International Journal of Theoretical and Applied Finance
Vol. 12, No. 6 (2009) 861–876

SMALL-TIME ASYMPTOTICS FOR IMPLIED VOLATILITY UNDER THE HESTON MODEL

MARTIN FORDE*

Department of Mathematical Sciences
Dublin City University, Glasnevin, Dublin 9, Ireland
Martin.Forde@dcu.ie

ANTOINE JACQUIER

Department of Mathematics, Imperial College
180 Queen's Gate, London
and
Zeliade Systems, 56 Rue Jean-Jacques Rousseau, Paris
antoine.jacquier08@imperial.ac.uk

Received 2 October 2008
Accepted 28 January 2009

We rigorize the work of Lewis (2007) and Durrleman (2005) on the small-time asymptotic behavior of the implied volatility under the Heston stochastic volatility model (Theorem 2.1). We apply the Gärtner-Ellis theorem from large deviations theory to the exponential affine closed-form expression for the moment generating function of the log forward price, to show that it satisfies a small-time large deviation principle. The rate function is computed as *Fenchel-Legendre transform*, and we show that this can actually be computed as a standard Legendre transform, which is a simple numerical root-finding exercise. We establish the corresponding result for implied volatility in Theorem 3.1, using well known bounds on the standard Normal distribution function. In Theorem 3.2 we compute the level, the slope and the curvature of the implied volatility in the small-maturity limit At-the-money, and the answer is consistent with that obtained by formal PDE methods by Lewis (2000) and probabilistic methods by Durrleman (2004).

Keywords: Implied volatility asymptotics; Heston; large deviation; small-time behavior.

1. Introduction

In recent years there has been an explosion of literature on small-time asymptotics for stochastic volatility models, see Hagan *et al.* [13], Berestycki *et al.* [4, 5], Henry-Labordère [14, 15] and Laurence [18]. All these articles are essentially higher order corrections and/or applications of the seminal work of Varadhan [23, 24], who showed that the small-time behavior of a diffusion process can be characterized in

*Corresponding author.

terms of an energy/distance function on a Riemmanian manifold, whose metric is given by the inverse of the diffusion coefficient (see also Dembo and Zeitouni [7]).

For a local volatility model, Berestycki *et al.* [4] showed that the implied volatility in the small-maturity limit is the *harmonic mean* of the local volatility. For a general stochastic volatility model satisfying certain growth conditions, Berestycki *et al.* [5] show that the implied volatility in the small-maturity limit can be characterized as a *viscosity solution* to a first order non-linear eikonal PDE. Viscosity solutions are a weak formulation for solutions to PDEs introduced by Crandall and Lions [6]. Viscosity solutions can be non-differentiable, but are in some sense sandwiched between a differentiable sub- and super-solution, which we call test functions (see Fleming and Soner [11]). In general, for optimal control problems, the value function is not smooth enough to satisfy the dynamic programming equation in the classical sense, but it is the unique viscosity solution to that equation.

The solution to this eikonal equation is the length of the shortest geodesic from a point to a line under the aforementioned metric, i.e., it is a *variable endpoint* variational problem. Durrleman [9] applied the *Legendre* transform to the eikonal equation for this shortest geodesic distance under the Heston model, and then solved the resulting ODE. Durrleman did not address the highly non-trivial issue of existence and uniqueness of solutions to this eikonal equation. Lewis [20] generalized the work of Durrleman, providing a formal derivation of the small-time implied volatility asymptotics for a general CEV(p)-volatility model, which nests the Heston model. Lewis computed the length of the aforementioned geodesics using conserved energy and momentum arguments that arise from integrating the geodesic equations, and a *transversality* condition for the variable endpoint problem.

Feng *et al.* [10] consider the asymptotic behavior of the Heston model in a regime where the maturity is small, but large compared to the mean reversion time for the volatility process. Using the fact that the Gärtner-Ellis theorem generalizes under Mosco convergence, they establish a Large deviation principle for the log stock price in this small-time, fast mean-reverting regime, and derive corresponding results for call option prices and implied volatility. Medvedev and Scaillet [22] consider a general stochastic volatility model and derive an asymptotic formula for small maturities in terms of a certain modified moneyness; this parametrization does not capture the effect of large deviations because the strike is forced to converge to the spot price as the maturity goes to zero. Alòs and Ewald [1] take a very novel approach, and use Malliavin calculus to derive an approximate pricing formula for European options under the Heston model when the maturity and the volatility-of-variance are $\ll 1$.

In this article, we provide a *rigorous* analysis of the small-time Heston stochastic volatility model. The Heston model is problematic because the singular coefficient $\sqrt{y}$ in the volatility-of-variance function is not Lipschitz with respect to the usual Euclidean metric, so we cannot readily apply the standard Freidlin-Wentzell (FW) theory of large deviations for stochastic differential equations. Moreover, it is non-trivial to establish whether the *comparison principle* holds for the Hamiltonian associated with the Heston process. We require the comparison principle to hold if we wish to apply the extensions of FW theory outlined in Feng and Kurtz [12] to establish a trajectory-level large deviation principle. In this article we take a simpler route, by analyzing the well known closed-form expression for the moment generating function of the log stock price for the Heston model, and using the Gärtner-Ellis theorem from large deviations theory to show that the log forward price satisfies a small-time large deviation principle (Theorem 2.1). In this way, we sidestep the problem of existence/uniqueness of classical or viscosity solutions to the eikonal equation and/or the variable endpoint calculus of variations problem associated with the Heston model. The rate function is computed as a *Fenchel-Legendre transform*, and we prove that this can actually be computed more efficiently as a Legendre transform. Using well known bounds on the standard Normal distribution function, we then convert this result into a statement about the limiting behavior of the implied volatility of a call option in the small-time regime (Theorem 3.1). It is not obvious how one might extend this approach to other stochastic volatility models, e.g., the SABR model in [13], where the log stock price mgf is not known in closed-form.

