{\sigma}_t \cdot \boldsymbol{\xi}_t(t, S_t)$ and $\boldsymbol{\sigma}_t \cdot \partial_K \boldsymbol{\xi}_t(t, S_t)$ in terms of implied volatilities. This will be done in several steps using Theorem 3.1. We are using it to give us a Taylor expansion of $\boldsymbol{\xi}_t(t, K)$ in K around S_t. To that end, we define the family of processes $\boldsymbol{\xi}_t^{\perp}(t+\theta, K)$ by

$$\boldsymbol{\xi}_t^{\perp}(t+\theta, K) = \boldsymbol{\xi}_t(t+\theta, K) - \frac{\boldsymbol{\sigma}_t \cdot \boldsymbol{\xi}_t(t+\theta, K)}{|\boldsymbol{\sigma}_t|^2}\boldsymbol{\sigma}_t. \tag{3.4}$$

The second term on the right-hand side is simply the orthogonal projection of $\boldsymbol{\xi}_t(t+\theta, K)$ onto $\boldsymbol{\sigma}_t$.

Proposition 3.6 *For almost every t,*

$$\boldsymbol{\sigma}_t \cdot \boldsymbol{\xi}_t(t, S_t) = |\boldsymbol{\sigma}_t| S_t \partial_K \Sigma_t(t, S_t),$$

$$S_t \boldsymbol{\sigma}_t \cdot \partial_K \boldsymbol{\xi}_t(t, S_t) = \frac{1}{2}\Big[|\boldsymbol{\sigma}_t|\big(S_t \partial_K \Sigma_t(t, S_t) + S_t^2 \partial_K^2 \Sigma_t(t, S_t)\big) - \big|\boldsymbol{\xi}_t^{\perp}(t, S_t)\big|^2\Big].$$

Proof The statement of Theorem 3.1 can be rewritten as

$$\Sigma_t(t, K) = \left|\left(1 - \frac{\boldsymbol{\sigma}_t \cdot \boldsymbol{\xi}_t(t, K)}{|\boldsymbol{\sigma}_t|^2}\ln\left(\frac{S_t}{K}\right)\right)\boldsymbol{\sigma}_t - \ln\left(\frac{S_t}{K}\right)\boldsymbol{\xi}_t^{\perp}(t, K)\right|,$$

which can be inverted as

$$\boldsymbol{\sigma}_t \cdot \boldsymbol{\xi}_t(t, K) = \frac{|\boldsymbol{\sigma}_t|}{\ln(S_t/K)}\left(|\boldsymbol{\sigma}_t| - \sqrt{\Sigma_t(t, K)^2 - \ln^2(S_t/K)\big|\boldsymbol{\xi}_t^{\perp}(t, K)\big|^2}\right).$$

Thanks to Assumption 6, the functions $K \mapsto \Sigma_t(t, K)$ and $K \mapsto \boldsymbol{\xi}_t(t, K)$ are $\mathcal{C}^2$ and $\mathcal{C}^1$, respectively, and their derivatives in K are equal to the limits of the derivatives as $\theta \to 0$. It simply remains to compute the value and the first derivative of $\boldsymbol{\sigma}_t \cdot \boldsymbol{\xi}_t(t, K)$ at S_t by expanding the right-hand side of the above equation. □

Proposition 3.6 gives us the crucial quantity $\boldsymbol{\sigma}_t \cdot \partial_K \boldsymbol{\xi}_t(t, S_t)$ in terms of $|\boldsymbol{\xi}_t^{\perp}(t, S_t)|$. However, we still need to relate $|\boldsymbol{\xi}_t^{\perp}(t, S_t)|$ to the implied volatilities. $|\boldsymbol{\xi}_t^{\perp}(t, S_t)|$ will depend on the joint quadratic variation of S and the at-the-money skew $\partial_K \Sigma_t(t, S_t)$. We first compute this joint quadratic variation in the next proposition.

