Markovian Models in the Stochastic Implied Volatility Framework

A Brace, B Goldys, J van der Hoek, R Womersley

24 Sept 2002

Abstract. A general framework for stochastic implied volatility (SIV) is shown to include all Markovian stochastic spot volatility models as special cases. Numerical experiments within the SIV framework indicate that while the initial implied volatility surface determines the marginal distribution of the underlying spot, the volatility of the implied volatility (volvol) determines its joint disribution.

Contents

1. Introduction

In this paper we study a general Markovian model of spot stochastic volatility (spot volatility model for short) from the point of view of the market model of stochastic implied volatility (SIV) as outlined in Section 2 and a previous working paper [1]. Our aim is to demonstrate that just as spot interest rate models like Vasicek fit within the general HJM framework, so do spot volatility models like Heston fit within the SIV framework. Moreover, just as HJM (which in general is Markov in the whole of the yield curve) leads to powerful models that do not fit in the spot framework, so does SIV (which in general is Markov in the underlying spot plus the whole of the implied volatility surface) generate a new class of models in which today's volatility smile or skew is input as a parameter.

In this paper we choose to work within a normal framework because of the existence of a simple one-factor Bachelier spot volatility model **B1** (short for Bachelier 1-Factor Model, see Section 3)) that is particularly suitable for numerical experiments. This B1 model is Markov in the underlying Brownian motion W_t and can produce significant negative implied volatility skews for reasonable values of its parameters. Similarly simple models in the Black-Scholes framework are severely cramped by the requirements that the spot be positive and the call option formula to be invertible for all strikes. For simplicity, we also take interest rates to be zero. Note, however, that we are not losing generality, because with little extra effort all the theoretical results of this paper can be shown to hold equally in the lognormal Black-Scholes framework, while [1] demonstrates how a zero-interest rate SIV model can be articulated to fit equities, FX and interest rates in the real world.

Within our chosen framework we will show in Section 4 that quite general spot volatility models satisfy all the stochastic differential equations (SDEs), and all the boundary and other conditions of the general SIV framework. In doing that, we will find ourselves working with the underlying **spot** S_t, and three volatilities - the **spot volatility** Θ_t, the **implied volatility** $\sigma_t(T,K)$, and the **volvol** or lognormal volatility of the implied volatility $u_t(T,K)$ - which are related via a "feedback" condition that "closes" the system. That closure is important because it puts such strong conditions on the volvol $u_t(T,K)$ that its volatility (the volatility of the volvol or volvolvol!) does not in turn become yet another risk parameter that must be taken into account.

Dupire's results on options and the distributions of the underlying spot, tell us that the initial implied volatility $\sigma_0(T,K)$ will determine the marginal densities of the spot S_T at maturity T. We **conjecture** that it is the volvol $u_t(T,K)$ that will determine the joint density function for the distribution of the spot. To support this conjecture we performed simulation experiments with a SIV model whose initial implied volatility $\sigma_0(T,K)$ was identical to that of B1 but whose volvol $u_t(T,K)$ was "flat" or independent of T. The results supported our conjecture because in this SIV model the prices of standard European call options (vanillas) were identical to those of B1, but the prices of digital cliquets (forward start options) which depend on the joint distribution, were different.

The authors would particularly like to thank Dr Ken Chow and Dr KK Leung of Cluster Technology[1] (Hong Kong) for producing the simulation results reported in this paper (see Section 5). Such results would have been very difficult to obtain without Cluster's cutting edge software which links PC boxes or workstations to form a parallel computer.

2. BACHELIER SIV FRAMEWORK

To set the scene, let us first recall the general Bachelier stochastic implied volatility (SIV) framework outlined in [1]. If W_t denotes one-dimensional Brownian motion under the arbitrage free measure $\mathbb{P}$, then in the Bachelier model with zero interest rates and constant spot volatility σ, the underlying spot S_t is a martingale with form

$$dS_t = \sigma dW_t \quad S_T = S_t + \sigma(W_T - W_t) = S_t + \sigma W_\tau, \quad \tau = T - t, \tag{1}$$

[1] Tel +852-2609-6007 <chowskei@clc.cuhk.edu.hk><kkleung@clc.cuhk.edu.hk>

and the time t price B_t of a call option of strike K maturing at T is

$$B_t = B(t, T, S_t, \sigma, K) = \mathbf{E}\left\{ [S_T - K]^+ \middle| \mathcal{F}_t \right\} = \sigma\sqrt{\tau}\mathbf{\Phi}\left(\frac{S_t - K}{\sigma\sqrt{\tau}}\right), \quad (2)$$

$$\text{where} \quad \mathbf{\Phi}(x) = \int_{-\infty}^{x} \mathbf{N}(u)\, du = \mathbf{N}_1(x) + x\mathbf{N}(x)$$

and $\mathbf{N}_1(\bullet)$ and $\mathbf{N}(\bullet)$ are the standard normal density and cumulative density functions respectively.

The function $\sigma \to B(t, T, S_t, \sigma, K)$ is monotonic increasing when the other variables (t, T, S_t, K) are fixed. So if $C_t = C(t, T, S_t, K)$ is the actual price as observed in the market of a European call option (with strike K and maturing at T), we can solve the equation

$$B(t, T, S_t, \sigma, K) = C(t, T, S_t, K)$$

for the option's (unique) **implied volatility**

$$\sigma_t = \sigma(t, T, S_t, K).$$

Because the relationship between price and implied volatility is one-one with

$$C_t = B(t, T, S_t, \sigma_t, K) \quad (3)$$

and volatility is a more "informative" variable, σ_t rather than C_t is normally quoted in the market as the "price" of the option.