This article is similar in spirit to Feng *et al.* [10]; in this article the term involving y_0 in the log stock price mgf dominates in the small-time regime, but in the [10] regime the asymptotics are dominated by the ergodic behavior of the variance process which is independent of y_0.

2. A Small-Time Large Deviation Principle for the Log Stock Price under the Heston Model

We work on a model $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ supporting two Brownian motions, and satisfying the usual conditions. All expectations are taken under $\mathbb{P}$ unless otherwise stated, and we assume that the interest rate r is constant.

Theorem 2.1. *Consider the Heston stochastic volatility model for a log forward price process X_t,*[1] *defined by the following stochastic differential equations*

$$\begin{cases} dX_t = -\frac{1}{2}Y_t dt + \sqrt{Y_t}dW_t^1, \\ dY_t = \kappa(\theta - Y_t)dt + \sigma\sqrt{Y_t}dW_t^2, \\ dW_t^1 dW_t^2 = \rho dt, \end{cases} \tag{2.1}$$

with $X_0 = x_0, Y_0 = y_0, \kappa, y_\infty, \sigma > 0, |\rho| < 1$ and $2\kappa y_\infty > \sigma^2$, so that $Y = 0$ is an unattainable barrier, where W and B are two correlated Brownian motions.[2]

[1] The interest rate does not materially affect any of the results in the article because we are only dealing with European-style contracts, so we work directly with the forward rate process.

[2] The Y process is the Cox-Ingersoll-Ross (CIR) diffusion (also known as the square root process) and it satisfies the Yamada-Watanabe condition (see page 291, Proposition 2.13 in Karatzas and

Table 1.

ρ	p_-	p_+
<0	$\frac{\arctan(\frac{\sqrt{1-\rho^2}}{\rho})}{\frac{1}{2}\sigma\sqrt{1-\rho^2}}$	$\frac{\pi+\arctan\left(\frac{\sqrt{1-\rho^2}}{\rho}\right)}{\frac{1}{2}\sigma\sqrt{1-\rho^2}}$
$=0$	$-\frac{\pi}{\sigma}$	$\frac{\pi}{\sigma}$
>0	$\frac{-\pi+\arctan\left(\frac{\sqrt{1-\rho^2}}{\rho}\right)}{\frac{1}{2}\sigma\sqrt{1-\rho^2}}$	$\frac{\arctan\left(\frac{\sqrt{1-\rho^2}}{\rho}\right)}{\frac{1}{2}\sigma\sqrt{1-\rho^2}}$.

Then $X_t - x_0$ satisfies a large deviation principle (LDP) as $t \to 0$, with rate function $\Lambda^*(x)$ equal to the Legendre transform of the continuous function $\Lambda : \mathbb{R} \to \mathbb{R}^+ \cup \{\infty\}$ given by

$$\Lambda(p) = \frac{y_0 p}{\sigma\left(\sqrt{1-\rho^2}\cot\left(\frac{1}{2}\sigma p\sqrt{1-\rho^2}\right) - \rho\right)} \quad \text{for } p \in (p_-, p_+)$$
$$= \infty \quad \text{for } p \notin (p_-, p_+), \tag{2.2}$$

where Table 1 shows how to compute the values of p_- and p_+

We also have that

$$-\lim_{t\to 0} t\log\mathbb{P}(X_t - x_0 > k) = -\lim_{t\to 0} t\log\mathbb{P}(X_t - x_0 \geq k)$$
$$= \inf_{\{x:x\geq k\}} \Lambda^*(x) = \Lambda^*(k) \tag{2.3}$$

for $k \geq 0$, because $\Lambda^*(x)$ is continuous and strictly increasing when $x \geq 0$, and

$$-\lim_{t\to 0} t\log\mathbb{P}(X_t - x_0 < k) = -\lim_{t\to 0} t\log\mathbb{P}(X_t - x_0 \leq k)$$
$$= \inf_{\{x:x\leq k\}} \Lambda^{*'}(x) = \Lambda^*(k) \tag{2.4}$$

for $k \leq 0$, because $\Lambda^*(x)$ is continuous and strictly decreasing when $x \leq 0$.

Proof. From Albrecher *et al.* [2], we have the following closed-form expression for the characteristic function $\phi(k,t)$ of $X_t - x_0$

$$\phi(k,t) = \mathbb{E}(e^{ik(X_t - X_0)})$$
$$= \exp\left\{\frac{\kappa\theta}{\sigma^2}\left[(\kappa - \rho\sigma ik - d)t - 2\log\left(\frac{1-g_2e^{-dt}}{1-g_2}\right)\right]\right\}$$
$$\times \exp\left\{\frac{y_0}{\sigma^2}(\kappa - \rho\sigma ik - d)\frac{1-e^{-dt}}{1-g_2e^{-dt}}\right\}, \tag{2.5}$$

Shreve [16]), so it admits a unique strong solution. The X process can be expressed as a stochastic integral of the Y process, so the X process is also well defined.

where

$$
\begin{aligned}
g_2 = g_2(k) &= \frac{\kappa - \rho\sigma i k - d(k)}{\kappa - \rho\sigma i k + d(k)},\\
d = d(k) &= \sqrt{(\kappa - \rho\sigma i k)^2 + \sigma^2(ik + k^2)},
\end{aligned}
\tag{2.6}
$$

and we take the principal branch for the complex log function. For the square root here, either of the two roots may be chosen, because the characteristic function is even in d. We now recall the following theorem