Proposition 3.7 *For every $\theta > 0$, $\partial_K \Sigma_t^2(t+\theta, S_t)$ is a semimartingale whose decomposition is given by*

$$\begin{aligned}
\partial_K \Sigma_t^2(t+\theta, S_t) = \partial_K \Sigma_0^2(\theta, S_0) &+ \int_0^t \partial_K \Bigg[2\Sigma_s \partial_T \Sigma_s - \Sigma_s^2 \boldsymbol{\sigma}_s \cdot \boldsymbol{\xi}_s \\
&+ \frac{1}{\theta}\big(\Sigma_s^2 - |\boldsymbol{\sigma}_s - \ln(S_s/K)\boldsymbol{\xi}_s|^2\big) + \frac{\theta}{4}\Sigma_s^4|\boldsymbol{\xi}_s|^2 + \Sigma_s^2|\boldsymbol{\xi}_s|^2 \\
&+ \frac{|\boldsymbol{\sigma}_s|^2}{2} S_s^2 \partial_K^2 \Sigma_s^2 + 2S_s \partial_K\big[\Sigma_s^2 \boldsymbol{\xi}_s\big] \cdot \boldsymbol{\sigma}_s\Bigg](s+\theta, S_s)\,\mathrm{d}s \\
&+ \int_0^t \big[S_s \partial_K^2 \Sigma_s^2 \boldsymbol{\sigma}_s + 2\partial_K\big(\Sigma_s^2 \boldsymbol{\xi}_s\big)\big](s+\theta, S_s) \cdot \mathrm{d}\boldsymbol{W}_s.
\end{aligned}$$

Proof Proposition 3.3 gives us that $\Sigma_t(t+\theta, K)$ is a $\mathcal{C}^3$ semimartingale for each $\theta > 0$. It follows from the definition of $\mathcal{C}^3$ semimartingales that $\partial_K \Sigma_t(t+\theta, K)^2$ is a $\mathcal{C}^2$ semimartingale whose semimartingale decomposition is given by interchanging derivatives with respect to K and integrals. More precisely,

$$\begin{aligned}\partial_K \Sigma_t^2(t+\theta, K) = \partial_K \Sigma_0^2(\theta, K) + \int_0^t \partial_K \Big[& 2\Sigma_s \partial_T \Sigma_s - \Sigma_s^2 \boldsymbol{\sigma}_s \cdot \boldsymbol{\xi}_s \\ & + \frac{1}{\theta}\big(\Sigma_s^2 - \big|\boldsymbol{\sigma}_s - \ln(S_s/K)\boldsymbol{\xi}_s\big|^2\big) \\ & + \frac{\theta}{4}\Sigma_s^4 |\boldsymbol{\xi}_s|^2 + \Sigma_s^2 |\boldsymbol{\xi}_s|^2 \Big](s+\theta, K)\,\mathrm{d}s \\ & + \int_0^t 2\partial_K \big[\Sigma_s^2 \boldsymbol{\xi}_s\big](s+\theta, K) \cdot \mathrm{d}\boldsymbol{W}_s.\end{aligned}$$

Exactly as in Proposition 3.4, we apply the Itô–Wentzell formula to get the semimartingale decomposition of $\partial_K \Sigma_t^2(t+\theta, S_t)$. The assumptions needed to apply Theorem 3.3.1 of [20] are satisfied thanks to Proposition 3.3. □

We now take the limit as $\theta \to 0$ in the previous proposition. It will lead to a new relation involving $\boldsymbol{\sigma}_t \cdot \boldsymbol{\xi}_t(t, S_t)$ and $\boldsymbol{\sigma}_t \cdot \partial_K \boldsymbol{\xi}_t(t, S_t)$, from which we find $|\boldsymbol{\xi}_t^\perp(t, S_t)|$.

Proposition 3.8 *For almost every t,*

$$\begin{aligned}\big|\boldsymbol{\xi}_t^\perp(t, S_t)\big|^2 = & -\frac{2}{|\boldsymbol{\sigma}_t|}\frac{\mathrm{d}}{\mathrm{d}t}\big\langle S_t, \partial_K \Sigma_t(t, S_t)\big\rangle + 2\big(S_t \partial_K \Sigma_t(t, S_t)\big)^2 \\ & - |\boldsymbol{\sigma}_t| S_t \partial_K \Sigma_t(t, S_t) - 3|\boldsymbol{\sigma}_t| S_t^2 \partial_K^2 \Sigma_t(t, S_t).\end{aligned}$$