To compute the partial derivatives of B_t that we will require, set

$$\xi = \sigma^2\tau \quad h = \xi^{-\frac{1}{2}}(S - K) \quad B(S, \xi; K) = \xi^{\frac{1}{2}}\mathbf{\Phi}\left(\xi^{-\frac{1}{2}}(S - K)\right), \quad (4)$$

and differentiate to get

$$\partial_S B = -\partial_K B = \mathbf{N}(h), \quad \partial_\xi B = \tfrac{1}{2}\xi^{-\frac{1}{2}}\mathbf{N}_1(h),$$

$$\partial_S^2 B = -\partial_S\partial_K B = \partial_K^2 B = \xi^{-\frac{1}{2}}\mathbf{N}_1(h),$$

$$\partial_\xi^2 B = \tfrac{1}{4}\xi^{-\frac{3}{2}}\left[h^2 - 1\right]\mathbf{N}_1(h), \quad \partial_S\partial_\xi B = -\partial_K\partial_\xi B = -\tfrac{1}{2}\frac{h}{\xi}\mathbf{N}_1(h).$$

Now introduce stochastic implied volatility by setting

$$\xi_t = \xi_t(T, K) = \sigma_t^2(T - t) = \sigma_t^2(T, K)(T - t),$$

and supposing the dynamics of S_t and ξ_t are determined by:

1. the **diffusions**

$$dS_t = \Theta_t dW_t^{(1)}, \tag{5}$$
$$d\xi_t = m_t (T, K, S_t, \Theta_t, \xi_t)\, dt + 2\xi_t\, u_t^* (T, K, S_t, \Theta_t, \xi_t)\, dW_t,$$

   where $W_t = \left(W_t^{(1)}, W_t^{(2)} \right)$ is 2-dimensional Brownian motion under $\mathbb{P}$, and the drift $m_t = m_t (T, K, S_t, \Theta_t, \xi_t)$ and volvol $u_t = u_t (T, K, S_t, \Theta_t, \xi_t)$ functions are assumed to be well behaved,

2. an **initial condition**

$$\xi_0 (T, K) = f (T, K),$$

3. the **feedback condition**

$$\xi_t (T, K)|_{T=t} = 0 \tag{6}$$

   which will force the Bachelier formula (2) to converge to $(S_T - K)^+$ as $t \to T$.

Remark 1. *Supposition (5) implies a lognormal model for the implied volatility because it is equivalent to*

$$d\sigma_t = (\text{drift})\, dt + \sigma_t\, u_t^* dW_t.$$

In this framework the actual call option must, because it is an asset, be a martingale under the arbitrage free measure $\mathbb{P}$. Therefore the SDE for C_t, obtained by applying Ito to (3), must have zero drift forcing the corresponding the drift in the SDE for ξ_t to be

$$m_t = \xi_t u_t^2 - [\Theta_t + (K - S_t)\, u_t]^2 .$$

Hence the system of equations specifying S_t, $\xi_t = \xi_t (T, K)$ and C_t becomes

$$dS_t = \Theta_t dW_t, \tag{7}$$
$$d\xi_t = \xi_t u_t^2 dt - [\Theta_t + (K - S_t)\, u_t]^2\, dt + 2\xi_t u_t dW_t,$$
$$dC_t = \mathbf{N}(h)\, \Theta_t dW_t + \xi_t^{\frac{1}{2}} \mathbf{N}_1 (h)\; u_t dW_t,$$
$$\xi_0 (T, K) = f (T, K), \qquad \xi_t (T, K)|_{t=T} = 0.$$

Assuming Θ_t and u_t are well defined by feedback, a formal solution to (7) is then

$$S_t = S_0 + \int_0^t \Theta_s dW_s, \tag{8}$$
$$\xi_t (T, K) = \exp (2M_t) [f (T, K) - Q_t (T, K)],$$
$$M_t = M_t (T, K) = \int_0^t u_s (T, K)\, dW_s - \frac{1}{2} u_s^2 (T, K)\, ds,$$
$$Q_t = Q_t (T, K) = \int_0^t \exp (2M_s) [\Theta_s + (K - S_s)\; u_s (T, K)]^2\, ds.$$

In conjunction with the feedback condition (6), these equations force $\xi_t(T,K)$ to be positive and show the spot volatility Θ_t, the implied volatility $\sigma_t(T,K)$, and the volvol $u_t(T,K)$ are jointly restricted by having to satisfy

$$f(T,K) = \int_0^T \exp(2M_s)\left[\Theta_s + (K - S_s)\, u_s(T,K)\right]^2 ds. \tag{9}$$

This strong condition implies

$$\partial_T \xi_t(T,K)|_{t=T} = \sigma_t^2(t,K) = \left|\Theta_t + (K - S_t)\, u_t(t,K)\right|^2, \tag{10}$$

and then

$$|\Theta_t| = |\sigma_t(t,S_t)|, \tag{11}$$

which says that the spot volatility is specified endogenously as being equal to the implied volatility of the immediately maturing at-the-money option. Notice, however, that the volvol $u_t(T,K)$ can in part be defined exogenously, just so long as the feedback conditions (6) and (9) are satisfied at $T = t$.

That suggests some exogenous specifications of volvol $u_t(T,K)$ are permissible in the sense that the underlying spot S_t is a true martingale and the model returns initial option prices, while other specification are not permissible, because S_t is only a local martingale and the model fails to return initial option prices. Distinguishing the cases will depend on an, as yet unidentified, existence theorem.

But clearly all Markov spot volatility models, in which the spot S_t is a-priori a true martingale, ought to fit within this framework. By showing their implied volatilities $\xi_t(T,K)$ satisfy (7) and their volvols $u_t(T,K)$ satisfy (9), we will demonstrate that they do.

3. B1 - A Simple Experimental Model

The Bachelier spot volatility model B1 that is the basis for our experiments, has underlying spot

$$\mathbf{B1} \quad S_t = S_0 + \alpha W_t + V_0\left[1 - \mathcal{E}(-\beta W_t)\right], \quad \alpha, \beta, V_0 > 0,$$

where W_t is Brownian motion under the arbitrage free measure. We considered other simple examples (including some in the Black-Scholes framework), but chose to concentrate on B1 mainly because it is a one-factor model, Markov in the underlying Brownian motion, which for reasonable values of its parameters, like

$$(S_0,\ \theta_0,\ \alpha,\ \beta) = (1.0,\ 20\%,\ 15\%,\ 50\%), \tag{12}$$

produces a significant negative skew. We derived expressions in B1 for the implied vol and volvol, including their limits as $t \to T$, and directly showed they satisfied the feedback condition (9). Note that satisfying the feedback condition (9) in B1 is an unambiguous exercise because we don't have to fit a vector to a scalar condition. Moreover programming one-factor models like B1 and its SIV variations is less complicated than for a two-factor model. The character of B1 as a Bachelier model with a shifted lognormal spot volatility is shown by its SDEs

$$dS_t = \Theta_t dW_t \quad d\Theta_t = -\beta\left(\Theta_t - \alpha\right) dW_t, \quad \alpha, \beta > 0,$$

which integrate to return

$$\begin{aligned} \Theta_t &= \alpha + \beta V_0 \mathcal{E}\left(-\beta W_t\right) = \alpha + \beta V_t, \quad \theta_0 = \alpha + \beta V_0 \quad V_0 > 0, \\ S_t &= S_0 + \alpha W_t + V_0\left[1 - \mathcal{E}\left(-\beta W_t\right)\right] = S_0 + \alpha W_t + V_0 - V_t. \end{aligned} \tag{13}$$