Theorem 2.2 (Lukacs [21, Theorem 7.1.1]). *If a characteristic function $\psi(k)$ is regular*[3] *in the neighborhood of $k = 0$, then it is also regular in a horizontal strip and can be represented in this strip by a Fourier integral. This strip is either the whole plane, or it has one or two horizontal boundary lines. The purely imaginary points on the boundary of the strip of regularity (if this strip is not the whole plane) are singular points of $\psi(z)$.*

(see also Lewis [19]). $\kappa > 0$ by assumption, so the square root function $\sqrt{z}$ that appears in the definition of $d(\cdot)$ is analytic at $z = \kappa^2$. Also, for t sufficiently small and $k = 0$, the argument being passed to the complex logarithm function in Eq. (2.5) tends to 1 where the log function is analytic. Thus, for t sufficiently small, the characteristic function of $X_t - x_0$ is analytic at $k = 0$, so substituting $k = -ip$ for $p \in \mathbb{R}$ in Eq. (2.5), we can apply Theorem 2.2 to obtain the analytic continuation

$$
\mathbb{E}(e^{p(X_t - x_0)}) = \exp\{C(t,p) + D(t,p)y_0\} \quad \text{for } t < T^*(p) \tag{2.7}
$$

$$
= +\infty \quad \text{for } t \geq T^*(p), \tag{2.8}
$$

where

$$
\begin{aligned}
C(t,p) &= \frac{\kappa\theta}{\sigma^2}\left[(\kappa - \rho\sigma p - d)t - 2\log\left(\frac{1 - g_2 e^{-dt}}{1 - g_2}\right)\right],\\
D(t,p) &= \frac{y_0}{\sigma^2}(\kappa - \rho\sigma p - d)\frac{1 - e^{-dt}}{1 - g_2 e^{-dt}},\\
g_2 = g_2(p) &= \frac{\kappa - \rho\sigma p - d(p)}{\kappa - \rho\sigma p + d(p)},\\
d = d(p) &= \sqrt{(\kappa - \rho\sigma p)^2 + \sigma^2(p - p^2)}
\end{aligned}
\tag{2.9}
$$

and

$$
\begin{aligned}
T^*(p) = \frac{2}{\sqrt{\sigma^2 p(p-1) - (\kappa - \rho\sigma p)^2}} &\left[\pi 1_{\{\rho\sigma p - \kappa < 0\}}\right.\\
&\left. + \arctan\left(\frac{\sqrt{\sigma^2 p(p-1) - (\kappa - \rho\sigma p)^2}}{\rho\sigma p - \kappa}\right)\right]
\end{aligned}
\tag{2.10}
$$

[3]A characteristic function $\psi(u)$ is regular if it is analytic and single-valued. $\psi(u)$ is analytic if there exists a function $A : \mathbb{C} \to \mathbb{C}$ which is analytic in some circle $z : |z| < r$, and $A(z) = \psi(z)$ for $z \in \mathbb{R}$ in a neighborhood of $k = 0$ on the real line.

is the critical explosion time for the pth moment if

$$\sigma^2 p(p-1) > (\kappa - \rho\sigma p)^2 \tag{2.11}$$

(the expression for $T^*(p)$ comes from Proposition 3.1, part (iii) in Andersen and Piterbarg [3], see also Keller-Ressel [17]). The moment explosion is caused by $\bar{b}\cot(\frac{1}{2}\bar{b}t) + b$ tending to zero as $t \searrow 0$. Setting $p \mapsto \frac{p}{t}$ and letting $t \searrow 0$ in Eq. (2.10), we see that condition (2.11) is satisfied for t sufficiently small, and we have

$$\begin{aligned} T^*\left(\frac{p}{t}\right) &\sim \frac{2t}{\sigma|p|\sqrt{1-\rho^2}}\left[\pi 1_{\{\rho p \le 0\}} + \operatorname{sgn}(p)\arctan\left(\frac{\sqrt{1-\rho^2}}{\rho}\right)\right] \\ &\qquad\qquad \text{as } t \to 0 \quad \text{for } \rho \neq 0,\ \ p \neq 0, \\ &\sim \frac{\pi t}{\sigma|p|} \quad \text{as } t \to 0 \quad \text{for } \rho = 0,\ \ p \neq 0, \\ &= \infty \quad \text{for } p = 0, \end{aligned} \tag{2.12}$$

where we have used the fact that $\arctan(x) \to -\frac{1}{2}\pi$ as $x \to -\infty$ to deal with the case $\rho = 0, p \neq 0$ (also note that we are not letting $\rho \to 0$, but rather we are fixing $\rho = 0$ and letting $t \to 0$). By careful examination of the entries of Table 1 we see that for any $p \in (p_-, p_+)$ we have $T^*(\frac{p}{t}) > t$ for t sufficiently small, so Eq. (2.7) is valid. Letting $t \searrow 0$ and $\theta = \sigma\bar{\rho}p$, we find that