Proof Exactly as in Proposition 3.5, we prove that the bounded variation part and the martingale part of the semimartingale decomposition of $\partial_K \Sigma_t^2(t+\theta, S_t)$ have limits as $\theta \to 0$. The limits can be computed as before. The joint quadratic variation of the limit and S is given by

$$\begin{aligned}&\big\langle \partial_K \Sigma_t(t, S_t), S_t \big\rangle \\ &\quad = \int_0^t \big[S_s \partial_K \Sigma_s \boldsymbol{\sigma}_s \cdot \boldsymbol{\xi}_s + |\boldsymbol{\sigma}_s| S_s \boldsymbol{\sigma}_s \cdot \partial_K \boldsymbol{\xi}_s + |\boldsymbol{\sigma}_s|^2 S_s^2 \partial_K^2 \Sigma_s\big](s, S_s)\,\mathrm{d}s.\end{aligned}$$

It remains to express $\boldsymbol{\sigma}_t \cdot \boldsymbol{\xi}_t(t, S_t)$ and $\boldsymbol{\sigma}_t \cdot \partial_K \boldsymbol{\xi}_t(t, S_t)$ using Proposition 3.6 and to differentiate with respect to t. □

We can now complete the proof of Theorem 3.2. We simply replace $\boldsymbol{\sigma}_t \cdot \boldsymbol{\xi}_t(t, S_t)$, $\boldsymbol{\sigma}_t \cdot \partial_K \boldsymbol{\xi}_t(t, S_t)$ with the expressions we have computed in Propositions 3.6 and 3.8 and use the fact that (see (3.4))

$$\big|\boldsymbol{\xi}_t(t, S_t)\big|^2 = \big|\boldsymbol{\xi}_t^\perp(t, S_t)\big|^2 + \big(\boldsymbol{\sigma}_t \cdot \boldsymbol{\xi}_t(t, S_t)\big)^2.$$

This yields

$$|\boldsymbol{\sigma}_t|^2 = |\boldsymbol{\sigma}_0|^2 + \int_0^t \big[4|\boldsymbol{\sigma}_s|\partial_T \Sigma_s(s,S_s) + 6|\boldsymbol{\sigma}_s|^2\big(S_s\partial_K \Sigma_s(s,S_s)\big)^2$$

$$+ 2|\boldsymbol{\sigma}_s|^3 S_s^2\partial_K^2 \Sigma_s(s,S_s)\big]\,\mathrm{d}s + \int_0^t \big[2\Sigma_s^2 \boldsymbol{\xi}_s^{\perp} + 4S_s\Sigma_s\partial_K \Sigma_s \boldsymbol{\sigma}_s\big](s,S_s)\cdot \mathrm{d}\boldsymbol{W}_s.$$

Now, thanks to Lévy's theorem, we define the new Wiener process $W^{\perp}$ as

$$W_t^{\perp} = \int_0^t \left[\frac{\boldsymbol{\xi}_t^{\perp}(t,S_t)}{|\boldsymbol{\xi}_t^{\perp}(t,S_t)|}\mathbf{1}_{\{|\boldsymbol{\xi}_t^{\perp}(t,S_t)|\neq 0\}} + \frac{\boldsymbol{\sigma}_s}{|\boldsymbol{\sigma}_s|}\mathbf{1}_{\{|\boldsymbol{\xi}_t^{\perp}(t,S_t)|=0\}}\right]\cdot \mathrm{d}\boldsymbol{W}_s.$$

This completes the proof of the theorem. □

4 The special case of stochastic volatility models

The goal of this section is to prove Proposition 4.2, which will give us sufficient conditions for Assumptions 1–6 to hold. This section makes use of [1] and [2]. We are going first to recall the precise statement of their result.

As in [2], we focus here on stochastic volatility models. More precisely, we consider a stock price process which follows a stochastic differential equation of the type

$$\frac{\mathrm{d}S_t}{S_t} = \sigma(t,S_t,y_t)\,\mathrm{d}W_t,$$
$$\mathrm{d}y_t = \gamma(t,y_t)\,\mathrm{d}t + \nu(t,y_t)\,\mathrm{d}Z_t,$$

where $W = Z^0$, $Z = (Z^1,\dots,Z^{n-1})$ are standard Wiener processes. We define the correlation matrix $\Omega = (\rho_{ij})_{1\le i,j\le n-1}$ by $\langle Z_t^i, Z_t^j\rangle = \rho_{ij}t$. γ is the vector of drift coefficients, and ν is the diffusion matrix. y_t takes values in $\mathbb{R}^{n-1}$, and we assume that $S_t \in (0,+\infty)$.