It is easy to show that in B1 the time t value of a **vanilla** option (European call option struck at K and maturing at $T = t + \tau$) is

$$\begin{aligned} C_t\left(T, K\right) &= \mathbf{E}\left\{\left[S_T - K\right]^+ \middle| \mathcal{F}_t\right\} = \mathbf{E}\left[f\left(X\right) - K\right]^+, \qquad X \sim \mathbf{N}\left(0,1\right) \\ &= \alpha\sqrt{\tau}\left\{\mathbf{N}_1\left(-c\right) + \frac{\left(S_t - K\right)}{\alpha\sqrt{\tau}}\mathbf{N}\left(-c\right)\right\} + V_t\left[\mathbf{N}\left(-c\right) - \mathbf{N}\left(-c - \beta\sqrt{\tau}\right)\right], \end{aligned} \tag{14}$$

where $f : \mathbb{R} \mapsto \mathbb{R}$ is the monotonic strictly increasing (because $f'\left(x\right) > 0$) function defined by

$$f\left(x\right) = S_t + \alpha\sqrt{\tau}x + V_t\left[1 - \exp\left(-\beta\sqrt{\tau}x - \frac{1}{2}\beta^2\tau\right)\right],$$

and $c = c\left(S_t, V_t\right)$ is the unique root of $f\left(c\right) = K$.

Because it is particularly sensitive to the joint distribution of S_t, we also want the present value of a **digital cliquet** $D_0 = D_0\left(T_1, T_2\right)$ which is a digital forward start option set at T_1 and paying at T_2:

$$\begin{aligned} D_0 = D_0\left(T_1, T_2\right) &= \mathbf{E}\mathbb{I}\left[S_{T_2} - S_{T_1}\right] \\ &= \mathbf{E}\left(\mathbf{E}\left\{\mathbb{I}\left[f\left(X\right) - S_{T_1}\right] \middle| \mathcal{F}_{T_1}\right\}\right), \quad X \sim \mathbf{N}\left(0,1\right) \\ &= \int_{-\infty}^{\infty} \mathbf{N}\left(-d\left(y\right)\right)\mathbf{N}_1\left(y\right) dy, \end{aligned}$$

where $d\left(y\right)$ is defined intrinsically by

$$\alpha\sqrt{\left(T_2 - T_1\right)}d\left(y\right) = V_0 \exp\left(-\beta\sqrt{T_1}y - \frac{1}{2}\beta^2 T_1\right)\left[\exp\left(\begin{array}{c} -\beta\sqrt{\left(T_2 - T_1\right)}d\left(y\right) \\ -\frac{1}{2}\beta^2\left(T_2 - T_1\right) \end{array}\right) - 1\right].$$

3.1. Implied Volatility. From (4) and (14), in B1 the stochastic implied volatility $\sigma_t = \sigma_t(T, K)$ away from the "leading edge" of the surface when $T > t$, is defined intrinsically by

$$B\left(S_t,\ \sigma_t^2(T-t),\ K\right) = C_t(T-t,\ K) = \mathbf{E}[f(X) - K]^+ \quad (T > t)\,. \tag{15}$$

To compute the implied volatility $\widetilde{\sigma}_t = \widetilde{\sigma}_t(K) = \sigma_t(T, K)|_{T=t}$ on the leading edge when $T = t$, first differentiate (15) partially with respect to τ, gather the largest terms, take limits as $\tau \to 0$ and simplify to get

$$\widetilde{\sigma}_t = \lim_{\tau \to 0} \frac{\alpha \mathbf{N}_1(c) + \beta V_t \mathbf{N}_1(c + \beta\sqrt{\tau})}{\mathbf{N}_1(h)} = \lim_{\tau \to 0} \left[\alpha + \beta V_t \exp\left(-\beta\sqrt{\tau}c\right)\right] \frac{\mathbf{N}_1(c)}{\mathbf{N}_1(h)}. \tag{16}$$

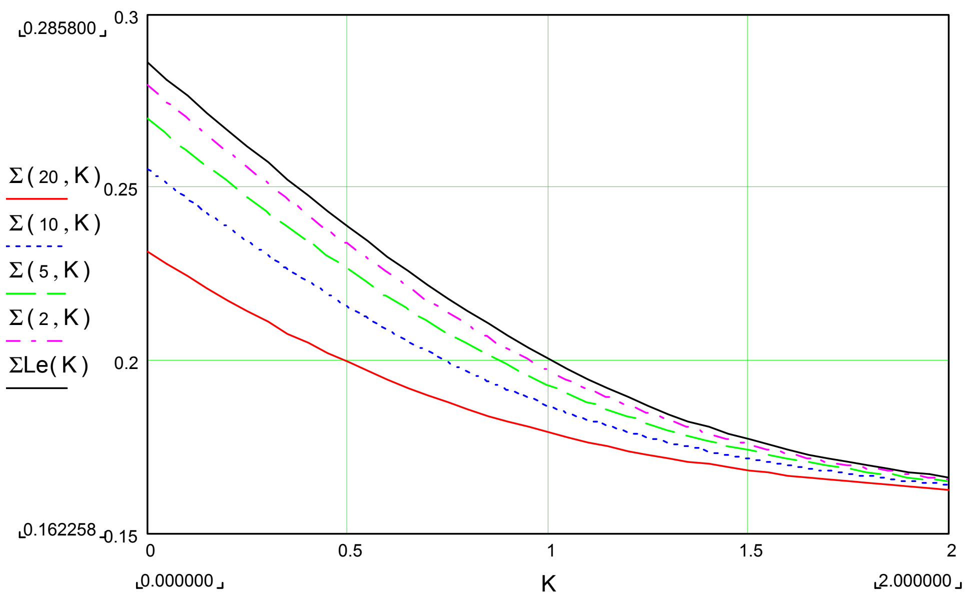
For $\widetilde{\sigma}_t$ to be a strictly finite positive number, that is $0 < \widetilde{\sigma}_t < \infty$, we must have

$$\lim_{\tau \to 0} \sqrt{\tau}c = -\frac{S_t - K}{\widetilde{\sigma}_t}, \tag{17}$$

which leads to the following intrinsic equation for $\widetilde{\sigma}_t$ in terms of K

$$(S_t - K)\left[1 - \frac{\alpha}{\widetilde{\sigma}_t}\right] + V_t\left[1 - \exp\left((S_t - K)\frac{\beta}{\widetilde{\sigma}_t}\right)\right] = 0. \tag{18}$$

The implied volatility surface $\sigma_t(T, K)$ for all maturities $T \geq t$ and strikes K can now be computed, and a selection of profiles is graphed below.