$$\begin{aligned} d\left(\frac{p}{t}\right) &\sim \frac{i\sigma\bar{\rho}p}{t}, \\ g_2\left(\frac{p}{t}\right) &\sim \frac{\rho + i\bar{\rho}}{\rho - i\bar{\rho}}, \\ ty_0 D\left(\frac{p}{t}\right) &\sim -\frac{y_0 p}{\sigma^2}(\rho\sigma + i\sigma\bar{\rho}) \cdot \frac{1 - e^{-i\sigma\bar{\rho}p}}{1 - \frac{\rho + i\bar{\rho}}{\rho - i\bar{\rho}} e^{-i\sigma\bar{\rho}p}}, \\ &= -\frac{y_0 p}{\sigma}(\rho + i\bar{\rho}) \cdot \frac{(\rho - i\bar{\rho})(e^{i\theta/2} - e^{-i\theta/2})}{(\rho - i\bar{\rho})e^{i\theta/2} - (\rho + i\bar{\rho})e^{-i\theta/2}} \\ &= -\frac{y_0 p}{\sigma} \cdot \frac{(\rho + i\bar{\rho})(\rho - i\bar{\rho})(e^{i\theta/2} - e^{-i\theta/2})}{\rho(e^{i\theta/2} - e^{-i\theta/2}) - i\bar{\rho}(e^{i\theta/2} + e^{-i\theta/2})} \\ &= \frac{y_0 p}{\sigma} \cdot \frac{1}{i\bar{\rho}\coth\left(\frac{1}{2}i\theta\right) - \rho} \\ &= \Lambda(p), \\ tC\left(\frac{p}{t}, p\right) &\sim t\frac{\kappa\theta}{\sigma^2}\left[(-\rho - i\bar{\rho})\sigma p - 2\log\left(\frac{1 - \frac{\rho + i\bar{\rho}}{\rho - i\bar{\rho}} e^{-i\theta}}{1 - \frac{\rho + i\bar{\rho}}{\rho - i\bar{\rho}}}\right)\right], \end{aligned} \tag{2.13}$$

where $\bar{\rho} = \sqrt{1-\rho^2}$. Thus

$$\lim_{t \to 0} t \log \mathbb{E}(e^{\frac{p}{t}(X_t - x_0)}) = \Lambda(p) \tag{2.14}$$

for $p \in (p_-, p_+)$ (see Table 1 for a definition of p_- and p_+), where $\Lambda(\cdot)$ is defined in Eq. (2.2). By direct verification, we see that $\Lambda(p) \to \infty$ as $p \nearrow p_+$ and as $p \searrow p_-$. We also need to verify that $\lim_{t\to 0} t \log \mathbb{E}(e^{\frac{p}{t}(X_t - x_0)}) = \infty$ for $p \notin (p_-, p_+)$. But this follows from the monotonicity result in Lemma B.1 when we let $t \to 0$. Thus Assumption A.1 is satisfied. Λ is smooth in (p_-, p_+) and

$$\begin{aligned}\Lambda'(p) &= \frac{y_0}{\sigma(\sqrt{1-\rho^2}\cot\frac{\theta}{2} - \rho)} + \frac{\sigma y_0 p(1-\rho^2)\csc^2\frac{\theta}{2}}{2\sigma(\sqrt{1-\rho^2}\cot\frac{\theta}{2} - \rho)^2},\\ \Lambda''(p) &= \frac{y_0(1-\rho^2)\csc^2\frac{\theta}{2}(1-\frac{\theta}{2}\cot\frac{\theta}{2})}{(\sqrt{1-\rho^2}\cot\frac{\theta}{2} - \rho)^2} + \frac{y_0\theta(1-\rho^2)^{\frac{3}{2}}\csc^4\frac{\theta}{2}}{2(\sqrt{1-\rho^2}\cot\frac{\theta}{2} - \rho)^3}.\end{aligned} \tag{2.15}$$

The first term in Eq. (2.15) is non-negative on (p_-, p_+) because $\frac{\theta}{2}\cot\frac{\theta}{2} \leq 1$ if θ in $(-2\pi, 2\pi)$, which will be the case when p lies in the interval (p_-, p_+) (by careful checking of the entries in the table in Theorem 2.1). The second term is also non-negative because it is a non-negative multiple of $\Lambda(p)$ itself which is non-negative, so $\Lambda''(p) > 0$ for $p \in (p_-, p_+)$. By direct inspection, we also have that $\Lambda(p) \to \infty$ and $|\Lambda'(p)| \to \infty$ as $p \nearrow p_+$ and $p \searrow p_-$. Thus $\Lambda(p)$ is convex, essentially smooth and lower semi-continuous, so (by the Gärtner-Ellis Theorem in Theorem A.1), $X_t - x_0$ satisfies the large deviation principle with rate function equal to the Fenchel-Legendre transform of $\Lambda(p)$. By the essential smoothness property of $\Lambda(p)$, the equation

$$\frac{\partial}{\partial p}(px - \Lambda(p))|_{p^*} = 0 \tag{2.16}$$

has a solution $p^* = p^*(x)$ in (p_-, p_+), which equivalently solves

$$x = \Lambda'(p^*). \tag{2.17}$$

But Λ' is continuous and strictly monotonically increasing on (p_-, p_+), so there is a unique $p^*(x)$. The unique minimum x^* of Λ^* occurs at $x^* = (\Lambda^{*'})^{-1}(0) = \Lambda'(0) = 0$, and $\Lambda^*(0) = 0$. For $p \in (p_-, p_+)$, $\Lambda'(p)$ is negative when $p < 0$, and is positive when $p > 0$. By a convex analysis result , $x = \Lambda'(p)$ for $p \in (p_-, p_+))$ if and only if $p = \Lambda^{*'}(x)$ (see the proof of Lemma 2.5 in Feng *et al.* [10] for a very similar analysis). Consequently, Λ^* is strictly increasing when $x > 0$ and strictly decreasing when $x < 0$, and Eqs. (2.3) and (2.4) follow . □

Remark 2.1. Note that $\Lambda(p)$ does not depend on the drift terms κ or y_∞, and this is typical in the Freidlin-Wentzell theory of small-time large deviations for diffusion processes.