If $(\mathcal{F}_t)_{t\ge 0}$ is the filtration generated by $(Z^0,\dots,Z^{n-1})$, then the value of a European call option with strike K and maturity T is given by the conditional expectation

$$C(t,S_t,y_t) = \mathbb{E}\big\{(S_T-K)^+\big|\mathcal{F}_t\big\}.$$

We introduce the new variables $\theta = T-t$, $x_1 = \ln(S/K)$, and $x_i = y_{i-1}$ for $i\ge 2$. In such models, the implied volatility is a deterministic function of time and the factors, i.e.,

$$\Sigma_t(T,K) = \tilde{\Sigma}(t,S_t,y_t,T,K).$$

We verify that Assumptions 1–6 hold when the implied volatility has this form.

We introduce the following notation. The correlation matrix is now $\Omega = (\omega_{ij})$ with $\omega_{ij} = \rho_{i-1,j-1}$, and i,j now run from 1 to n. The diffusion matrix is given by $M = (m_{ij})$ with $m_{11} = \sigma$, $m_{1k} = m_{k1} = 0$ if $k\neq 1$, and $m_{ij} = \nu_{i-1,j-1}$ if $2\le i,j\le n$. The drifts are defined as $q_1 = -\frac{1}{2}\sigma^2(t,S,y)$, $q_i = \gamma_{i-1}$ for $i = 2,\dots,n$. We also define $\tilde{\gamma}_1 = 0$ and $\tilde{\gamma}_i = q_i + \omega_{1i}\sigma\nu_{i-1}$.

Note that $\boldsymbol{q}$, $\tilde{\boldsymbol{\gamma}}$ are vectors and functions of $(\theta, \boldsymbol{x})$ whereas M is a matrix and also a function of $(\theta, \boldsymbol{x})$. Let us make the following assumption:

Assumption 7 As functions of $(\theta, \boldsymbol{x}) \in (0, T) \times \mathbb{R}^n$, each component of the vector $\boldsymbol{q}$ and the matrix $M\Omega M^\top$ are $\mathcal{C}^\infty$. Also, there exists a constant C such that for all $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^n$ and all $\theta \in (0, T)$,

$$C^{-1}\big(1+|\boldsymbol{x}|\big)^{-2}|\boldsymbol{v}|^2 \le \big(M\Omega M^\top(\theta, \boldsymbol{x})\boldsymbol{v}\big) \cdot \boldsymbol{v} \le C\big(1+|\boldsymbol{x}|\big)^2|\boldsymbol{v}|^2.$$

Here and throughout the rest of the paper, we denote by ∂_θ the differentiation with respect to θ and by ∇ the gradient with respect to the variable $\boldsymbol{x}$. ∇^2 will denote the Hessian matrix, Tr the trace, and $^\top$ transpose. Finally, for any two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, $\boldsymbol{a} \otimes \boldsymbol{b}$ is the matrix whose ijth entry is $a_i b_j$.

With this notation, we are in a position to state the result taken from Proposition 1.2 in [2].

Proposition 4.1 *Under Assumption 7, the implied volatility function is the unique solution* $\tilde{\Sigma}(\theta, \boldsymbol{x}) \in W^{1,2,\infty}_{\mathrm{loc}}((0, T) \times \mathbb{R}^n)$ *of the well-posed nonlinear degenerate parabolic initial value problem*

$$\partial_\theta\big(\theta\tilde{\Sigma}^2\big) = H\big(\theta, \boldsymbol{x}, \tilde{\Sigma}, \nabla\tilde{\Sigma}, \nabla^2\tilde{\Sigma}\big) \tag{4.1}$$

with the initial condition $\tilde{\Sigma}(0, \boldsymbol{x}) = \tilde{\Sigma}^0(\boldsymbol{x})$, *where* $\tilde{\Sigma}^0$ *is the unique solution to*

$$H\big(0, \boldsymbol{x}, \tilde{\Sigma}, \nabla\tilde{\Sigma}, \nabla^2\tilde{\Sigma}\big) = 0.$$