Implied vol profiles for 0-20 year maturities (1.0, 20%, 15%, 50%)

3.2. Volvol and Feedback. For the volvol $u_t(T,K)$ away from the leading edge when $T>t$, differentiate (15) stochastically to obtain

$$u_t(T,K)=\frac{\alpha\left[\mathbf{N}(-c)-\mathbf{N}(h)\right]+\beta V_t\left[\mathbf{N}\left(-c-\beta\sqrt{\tau}\right)-\mathbf{N}(h)\right]}{\sigma_t\sqrt{\tau}\mathbf{N}_1(h)}\quad(T>t). \tag{19}$$

To verify feedback condition (9), differentiate again but now partially with respect to τ, gather the larger terms, take limits as $\tau\to 0$, and simplify to get

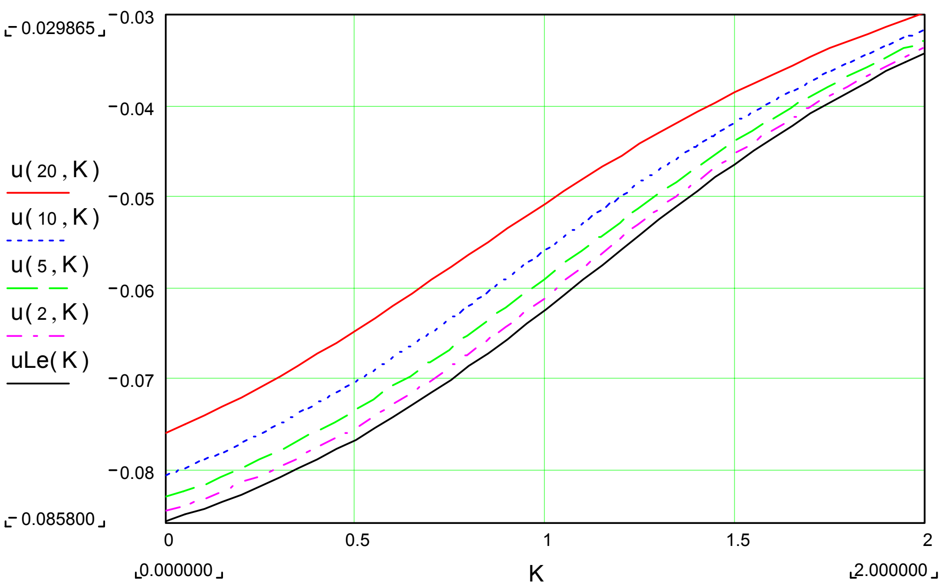
$$\left\{\Theta_t+(K-S_t)u_t\right\}\frac{(S_t-K)}{\widetilde{\sigma}_t}=-\lim_{\tau\to 0}\tau^{\frac{1}{2}}c\left\{\alpha+\beta V_t e\left(-\beta\sqrt{\tau}c\right)\right\}\frac{\mathbf{N}_1(c)}{\mathbf{N}_1(h)},$$
$$\text{or}\qquad \Theta_t+(K-S_t)\widetilde{u}_t(K)=\widetilde{\sigma}_t(K)$$

from (16) and (17).

The volvol $\widetilde{u}_t=\widetilde{u}_t(K)=u_t(t,K)$ on the leading edge when $T=t$ is then

$$\widetilde{u}_t(K)=\begin{cases}\dfrac{\widetilde{\sigma}_t(K)-\Theta_t}{(K-S_t)} & K\neq S_t\\ -\dfrac{1}{2}\dfrac{\beta^2 V_t}{\Theta_t} & K=S_t\end{cases}. \tag{20}$$

permitting computation of $u_t(T,K)$ for all $T\geq t$. A selection of volvol profiles appears below.



Volvol profiles for 0-20 year maturities (1.0, 20%, 15%, 50%)

4. General Spot Volatility Models

In this section we introduce the general spot volatility model (S_t, Θ_t) defined by

$$\begin{cases} dS_t = \Theta_t dW_t^{(1)}, \\ d\Theta_t = f\left(S_t, \Theta_t\right) dt + g_1\left(S_t, \Theta_t\right) dW_t^{(1)} + g_2\left(S_t, \Theta_t\right) dW_t^{(2)}, \\ S_0 = s, \theta_0 = \theta, \end{cases} \tag{21}$$

which is a two-dimensional diffusion that in the Black-Scholes case covers most of the spot volatility models used in finance. We will show that it fits within the general SIV framework by demonstrating that the SDEs (7) are satisfied, they have solution (8) and the feedback conditions (6), (9) and (10) are obeyed.

Putting

$$g\left(s, \theta\right) = \begin{pmatrix} \theta & 0 \\ g_1(s,\theta) & g_2(s,\theta) \end{pmatrix}, \quad X_t = \begin{pmatrix} S_t \\ \Theta_t \end{pmatrix}, \quad W_t = \begin{pmatrix} W_t^{(1)} \\ W_t^{(2)} \end{pmatrix}$$

and denoting (with a slight abuse of notation) the vector $\begin{pmatrix} 0 \\ f(s,\theta) \end{pmatrix}$ by $f(s,\theta)$, we can rewrite (21) in the form

$$\begin{cases} dX_t = f\left(X_t\right) dt + g\left(X_t\right) dW_t, \\ X_0 = x. \end{cases}$$

If the following conditions are satisfied

Condition 2. *There exists* $c > 0$ *such that*

$$|f(x) - f(y)| + |g(x) - g(y)| \leq c|x - y|, \quad x, y \in \mathbb{R}^2,$$

$$< g(x)h, h > \geq c|h|^2, \quad h, x \in \mathbb{R}^2.$$

then (X_t) is a Markov process with the generator

$$LG = \frac{1}{2}\left(gg^* D^2 G\right) + \langle f, DG \rangle, \quad G \in C^2\left(\mathbb{R}^2\right),$$

where $\langle \cdot, \cdot \rangle$ stands for the inner product in $\mathbb{R}^2$ and DG denotes the Fréchet derivative of G.