3. Small-Time Behavior of Call Options and Implied Volatility

Two useful corollaries of Theorem 2.1 are the following rare event estimates for pricing out-of-the-money call and put options of small maturity.

Corollary 3.1. *We have the following small-time behavior for out-of-the-money call options on* $S_t = e^{X_t}$

$$-\lim_{t\to 0} t \log \mathbb{E}(S_t - K)^+ = \Lambda^*(x), \tag{3.1}$$

where $x = \log(\frac{K}{S_0}) \geq 0$ *is the log-moneyness.*

Corollary 3.2. *We have the following small-time behavior for out-of-the-money put options on* $S_t = e^{X_t}$

$$-\lim_{t\to 0} t \log \mathbb{E}(K - S_t)^+ = \Lambda^*(x), \tag{3.2}$$

for $x = \log(\frac{K}{S_0}) \leq 0$*.*

Proof. (of Corollary 3.1)

(i) We first deal with the lower bound. From a drawing, we see that for any $\delta > 0$, we have

$$\mathbb{E}(S_t - K)^+ \geq \delta \mathbb{P}(S_t > K + \delta). \tag{3.3}$$

Then by Theorem 2.1 we have that

$$\begin{aligned}\liminf_{t\to 0} t \log \mathbb{E}(S_t - K)^+ &\geq \liminf_{t\to 0} [t \log \delta + t \log \mathbb{P}(S_t > K + \delta)] \\ &\geq -\Lambda^* \left(\log \frac{K+\delta}{S_0}\right). \end{aligned} \tag{3.4}$$

Take $\delta \to 0+$. By continuity of $\Lambda^*(x)$, we have the desired lower bound.

(ii) To obtain the desired upper bound, we note (by Hölder's inequality) that for any $p, q > 1$ with $p^{-1} + q^{-1} = 1$, we have

$$\mathbb{E}(S_t - K)^+ = \mathbb{E}[(S_t - K)^+ 1_{\{S_t \geq K\}}] \leq [\mathbb{E}((S_t - K)^+)^p]^{1/p} \mathbb{E}(1^q_{\{S_t \geq K\}})^{1/q} \tag{3.5}$$

$$= [\mathbb{E}((S_t - K)^+)^p]^{1/p} \mathbb{P}(S_t \geq K)^{1/q}.$$

$$\leq [\mathbb{E}(S_t^p)]^{1/p} \mathbb{P}(S_t \geq K)^{1-1/p}. \tag{3.6}$$

Taking logs and multiplying by t we obtain

$$t \log \mathbb{E}(S_t - K)^+ \leq \frac{t}{p} \log \mathbb{E}(S_t^p) + t\left(1 - \frac{1}{p}\right) \log \mathbb{P}(S_t \geq K). \tag{3.7}$$

We claim that $t \log \mathbb{E}(S_t^p) \to 0$ as $t \to 0$. Consider $\delta > 0$. For t sufficiently small, $\frac{\delta}{t} > p$. Then we have

$$\limsup_{t\to 0} t \log \mathbb{E}\left(\left(\frac{S_t}{S_0}\right)^p\right) \leq \limsup_{t\to 0} t \log \mathbb{E}\left(\left(\frac{S_t}{S_0}\right)^{\delta/t}\right) = \Lambda(\delta), \tag{3.8}$$

by Lemma B.1, and

$$\liminf_{t\to 0} t\log\mathbb{E}\left(\left(\frac{S_t}{S_0}\right)^p\right) \geq \liminf_{t\to 0} t\log\left(\mathbb{E}\left(\frac{S_t}{S_0}\right)\right)^p = 0 = \Lambda(0) \tag{3.9}$$

by Jensen's inequality. But Λ is continuous, so the claim is verified. If we then take $\lim_{p\to\infty}$ on both sides of Eq. (3.7), we have (by Theorem 2.1) the upper bound

$$\limsup_{t\to 0} t\log\mathbb{E}(S_t-K)^+ \leq -\Lambda^*\left(\log\frac{K}{S_0}\right). \tag{3.10}$$

□

The proof of Corollary 3.2 follows by a similar argument.

Using the put-call parity, we can combine Corollaries 3.1 and 3.2 to obtain the following result.

Corollary 3.3. *For all $x\in\mathbb{R}$*

$$\begin{aligned} t\log(\mathbb{E}(S_t-K)^+ - (S_0-K)^+) &= t\log(\mathbb{E}(K-S_T)^+ - (K-S_0)^+)\\ &= \Lambda^*(x), \end{aligned} \tag{3.11}$$

where $x=\log\frac{K}{S_0}$.

We can also compute the asymptotic implied volatility as follows

Theorem 3.1. *We have the following asymptotic behavior for the implied volatility $\sigma_t=\sigma_t(x)$ of a European call option on $S_t=e^{X_t}$, with strike $K=S_0e^x$ and $x\in\mathbb{R}, x\neq 0$, as $t\to 0$*

$$I(x)=\lim_{t\to 0}\sigma_t(x)=\frac{x}{\sqrt{2\Lambda^*(x)}}. \tag{3.12}$$

Proof. We first assume that $x>0$. We first establish the lower bound for $\sigma_t(x)$. Let $n(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$, and $\Phi(z)=\int_{-\infty}^{z} n(u)du$ denote the standard cumulative Normal distribution function. Using the classical notation for the Black-Scholes formula, we set $d_1=\frac{-x+\frac{1}{2}\sigma_t^2 t}{\sigma_t\sqrt{t}}, d_2=d_1-\sigma_t\sqrt{t}$. By Corollary 3.1 and the definition of implied volatility, we know that for all $\delta, C>0$, there exists a $t^*=t^*(\delta)>0$ such that for all $t<t^*$ we have

$$\begin{aligned} e^{-(\Lambda^*(x)+\delta)/t} &\leq \mathbb{E}(S_t-K)^+\\ &= S_0\Phi(d_1)-S_0e^x\Phi(d_2)\\ &\leq S_0\Phi^c(-d_1)\\ &\leq S_0\cdot\frac{1}{|d_1|}n(d_1)\\ &\leq S_0C\cdot n(d_1), \end{aligned} \tag{3.13}$$