The function H *is given by*

$$\begin{aligned} H = {} & \mathrm{Tr}\bigg(M\Omega M^\top\tilde{\Sigma}^2\bigg(\nabla\bigg(\frac{x_1}{\tilde{\Sigma}}\bigg) \otimes \nabla\bigg(\frac{x_1}{\tilde{\Sigma}}\bigg) - \frac{\theta^2}{4}\tilde{\Sigma}^2\nabla\tilde{\Sigma} \otimes \nabla\tilde{\Sigma}\bigg)\bigg) \\ & + \theta\tilde{\Sigma}\gamma \cdot \nabla\tilde{\Sigma} + \theta\tilde{\Sigma}\,\mathrm{Tr}\big(M\Omega M^\top\nabla^2\tilde{\Sigma}\big). \end{aligned}$$

Here $W^{1,2,\infty}_{\mathrm{loc}}((0, T) \times \mathbb{R}^n)$ denotes the Sobolev space of functions that are almost everywhere once differentiable in θ on $(0, T)$ and twice in $\boldsymbol{x}$ on $\mathbb{R}^n$, and whose derivatives are bounded on every compact in $(0, T) \times \mathbb{R}^n$.

This result shows in particular that the function $\tilde{\Sigma}$ is continuous on $[0, T) \times \mathbb{R}^n$. Since

$$\Sigma_t(T, K) = \tilde{\Sigma}\big(T - t, \ln(S_t/K), y_t\big),$$

it is clear that the limit of $\Sigma_t(t + \theta, K)$ as θ goes to 0 exists a.s. for every t and K. This is exactly what is needed to check Assumption 5. Moreover,

$$\int_0^t \sup_{K\in\mathcal{K}} \sup_{0<\theta\le\bar{\theta}} \Sigma_s(s + \theta, K)^p \, \mathrm{d}s < +\infty \quad \text{a.s.},$$

since the integrand is actually a continuous function of time. This is exactly what is needed to check Assumption 6.

In view of this discussion, we now state the following proposition that gives sufficient conditions for Assumptions 1–6 to hold.

Proposition 4.2 *Under Assumption 7, $\tilde{\Sigma}$ is $\mathcal{C}^\infty$ on $(0, T) \times \mathbb{R}^n$. Under Assumption 7 and assuming that the derivatives of $\tilde{\Sigma}$ are continuous on $[0, T) \times \mathbb{R}^n$, Assumptions 1–6 hold true.*

Proof Assumption 3 is Proposition 1.1 in [2].

Assumption 4 is about regularity of the implied volatility surface and its volatility. It follows from the regularity of the law of S_T and of its first variation process. Under Assumption 7, this law has a density that is $\mathcal{C}^\infty$. Details can be found in [13], Sect. 2.6.

As for Assumptions 5 and 6, they follow from the same argument presented before the statement of this proposition applied to the corresponding derivatives. The implied volatility's volatility vector is found by an application of Itô's formula as

$$\boldsymbol{\xi}_t(T, K) = \frac{M(T-t, \ln(S_t/K), y_t)\nabla\tilde{\Sigma}(T-t, \ln(S_t/K), y_t)}{\tilde{\Sigma}(T-t, \ln(S_t/K), y_t)}.$$

Therefore the regularity of the implied volatility's volatility vector follows from that of the implied volatility surface itself. The limits in Assumption 5 actually hold a.s. □

To be complete, we should now look for conditions under which the derivatives of $\tilde{\Sigma}$ are continuous up to the boundary $\{\theta = 0\}$. We do not go into the details here. It should be noted, however, that these derivatives are solutions to parabolic PDEs obtained by differentiating (4.1). These parabolic PDEs are simpler because they are linear. The techniques developed in [2] can be used to prove that these are well-posed problems.

Finally, let us mention that in some concrete examples like the constant elasticity of variance (CEV) model or the Heston model, one can prove the analyticity of the function $\tilde{\Sigma}$, which of course implies the corresponding regularity. Details can be found in [13], Sect. 7.

5 Application: from spot to implied volatilities

There are several interesting applications of Theorem 3.2. The most striking one is that it allows for a converse result. More precisely, we can start from the spot volatility semimartingale decomposition and get an approximation of the implied volatility smile.