By the Markov property and putting $\tau = T - t$, the time t price of a European call option with constant volatility σ is

$$\begin{aligned} B_t = B\left(t, T, s, \sigma, K\right) &= \mathbf{E}\left\{ \left(S_T - K\right)^+ \middle| \, S_t = s \right\}, \\ &= B\left(\tau, s, \sigma, K\right) = \sigma\sqrt{\tau}\boldsymbol{\Phi}\left(\frac{s-K}{\sigma\sqrt{\tau}}\right), \\ &= B\left(\sigma\sqrt{\tau}, s\right), \end{aligned} \tag{22}$$

giving

Lemma 3. $B\left(t, T, s, \sigma, K\right) \in C\left([0, T] \times \mathbb{R}_+^4\right)$.

The following fact is well known:

Lemma 4. *We have*

$$\mathbf{E} \sup_{t \leq T} \Theta_t^2 < \infty,$$

and therefore the process (S_t) *is a square-integrable martingale.*

Again by the Markov property the time t price of a European call option with stochastic spot volatility Θ_t is

$$\begin{aligned} C_t = C(t, T, s, \theta, K) &= \mathbf{E}\left(\left(S_T - K\right)^+ \middle| \, S_t = s, \Theta_t = \theta \right) \\ &= \mathbf{E}\left(\left(S_\tau - K\right)^+ |S_0 = s, \theta_0 = \theta \right) = C(\tau, s, \theta, K). \end{aligned}$$

So C satisfies the backward Kolmogorov equation:

$$\begin{cases} \frac{\partial C}{\partial \tau}(\tau, s, \theta, K) = LC(\tau, s, \theta, K), \\ C(0, s, \theta, K) = (s-K)^+, \end{cases}$$

for each K. Moreover because the function $\sigma \to B(\tau, s, \sigma, K)$ is monotonic increasing for each (τ, s, K), the equation

$$C(\tau, s, \theta, K) = B(\tau, s, \sigma, K)$$

has a unique solution that defines the implied volatility $\sigma_\tau = \sigma(\tau, s, \theta, K)$ and which satisfies

$$C(\tau, s, \theta, K) = B(\tau, s, \sigma_\tau, K) = \sigma_\tau\sqrt{\tau}\Phi\left(\frac{s-K}{\sigma_\tau\sqrt{\tau}}\right) = B\left(\sigma_\tau\sqrt{\tau}, s\right) = B\left(z_\tau, s\right) \tag{23}$$

where

$$z_\tau = z(\tau, s, \theta, K) = \sqrt{\tau}\sigma(\tau, s, \theta, K) = \sqrt{\tau}\sigma_\tau. \tag{24}$$

Lemma 5. *We have* $z \in C^{1,2}\left((0,\infty)\times\mathbb{R}^3\right)$ *and for any* $\theta, K \in \mathbb{R}$

$$\lim_{\tau\to 0} z(\tau, s, \theta, K) = 0.$$

So putting $z(0,s,\theta,K)=0$ *for all* $s,\theta,K\in\mathbb{R}$ *we obtain* $z \in C\left([0,\infty)\times\mathbb{R}^3\right)$.

Proof. It follows from the Implicit Function Theorem that $z \in C^{1,2}\left((0,\infty)\times\mathbb{R}^3\right)$. Then because

$$\lim_{\tau\to 0} C(\tau,s,\theta,K) = (s-K)^+, \quad s,\theta,K\in\mathbb{R},$$

(23) yields

$$\lim_{\tau\to 0} z(\tau, s, \theta, K) = 0.$$

4.1. Dynamics of the Implied Volatility Surfaces. The next result shows that the process $\xi_t = \xi\left(T-t, S_t, \Theta_t\right)$ solves the SDE (7) for each $T>0$. In particular, it follows from Theorem 6 below that the process (S_t, Θ_t, ξ_t) when considered on $[0,T)$ is Markov in $\mathbb{R}^3$ for each $T>0$.

Theorem 6. *For each* $T>0$ *the process* (ξ_t) *is a unique solution of the equation*

$$\begin{aligned} d\xi_t = &\left(\left|\left(g^*\nabla U\right)\left(T-t,S_t,\Theta_t\right)\right|^2 \xi_t - \left|\Theta_t - \left(S_t-K\right)\left(g^*\nabla U\right)\left(T-t,S_t,\Theta_t\right)\right|^2\right)dt \\ &+2\xi_t\left\langle \left(g^*\nabla U\right)\left(T-t,S_t,\Theta_t\right), dW_t\right\rangle, \end{aligned} \tag{25}$$

where

$$U(\tau,s,\theta) = \frac{1}{2}\log z^2(\tau,s,\theta).$$

Proof. Using (23) and (24), write

$$Z_t = z\left(T-t,S_t,\Theta_t\right), \quad B\left(Z_t,S_t\right) = C\left(T-t,S_t,\Theta_t\right), \quad t\le T.$$

Because C is a martingale, applying Ito we obtain

$$\frac{\partial B}{\partial z}\left(Z_t,S_t\right)dZ_t + \frac{\partial B}{\partial s}\left(Z_t,S_t\right)dS_t + \frac{1}{2}\frac{\partial^2 B}{\partial z^2}\left(Z_t,S_t\right)d\left\langle Z\right\rangle_t + \frac{1}{2}\Theta_t^2\frac{\partial^2 B}{\partial s^2}\left(Z_t,S_t\right) + \frac{\partial^2 B}{\partial z\partial s}\left(Z_t,S_t\right)d\left\langle Z,S\right\rangle_t$$

$$= \Theta_t\frac{\partial C}{\partial s}\left(T-t,S_t,\Theta_t\right)dW_t^{(1)} + \frac{\partial C}{\partial\theta}\left(T-t,S_t,\Theta_t\right)\left(g_1\left(S_t,\Theta_t\right)dW_t^{(1)} + g_2\left(S_t,\Theta_t\right)dW_t^{(2)}\right)$$

or

$$\frac{\partial B}{\partial z}\left(Z_t,S_t\right)dZ_t = \left(\Theta_t\frac{\partial C}{\partial s}\left(T-t,S_t,\Theta_t\right) + g_1\left(S_t,\Theta_t\right)\frac{\partial C}{\partial\theta}\left(T-t,S_t,\Theta_t\right) - \Theta_t\frac{\partial B}{\partial s}\left(Z_t,S_t\right)\right)dW_t^{(1)}$$

$$+g_2\left(S_t,\Theta_t\right)\frac{\partial C}{\partial\theta}\left(T-t,S_t,\Theta_t\right)dW_t^{(2)}+H_tdt.$$