The last two lines follow from the standard estimate on the normal distribution in section 14.8 of Williams [25] (see Appendix C), and the fact that the *dimensionless* *implied variance* $\sigma_t^2 t \to 0$ as $t \to 0$, because the call price tends to zero as $t \to 0$. Taking logs of both sides and multiplying by t, we can find a $t^{**} = t^{**}(\delta) \le t^*$ such that for all $t < t^{**}$ we have

$$-(\Lambda^*(x) + \delta) \le t \log(S_0 C) - \frac{t}{2}\log(2\pi) - \frac{(x - \frac{1}{2}\sigma_t^2 t)^2}{2\sigma_t^2} \le -\frac{x^2}{2\sigma_t^2} + \delta,$$

and the lower bound follows. For the upper bound, consider $\delta > 0$. Again, by Corollary 3.1, there exists an $t(\delta) > 0$ such that for all $t < t(\delta)$ we have

$$\begin{aligned} e^{-(\Lambda^*(x)-\delta)/t} &\ge \mathbb{E}(S_t - K)^+ \\ &= S_0\Phi(d_1) - S_0 e^x \Phi(d_2) \\ &= S_0\Phi^c(-d_1) - S_0 e^x \Phi^c(-d_2). \end{aligned} \tag{3.14}$$

Set $d_{2,\delta} = \frac{-x-\delta-\frac{1}{2}\sigma_t^2 t}{\sigma_t\sqrt{t}}$. Similar to Corollary 3.1 or by drawing a picture, we can now use the fact that one call option of strike K is worth more than $K(e^\delta - 1)$ digital call options of strike Ke^δ, and invoke the lower bound on the normal distribution in Appendix C and the fact that $\sigma_t^2 t \to 0$ as $t \to 0$ to see that

$$\begin{aligned} e^{-(\Lambda^*(x)-\delta)/t} &\ge K(e^\delta - 1)\Phi^c(-d_{2,\delta}) \\ &\ge K(e^\delta - 1)n(d_{2,\delta})\left(\frac{1}{|d_{2,\delta}|} + |d_{2,\delta}|\right)^{-1} \\ &\ge K(e^\delta - 1)n(d_{2,\delta}) \cdot \frac{\sigma_t\sqrt{t}}{x+\delta}(1-\delta) \\ &\ge K(e^\delta - 1)n(d_{2,\delta}) \cdot e^{-\delta\frac{x+\delta}{\sigma_t\sqrt{t}}}(1-\delta). \end{aligned} \tag{3.15}$$

Taking logs of both sides and multiplying by t, we can further find a $t^{**} \le t^*$ such that for all $t < t^{**}$ we have

$$\begin{aligned} -(\Lambda^*(x) - \delta) \ge t\log(K) + t\log(e^\delta - 1) - \frac{t}{2}\log(2\pi) - \frac{(x+\delta+\frac{1}{2}\sigma_t^2 t)^2}{2\sigma_t^2} \\ - \sqrt{t}\delta\frac{(x+\delta)}{\sigma_t} + t\log(1-\delta) \\ \ge -\frac{(x+\delta)^2}{2\sigma_t^2} - \delta. \end{aligned} \tag{3.16}$$

We proceed similarly for $x < 0$. □

3.1. *The level, slope and curvature of the small-time implied volatility at-the-money*

Note that Theorem 3.1 does not deal with the at-the-money case $x = 0$, because the LDP bounds are useless in this case. Here, we can either first set $x = 0$, and then let $t \to 0$, or vice versa. The former requires a central limit theorem type