Assume that the spot dynamics are given by

$$\begin{aligned} \frac{\mathrm{d}S_t}{S_t} &= \sigma_t \,\mathrm{d}W_t, \\ \mathrm{d}\sigma_t^2 &= \delta_t \,\mathrm{d}t + \nu_t \,\mathrm{d}W_t + \nu_t^\perp \,\mathrm{d}W_t^\perp \end{aligned} \tag{5.1}$$

for some processes δ, ν, and $\nu^{\perp}$. We have chosen a model driven by two independent scalar Brownian motions W and $W^{\perp}$, but one could easily deal with more involved models. To reconcile with our previous notation, $\boldsymbol{\sigma}_t = (\sigma_t, 0)$. The first derivatives of the implied volatility surface are given by the formulas

$$
\begin{aligned}
\Sigma_t(t, S_t) &= \sigma_t, \\
\frac{\partial \Sigma_t}{\partial K}(t, S_t) &= \frac{\nu_t}{4\sigma_t^2 S_t}, \\
\frac{\partial^2 \Sigma_t}{\partial K^2}(t, S_t) &= \frac{1}{2\sigma_t^2 S_t^2}\left[\frac{1}{3\sigma_t^2 S_t}\frac{\mathrm{d}}{\mathrm{d}t}\langle \nu, S\rangle_t + \frac{(\nu_t^{\perp})^2}{3\sigma_t^3} - \frac{\nu_t^2}{4\sigma_t^3} - \frac{\nu_t}{2}\right], \\
\frac{\partial \Sigma_t}{\partial T}(t, S_t) &= \frac{1}{4}\left[\frac{\delta_t}{\sigma_t} - \frac{1}{3\sigma_t^2 S_t}\frac{\mathrm{d}}{\mathrm{d}t}\langle \nu, S\rangle_t - \frac{(\nu_t^{\perp})^2}{3\sigma_t^3} - \frac{5\nu_t^2}{4\sigma_t^3} + \frac{\nu_t}{2}\right].
\end{aligned}
$$

Higher-order derivatives can also be computed (see [13]). The first equation is a consequence of Theorem 3.1, and the other three are obtained by identifying the semi-martingale decomposition of σ_t^2 in (5.1) with that of Theorem 3.2.

Such formulas provide qualitative understanding of stochastic volatility models. Indeed, interesting quantities and dynamics can be found directly from the model equations completely bypassing simulation and numerical methods.

It is interesting to ask how our result relates to the previous result of Dupire [12] in the case of a Markov spot process. Dupire gives a formula for the entire deterministic volatility function $\sigma(t, S)$. Instead, we get its local shape, namely its derivatives in t and S at the current spot value and current time. Expressions for these derivatives at low order can be found in [13]. Of course, they agree with Dupire's formula. In particular our formulas above contain the practitioners' so-called "1/2 slope rule." Indeed, for a Markov spot process, where $\sigma_t = \sigma(t, S_t)$,

$$
\nu_t = 2\sigma(t, S_t)^2 S_t \frac{\partial \sigma}{\partial S}(t, S_t),
$$

and at first order in K,

$$
\Sigma_t(t, K) \simeq \sigma(t, S_t) + (K - S_t)\frac{1}{2}\frac{\partial \sigma}{\partial S}(t, S_t) \simeq \frac{\sigma(t, S_t) + \sigma(t, K)}{2}.
$$

Note also that the formulas above imply a very simple arbitrage lower bound on the volatility of volatility. If we let the volatility of volatility be

$$
\mathrm{Vvol}_t^2 = \frac{1}{\mathrm{d}t}\left\langle \frac{\mathrm{d}\sigma_t}{\sigma_t} \right\rangle,
$$

then

$$
\mathrm{Vvol}_t \geq S_t \left| \frac{\partial \Sigma_t}{\partial K}(t, S_t) \right|.
$$

Further and more elaborate applications of these relationships can be found in [15].

6 Conclusion

The present paper derives the spot volatility dynamics from the implied volatility dynamics. It particular, it shows the role played by the implied volatility surface as the *driver* of the spot volatility process. We tried to develop a new framework where the implied volatility is not an output of a pricing model but rather is the input from which one finds a good model.