Putting

$$u_t^{(1)}=\frac{\Theta_t\left(\frac{\partial C}{\partial s}\left(T-t,S_t,\Theta_t\right)-\frac{\partial B}{\partial s}\left(Z_t,S_t\right)\right)+g_1\left(S_t,\Theta_t\right)\frac{\partial C}{\partial\theta}\left(T-t,S_t,\Theta_t\right)}{Z_t\frac{\partial B}{\partial z}\left(Z_t,S_t\right)},$$

and

$$u_t^{(2)}=\frac{g_2\left(S_t,\Theta_t\right)\frac{\partial C}{\partial\theta}\left(T-t,S_t,\Theta_t\right)}{Z_t\frac{\partial B}{\partial z}\left(Z_t,S_t\right)},$$

it is easy to check that

$$H_t=-\frac{1}{2Z_t}\left(\left(S_t-K\right)^2\left|u_t\right|^2+\Theta_t^2-2\Theta_tu_t^{(1)}\left(S_t-K\right)\right)\phi\left(\frac{S_t-K}{Z_t}\right),$$

and therefore,

$$d\xi_t=\left(\left|u_t\right|^2\xi_t-\left|\Theta_t-\left(S_t-K\right)u_t\right|^2\right)dt+2\xi_tu_t^{(1)}dW_t^{(1)}+2\xi_tu_t^{(2)}dW_t^{(2)}.$$

Let

$$\tilde{u}^{(1)}(\tau,s,\theta)=\frac{\theta\left(\frac{\partial C}{\partial s}\left(\tau,s,\theta\right)-\frac{\partial B}{\partial s}\left(z_\tau,s\right)\right)+g_1\left(s,\theta\right)\frac{\partial C}{\partial\theta}\left(\tau,s,\theta\right)}{z_\tau\frac{\partial B}{\partial z}\left(z_\tau,s\right)},$$

$$\tilde{u}^{(2)}(\tau,s,\theta)=\frac{g_2\left(s,\theta\right)\frac{\partial C}{\partial\theta}\left(\tau,s,\theta\right)}{z_\tau\frac{\partial B}{\partial z}\left(z_\tau,s\right)}.$$

By (23)

$$\frac{\partial C}{\partial\theta}\left(\tau,s,\theta\right)=\frac{\partial B}{\partial z}\frac{\partial z_\tau}{\partial\theta},\qquad\frac{\partial C}{\partial s}\left(\tau,s,\theta\right)=\frac{\partial B}{\partial z}\frac{\partial z_\tau}{\partial s}+\frac{\partial B}{\partial s},$$

and therefore

$$\tilde{u}^{(1)}(\tau,s,\theta)=\frac{\theta\frac{\partial B}{\partial z}\frac{\partial z_\tau}{\partial s}+g_1\left(s,\theta\right)\frac{\partial B}{\partial z}\frac{\partial z_\tau}{\partial\theta}}{z_\tau\frac{\partial B}{\partial z}\left(z_\tau,s\right)}=\theta\frac{1}{z_\tau}\frac{\partial z_\tau}{\partial s}+g_1\left(s,\theta\right)\frac{1}{z_\tau}\frac{\partial z_\tau}{\partial\theta},$$

$$\tilde{u}^{(2)}(\tau,s,\theta)=\frac{g_2\left(s,\theta\right)\frac{\partial B}{\partial z}\frac{\partial z_\tau}{\partial\theta}}{z_\tau\frac{\partial B}{\partial z}\left(z_\tau,s\right)}=g_2\left(s,\theta\right)\frac{1}{z_\tau}\frac{\partial z_\tau}{\partial\theta}.$$

Hence,

$$\tilde{u}^{(1)}(\tau,s,\theta)=\theta\tfrac{\partial U}{\partial s}(\tau,s,\theta)+g_1\left(s,\theta\right)\tfrac{\partial U}{\partial\theta}(\tau,s,\theta),$$
$$\tilde{u}^{(2)}(\tau,s,\theta)=g_2\left(s,\theta\right)\tfrac{\partial U}{\partial\theta}(\tau,s,\theta),$$

or

$$\begin{pmatrix}\tilde{u}^{(1)}(\tau)\\\tilde{u}^{(2)}(\tau)\end{pmatrix}=g^*\nabla U(\tau),$$

and the theorem follows.

Let

$$M_t = \int_0^t \langle (g^* \nabla U)(T-t, S_t, \Theta_t), dW_t \rangle, \quad t < T, \qquad \rho_t = \exp\left(M_t - \frac{1}{2}\langle M \rangle_t \right).$$

The next Theorem 7 shows that the solution to (25) is of form (8) and satisfies the feedback conditions (6), (9) and (10).

Theorem 7. *For any $T > 0$ and $K \in \mathbb{R}$ we have for all $t < T$,*

$$\xi_t(T,K) = \int_t^T \left(\frac{\rho_t}{\rho_s}\right)^2 |\Theta_s - (S_s - K)(g^* \nabla U)(T-s, S_s, \Theta_s)|^2 \, ds,$$

where we use the notation Θ_s for the vector $(\Theta_s, 0)$ as well. In particular,

$$f(T,K) = \int_0^T \rho_s^{-2} |\Theta_s - (S_s - K)(g^* \nabla U)(T-s, S_s, \Theta_s)|^2 \, ds.$$

Proof. By the Ito Lemma we have for any $t < T$

$$\xi_t(T,K) = \rho_t^2 \left(f(T,K) - \int_0^t \rho_s^{-2} |\Theta_s - (S_s - K)(g^* \nabla U)(T-s, S_s, \Theta_s)|^2 \, ds \right). \tag{26}$$

Since $\xi_t(T,K) \geq 0$ we find that

$$\int_0^t \rho_s^{-2} |\Theta_s - (S_s - K)(g^* \nabla U)(T-s, S_s, \Theta_s)|^2 \, ds \leq f(T,K), \tag{27}$$

hence

$$\int_0^T \rho_s^{-2} |\Theta_s - (S_s - K)(g^* \nabla U)(T-s, S_s, \Theta_s)|^2 \, ds < \infty.$$