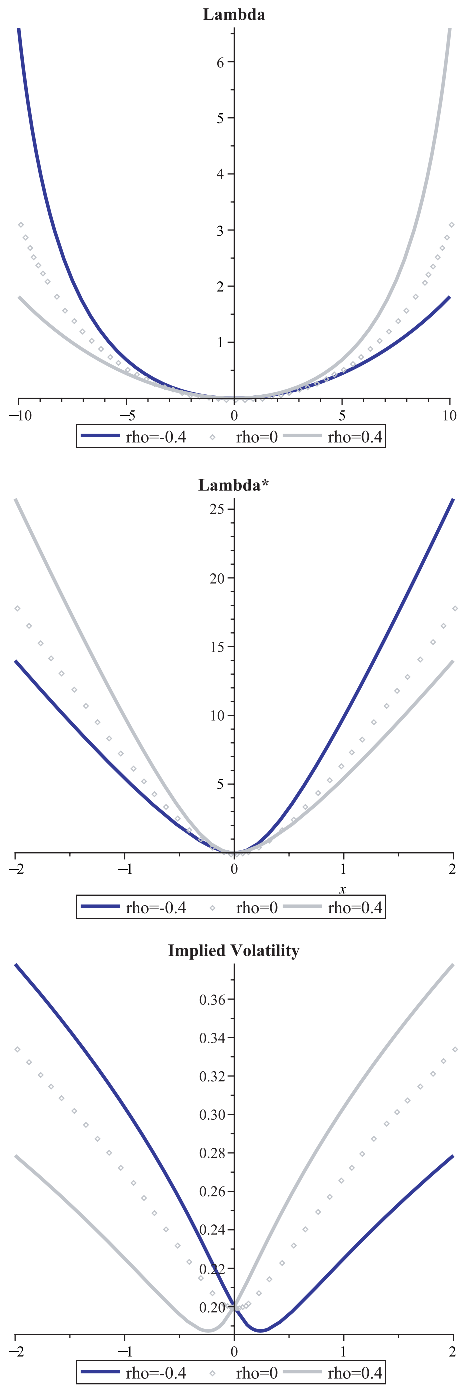

Fig. 1. Here we have plotted $\Lambda(p)$, the rate function $\Lambda^*(x)$ and the asymptotic implied volatility $I(x)$ for the Heston model with the parameters $y_0 = .04 = y_\infty = .04, \sigma = .20$ and $\rho = -0.4, 0.0.4$. We have used the bisection method to numerically compute the Legendre transform $\Lambda^*(x)$.

argument using Lévy's convergence theorem which does not involve large deviations theory; we omit the details for brevity. For the latter, we can appeal to the following result.

Theorem 3.2. *The asymptotic implied volatility $I(x)$ has the following expansion around $x = 0$*

$$I(x) = \sqrt{y_0}\left[1 + \frac{1}{4}\rho z + \left(\frac{1}{24} - \frac{5}{48}\rho^2\right) z^2 + O(z^3)\right], \tag{3.17}$$

where $z = \frac{\sigma x}{y_0}$.

Remark 3.1. Note that $I''(0) < 0$ for $\rho^2 > \frac{2}{5}$.

Proof. Λ' is a diffeomorphism on (p_-, p_+) and near $p = 0$, we have

$$\begin{aligned}\Lambda(p) &= \frac{1}{2}y_0 p^2 + \frac{1}{4}y_0\rho\sigma p^3 + \frac{1}{24}y_0\sigma^2(1 + 2\rho^2)p^4 + O(p^5),\\ \Lambda'(p) &= y_0 p + \frac{3}{4}y_0\rho\sigma p^2 + \frac{1}{6}y_0\sigma^2(1 + 2\rho^2)p^3 + O(p^4).\end{aligned} \tag{3.18}$$

Recall the following expressions for the first and second derivatives of the inverse of a smooth, invertible function $y(x)$

$$\begin{aligned}\frac{dy}{dx} &= \frac{1}{\frac{dx}{dy}}.\\ \frac{d^2y}{dx^2} &= -\frac{d^2x}{dy^2}\left(\frac{dy}{dx}\right)^3.\end{aligned} \tag{3.19}$$

Using these equations and setting $x = \Lambda'(p^*(x))$, we find that

$$\begin{aligned}p^{*'}(0) &= \frac{1}{\Lambda''(\Lambda'^{-1}(0))} = \frac{1}{\Lambda''(0)} = \frac{1}{y_0},\\ p^{*''}(0) &= -\frac{\Lambda'''(0)}{(\Lambda''(0))^3} = -\frac{3\rho\sigma}{2y_0^2}.\end{aligned} \tag{3.20}$$

Thus

$$p^*(x) = \frac{1}{y_0}x - \frac{3\rho\sigma}{4y_0^2}x^2 + O(x^3), \tag{3.21}$$

and

$$\begin{aligned}I(x) &= \frac{x}{\sqrt{2\Lambda^*(x)}} = \frac{1}{\sqrt{2(p^*(x)x - \Lambda(p^*(x)))}}\\ &= \sqrt{y_0}\left[1 + \frac{1}{4}\rho z + \left(\frac{1}{24} - \frac{5}{48}\rho^2\right) z^2 + O(z^3)\right].\end{aligned} \tag{3.22}$$

□

Remark 3.2. The $O(x^3)$ term for $p^*(x)$ does not affect the final answer for $I(x)$ up to $O(z^3)$.

Remark 3.3. If we set $V(x) = I(x)^2$, then

$$V(x) = y_0 \left[1 + \frac{1}{2}\rho z + \left(\frac{1}{12} - \frac{7}{48}z^2\right) + O(z^3)\right] \tag{3.23}$$

and this agrees with the formula on page 127 in Lewis [19], and section 3.1.2. in Durrleman [8].

Appendix

A. *The Gärtner-Ellis theorem*

Consider a sequence of random variables $X_n \in \mathbb{R}^d$, where X_n possesses the law μ_n and logarithmic moment generating function

$$\Lambda_n(\lambda) = \log \mathbb{E}(e^{\langle \lambda, X_n \rangle}). \tag{A.1}$$

Definition A.1 (Definition 2.3.5 in Dembo and Zeitouni [7]). Let $\mathcal{D}_\Lambda = \{\lambda \in \mathbb{R}^d : \Lambda(\lambda) < \infty\}$. A convex function $\Lambda : \mathbb{R}^d \to (-\infty, \infty]$ is said to be essentially smooth if

- The interior $\mathcal{D}^0_\Lambda$ of $\mathcal{D}_\Lambda$ is non-empty.
- $\Lambda(\cdot)$ is differentiable throughout $\mathcal{D}^0_\Lambda$.
- $\Lambda(\cdot)$ is steep, namely $\lim_{n\to\infty} |\nabla\Lambda(\lambda_n)| = \infty$ whenever $\{\lambda_n\}$ is a sequence in $\mathcal{D}^0_\Lambda$ converging to a boundary point of $\mathcal{D}^o_\lambda$.