The derivation of the semimartingale decomposition of the spot volatility is done through an asymptotic analysis of the implied volatility surface for short maturities. We rely on stochastic analysis on Wiener space techniques, and no Markov assumptions are needed. This allows for great flexibility in applications.

References

1. Berestycki, H., Busca, J., Florent, I.: Asymptotics and calibration of local volatility models. Quant. Finance **2**, 61–69 (2002)
2. Berestycki, H., Busca, J., Florent, I.: Computing the implied volatility in stochastic volatility models. Commun. Pure Appl. Math. **57**, 1352–1373 (2004)
3. Bowie, J., Carr, P.: Static simplicity. Risk **7**(8), 44–50 (1994)
4. Brace, A., Goldys, B., Klebaner, F., Womersley, R.: Market model for stochastic implied volatility with application to the BGM model. Technical Report, Department of Statistics, University of New South Wales (2001). http://web.maths.unsw.edu.au/~sw/Finance/svol.pdf
5. Breeden, D., Litzenberger, R.: Prices of state-contingent claims implicit in option prices. J. Bus. **51**, 621–651 (1978)
6. Carmona, R., Nadtochiy, S.: Local volatility dynamic models. Finance Stoch. **13**, 1–48 (2009)
7. Carr, P.: A survey of preference free option valuation with stochastic volatility. In: Risk's 5th Annual European Derivatives and Risk Management Congress, Paris (2000)
8. Cont, R., Da Fonseca, J.: Deformation of implied volatility surfaces: an empirical analysis. In: Takayasu, H. (ed.) Empirical Science of Financial Fluctuations, pp. 230–239. Springer, Berlin (2001)
9. Delbaen, F., Schachermayer, W.: A general version of the fundamental theorem of asset pricing. Math. Ann. **300**, 463–520 (1994)
10. Derman, E.: Regimes of volatility. Risk **12**(4), 55–59 (1999)
11. Derman, E., Kani, I.: Riding on a smile. Risk **7**(2), 32–39 (1994)
12. Dupire, B.: Pricing with a smile. Risk **7**(1), 18–20 (1994)
13. Durrleman, V.: From implied to spot volatilities. Ph.D. Thesis, Department of Operations Research & Financial Engineering, Princeton University (2004). http://www.cmap.polytechnique.fr/~valdo/papers/dissertation.pdf
14. Durrleman, V.: Convergence of at-the-money implied volatilities to the spot volatility. J. Appl. Probab. **45**, 542–550 (2008)
15. Durrleman, V., El Karoui, N.: Coupling smiles. Quant. Finance **8**, 573–590 (2008)
16. Gatheral, J.: The Volatility Surface: A Practitioner's Guide. Wiley Finance, New York (2006)
17. Hagan, P., Kumar, D., Lesniewski, A., Woodward, D.: Managing smile risk. Wilmott Magazine, pp. 84–108, September 2002
18. Heath, D., Jarrow, R., Morton, A.: Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. Econometrica **60**, 77–105 (1992)
19. Jacod, J., Protter, P.: Risk neutral compatibility with option prices. Finance Stoch. (to appear)
20. Kunita, H.: Stochastic Flows and Stochastic Differential Equations. Cambridge University Press, Cambridge (1990)
21. Ledoit, O., Santa-Clara, P.: Relative pricing of options with stochastic volatility. Technical Report, UCLA, (1999). http://repositories.cdlib.org/anderson/fin/9-98/
22. Lee, R.: Implied volatility: statics, dynamics, and probabilistic interpretation. In: Baeza-Yates, R., et al. (eds.) Recent Advances in Applied Probability, pp. 241–268. Springer, Berlin (2005)
23. Rubinstein, M.: Implied binomial trees. J. Finance **49**, 771–818 (1994)

24. Schönbucher, P.: A market model for stochastic implied volatility. Philos. Trans. R. Soc. Lond. Ser. A **357**(1758), 2071–2092 (1999)
25. Schweizer, M., Wissel, J.: Arbitrage-free market models for option prices: The multi-strike case. Finance Stoch. **12**, 469–505 (2008)
26. Schweizer, M., Wissel, J.: Term structures of implied volatilities: Absence of arbitrage and existence results. Math. Finance **18**, 77–114 (2008)