Assume that inequality in (27) is strict. Then by Lemma 5 and (26) we obtain

$$\lim_{t \to T} \rho_t = 0,$$

and therefore by the Strong Law of Large Numbers for Local Martingales, see for example [3],

$$\langle M \rangle_T = \int_0^T |(g^* \nabla U)(T-t, S_t, \Theta_t)|^2 \, dt = \infty,$$

which yields contradiction:

$$\int_0^T \rho_s^{-2} |\Theta_s - (S_s - K)(g^* \nabla U)(T-s, S_s, \Theta_s)|^2 \, ds = \infty.$$

Hence ρ_t does not converge to zero and since $\xi_T(T,K) = 0$, we obtain

$$f(T,K) = \int_0^T \rho_s^{-2} \left|\Theta_s - (S_s - K)(g^* \nabla U)(T-s, S_s, \Theta_s)\right|^2 ds,$$

and then

$$\xi_t(T,K) = \int_t^T \left(\frac{\rho_t}{\rho_s}\right)^2 \left|\Theta_s - (S_s - K)(g^* \nabla U)(T-s, S_s, \Theta_s)\right|^2 ds.$$

5. Simulation Experiments

Our aim in this section is to test our conjecture that the volvol determines the joint distribution, by simulating spots, vanillas and digital cliquets in three ways:
(A) Using B1, directly from the defining equation (13) for S_t.
(B) Within the SIV framework and with B1's initial implied volatility surface as given by (15) and (16), but using two different types of volvol function

$$u_t(T,K) = \begin{cases} u_t(t,K) & \textbf{Flat} \\ u_t(t,K)\left[1 + (T-t)\tan 10^\circ\right] & \textbf{Up-10} \end{cases}. \tag{28}$$

(C) Within the SIV framework, using B1's initial implied volatility surface, and B1's volvol as defined by (19) and (20).

Method-A is straightforward; directly simulate the values W_{T_1} and W_{T_2} of Brownian motions at T_1 and T_2 from

$$W_{T_1} \sim \mathbf{N}\left(0, {T_1}^2\right) \quad \text{and} \quad W_{T_2} - W_{T_1} \sim \mathbf{N}\left(0, (T_2 - T_1)^2\right),$$

and compute S_{T_2} and S_{T_1}.

Methods B and C are more complicated. Because the general SIV model is Markov in both the underlying spot and the implied volatility surface we need to carry S_t and the surface $\xi_t(T,K)$, and also compute $u_t(T,K)$ at each timestep. Assuming the same time grid for both t and T, proceed as follows:

1. Suppose at time t we have got S_t and $\xi_t(T,K)$ for $T \geq t$ and now want to simulate $S_{t+\Delta t}$ and $\xi_{t+\Delta t}(T,K)$ for $T \geq t + \Delta t$ at time $t + \Delta t$.

2. At time t produce the next increment ΔW_t of Brownian motion for use over the current time interval $[t, t+\Delta t]$.

3. From the surface $\xi_t(T,K)$ numerically compute $\sigma_t(t,K) > 0$ from

$$\sigma_t^2(t,K) = \partial_T \xi_t(T,K)\big|_{t=T} = \frac{\xi_t(t+\Delta t, K) - 0}{\Delta t}$$
$$= \frac{\exp(2M_t(t+\Delta t, K))}{\Delta t}\left\{f(t+\Delta t, K) - D_t(t+\Delta t, K)\right\}.$$

4. Find Θ_t by setting $K = S_t$ in (9)

$$\Theta_t = \sigma_t(t, S_t).$$

5. Define the dependence of the volvol $u_t(T,K)$ on maturity T in a way that is compatible with (9)

$$[\Theta_t + (K - S_t)\, u_t(t,K)] = \sigma_t(t,K).$$

6. Use (8) to increment, evaluating all integrands off the left-hand sides of intervals to ensure convergence,

$$S_{t+\Delta t} = S_t + \Theta_t \Delta W_t,$$
$$M_{t+\Delta t}(T,K) = M_t(T,K) + u_t(T,K)\,\Delta W_t - \frac{1}{2}u_t^2(T,K)\,\Delta t,$$
$$Q_{t+\Delta t}(T,K) = Q_t(T,K) + \exp(-2M_t(T,K))\,[\Theta_t + (K - S_t)\, u_t(T,K)]^2\,\Delta t,$$
$$\xi_{t+\Delta t}(T,K) = \exp(2M_{t+\Delta t}(T,K))\,[f(T,K) - Q_{t+\Delta t}(T,K)].$$

7. With $S_{t+\Delta t}$ and $\xi_{t+\Delta t}(T,K)$ now computed at time $t+\Delta t$, return to Step-1, reset t to $t+\Delta t$ and iterate to generate trajectories of S_t.

The appropriate volvol function is inserted at Steps 5&6; Flat or Up-10 in **Method-B**, B1's in **Method-C**.

5.1. Technical Points. The simulation was carried out on a cluster of SunBlade 1000 workstations with 32 UltraSPRAC III CPUs running at 750 MHz. To get the sort of convergence and accuracy we wanted, some 6400 time steps per trajectory were necessary at a cost of about 16 hours to simulate 50k paths. While this computational load may appear to be an obstacle to practical implementation, there are several mitigating factors :

1. Any practical SIV model can probably be approximated by a low-dimensional Markov model, which could be used for variance reduction and to roughly predict trajectories. With good variance reduction far fewer trajectories than the above are needed for a usable market price Time and strike grids might also be tailored to the trajectory and made sparser, while some trajectories could be rejected (like those finishing well out-of-the-money in the case of a vanilla).

2. The speed of simulation is linearly proportional to the clock speed of the CPU. Currently CPUs with clock speeds at least 3 times higher than the 750 MHz CPUs used for our experiments, are widely available. A cluster of 64 such "commodity-off-the-shelf hardware" can be bought for the rounding error on the daily P&L of a trading desk, and would improve speeds by a factor of six.

3. Path generation occupies most of the time needed to simulate the price of an option. But because the SIV model is cast in terms of moneyness rather than absolute spot level, movement in spot levels can be handled without regeneration of the paths and only changes in the initial implied volatility surface need trigger resimulation. Therefore a set of paths might be pregenerated to price different derivatives during the course of a day.

4. Because implied volatilities require surfaces indexed by strike and maturity while yield curves require just vectors indexed by maturity, simulating interest rates in a SIV framework will only be marginally more time consuming than simulating FX or equities.

Collaborative numerical work with Cluster Technology is in progress, and we expect simulation times to be reduced considerably.

5.2. Effect of changing the volvol. To support our conjectures about the connection between the volvol and distributions, we first investigated differences in pricing by Methods B and A. Using Flat and Up-10 volvol functions (28), we simulated $50k$ differences and report the test statistic

$$\sigma_{BA} = \frac{\text{Mean}\left\{(\text{Price-B}) - (\text{Price-A})\right\}}{\frac{\text{StDev}\{(\text{Price-B})-(\text{Price-A})\}}{\sqrt{N}}},$$

which allows us to conclude at a 95% confidence level that Price-B equals Price-A.if $\sigma_{BA} < 2$.

The results for the spot, vanillas and cliquets in the **Flat** case were

Vanilla ($T =$	1	2	3	4	5
S_T	-0.80	-0.55	0.77	-0.16	-0.36
$\Delta K = .10$	2.16	1.10	-0.85	-0.81	0.21
.25	1.85	1.81	0.36	1.25	0.78
.50	-0.47	-0.08	0.43	-0.49	-0.09
.75	-1.90	-0.78	-0.19	-0.94	-0.12
.90	-1.83	-1.66	-0.08	-1.06	-1.03

Cliquet ($T_2 =$	2	3	4	5
$T_1 = 1$	-3.75	-4.52	-5.16	-3.41
2		-3.17	-5.22	-5.58
3			-4.81	-6.02
4				-7.49

while in the **Up-10** case the results for the spot were

Spot ($T =$	1	2	3	4	5
S_T	-0.12	1.47	4.35	6.69	5.81

On notionals of $\$1M$, sample standard deviations for spots and vanillas were of the order of \$10, and for digital cliquets of the order of \$100.

We deduce that:

1. Some specifications of volvol make the model untenable; assets are no longer martingales,
2. When the model is tenable, the marginal distribution of the underlying is determined by the implied vol, while the joint distribution is determined by the volvol.

5.3. Control Experiment. To test the accuracy of our numerical procedures in the SIV framework, we also investigated, whether Methods C and A produce, as they ought, the same prices for spots, vanillas, and cliquets.

That presented the computationally exhausting problem of simulating $50k$ trajectories, with each trajectory requiring at each of $6.4k$ separate timesteps computation of B1's volvol surface as defined in equations (19) and (20). We got around the problem by linearly interpolating between the leading and far edges of the volvol surface at Step-6, and report the consequent test statistic

$$\sigma_{BA} = \frac{\text{Mean}\{(\text{Price-C}) - (\text{Price-A})\}}{\frac{\text{StDev}\{(\text{Price-C})-(\text{Price-A})\}}{\sqrt{N}}},$$

for spot and at-the-money vanillas

Vanilla ($T =$	1	2	3	4	5
S_T	-0.55	0.82	-0.69	-0.59	0.22
$\Delta K = .50$	2.16	1.10	-0.85	-0.81	0.21

and for cliquets

Cliquet ($T_2 =$	2	3	4	5
$T_1 = 1$	-1.79	1.41	-1.43	0.30
2		0.53	0.55	-0.96
3			1.23	-0.73
4				-0.29

If we could reproduce these sorts of numbers for all away-from-the-money vanillas, we could unequivocally claim that our numerical procedures were accurate. But with linear interpolation of the volvol surface, results for well away-from-the-money options were clearly skewed and off by up to 6 sample standard deviations. A more thorough investigation using quadratic interpolation of the volvol surface significantly improved those numbers, but proved expensive in computing time and is deferred. This slightly inconclusive result should be kept in perspective by noting that errors of 6 versus 2 sample standard deviations, means errors of about \$15 versus \$5 on vanilla options with face values of \$1M.

6. Conclusion

To be able to experiment with a tractable spot volatility model, we chose to work in a Bachelier framework, but all our results apply equally in the lognormal Black-Scholes framework. We proved that quite general Bachelier spot volatility models fit within the Bachelier stochastic implied volatility (SIV) framework, satisfying all the required stochastic differential equations and feedback conditions. In addition we constructed within the SIV framework a simple one-factor model, which is Markov in the spot and the whole of the implied volatility surface, and in which the prices of assets appear to be martingales under the arbitrage free measure for certain specifications of the volvol. Articulation of the SIV model now requires some of the following issues to be addressed.

We demonstrated that while the marginal distribution of the spot is determined by the implied volatility, its joint distribution is controlled by the volvol, and find support for that notion in anecdotes about some banks pricing FX exotics by matching their risk characteristics to a linear combination of vanillas (for correct marginals) and barriers (for correct joint). A worthwhile result would be to connect the volvol to the joint distribution similarly to the way Dupire's results connect the implied volatility to the marginal distribution.

Note that the problem of ensuring initial implied volatility surfaces are "feasible" was sidestepped in this paper. Exactly what feasible means requires clarification, but at the very least the initial implied volatility surface must for all maturities generate marginal probability distributions for the spot which are consistent with the spot being a martingale (see [4] for work along these lines). By taking our initial volatility surface from B1 we a-priori ensure that it is feasible, and it is interesting to hear anecdotal evidence that some banks make a practice of interpolating their implied volatilities in just this fashion using a selection of spot volatility models.

In contrast, the attractive "toy model" described in [1] (where in a Bachelier framework the evolving implied volatility surface is always quadratic in the moneyness) is not feasible because the putative distributions do not integrate to unity. Analysis of this example shows that one necessary condition for an initial implied volatility surface to be feasible is that it be sublinear in moneyness at infinity.

With or without the feasibility issue settled, an obvious next step is to continue in a one-factor Bachelier framework with an initial volatility surface defined by B1 and identify necessary and sufficient conditions (related to the maturity dependence of the volvol) for the underlying spot to be a true martingale. Then settle the feasibility issue and repeat the proof in general Bachelier and Black-Scholes frameworks.

Working in a multi-factor SIV model will involve the extra problem of how to distribute the vector volvol (as opposed to scalar volvol in the one-factor case) into its components in a manner consistent with the scalar feedback condition (9), and then how to specify the dependence of those components on maturity. Probably experiments based on a simple (like B1) two-factor model will help provide answers in this more general setting.

References

[1] A Brace, B Goldys, F Klebaner & R Womersley (2001) "Market Model of Implied Volatility with Application to BGM" *Working Paper UNSW.*

[2] Carr P (2000) "A survey of preference free option valuation with stochastic volatility" *April Risk Conference, Paris.*

[3] Liptser "Theory of Martingales", Kluwer

[4] Madan D & Yor M (2001) "Making Markov Martingales Meet Marginals: With Explicit Constructions" *Working Paper.*

[5] Schonbucher PJ (1998) "A market model for stochastic implied volatility" *Working paper, University Bonn.*