Assumption A.1 (Assumption 2.3.2 in Dembo and Zeitouni [7]). For each $\lambda \in \mathbb{R}^d$, we assume that the logarithmic moment generating function, defined as the limit

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda) \tag{A.2}$$

exists as an extended real number. Further, the origin belongs to the interior of $\mathcal{D} = \{\lambda \in \mathbb{R}^d : \Lambda(\lambda) < \infty\}$.

We now recall the Gärtner-Ellis theorem:

Theorem A.1 (see Theorem 2.3.6 in Dembo and Zeitouni [7]). *Let Assumption* A.1 *hold. Then if* $\Lambda(\lambda)$ *is essentially smooth and lower semicontinuous, the sequence of random variables* X_n *satisfies the large deviation principle with rate function* $\Lambda^*(x)$, *which is the Fenchel-Legendre transform of* F, *defined by the variational formula*

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - \Lambda(\lambda)). \tag{A.3}$$

Lemma A.2 (Lemma 2.3.9 (a) in Dembo and Zeitouni [7]). *Let Assumption* A.1 *hold. Then* Λ *is a convex function,* $\Lambda(\lambda) > -\infty$ *everywhere, and* $\Lambda^*(x)$ *is a good convex rate function. In one dimension, this means that* $\Lambda^*(x)$ *is non-decreasing in the region* $x > 0$.

Remark A.1. Since $\mu_n(\Omega) = 1$, it is necessary that $\inf_x I(x) = 0$. If $\Lambda^*(x)$ is a good rate function, then the level set $\{x : \Lambda^*(x) \le \alpha\}$ is compact, so we know that the infimum is attained on this compact set because Λ^* is lower semicontinuous, i.e., there exists at least one point x^* for which $\Lambda^*(x^*) = 0$ (see pages 5 and 6 in Dembo and Zeitouni [7]).

B. *Monotonicity of the logarithmic transform*

Lemma B.1. *Let $S_t = e^{X_t}$ be a non-negative martingale with $S_0 = 1$, and let $|p_1| < |p_2|$ with* $\operatorname{sgn}(p_1) = \operatorname{sgn}(p_2)$. *Then*

$$\mathbb{E}(e^{\frac{p_1}{t}X_t}) \le \mathbb{E}(e^{\frac{p_2}{t}X_t}) \tag{B.1}$$

for t sufficiently small.

Proof. (i) We first consider $0 < p_1 < p_2 < \infty$. Let $S_t = e^{X_t}$ and assume S_t is a martingale and $S_0 = 1$. For t sufficiently small, by the monotonicity of the $\mathcal{L}^p$ norm, we have

$$\mathbb{E}(S_t^{\frac{p_1}{t}})^{\frac{t}{p_1}} < \mathbb{E}(S_t^{\frac{p_2}{t}})^{\frac{t}{p_2}}, \tag{B.2}$$

or

$$\mathbb{E}(S_t^{\frac{p_1}{t}}) < \mathbb{E}(S_t^{\frac{p_2}{t}})^{\frac{p_1}{p_2}}. \tag{B.3}$$

For t sufficiently small, $\frac{p_2}{t} > 1$, so by Jensen's inequality, we have

$$\mathbb{E}(S_t^{\frac{p_2}{t}}) \ge \mathbb{E}(S_t)^{\frac{p_2}{t}} = 1. \tag{B.4}$$

Thus

$$\mathbb{E}(S_t^{\frac{p_2}{t}})^{\frac{p_1}{p_2}} \le \mathbb{E}(S_t^{\frac{p_2}{t}}), \tag{B.5}$$

and we are done.

(ii) We now consider $-\infty < p_2 < p_1 < 0$. We wish to prove that

$$\mathbb{E}(e^{\frac{p_1}{t}X_t}) \le \mathbb{E}(e^{\frac{p_2}{t}X_t}) \tag{B.6}$$

for t sufficiently small. For t sufficiently small, by the monotonicity of the $\mathcal{L}^p$ norm, we have

$$\mathbb{E}\left(\left(\frac{1}{S_t}\right)^{\frac{|p_1|}{t}}\right)^{\frac{t}{|p_1|}} < \mathbb{E}\left(\left(\frac{1}{S_t}\right)^{\frac{|p_2|}{t}}\right)^{\frac{t}{|p_2|}}, \tag{B.7}$$

or

$$\mathbb{E}\left(\left(\frac{1}{S_t}\right)^{\frac{|p_1|}{t}}\right) < \mathbb{E}\left(\left(\frac{1}{S_t}\right)^{\frac{|p_2|}{t}}\right)^{\frac{|p_1|}{|p_2|}}. \tag{B.8}$$

For t sufficiently small, $\frac{|p_2|}{t} > 1$, so by Jensen's inequality, we have

$$\mathbb{E}\left(\left(\frac{1}{S_t}\right)^{\frac{|p_2|}{t}}\right) \geq \mathbb{E}\left(\frac{1}{S_t}\right)^{\frac{|p_2|}{t}} \geq 1, \tag{B.9}$$

because $\frac{1}{S_t}$ is a submartingale. Thus

$$\mathbb{E}\left(\left(\frac{1}{S_t}\right)^{\frac{|p_2|}{t}}\right)^{\frac{|p_1|}{|p_2|}} \leq \mathbb{E}\left(\left(\frac{1}{S_t}\right)^{\frac{|p_2|}{t}}\right), \tag{B.10}$$

and we are done. □

C. *Estimates for the standard normal distribution function*

Let $x > 0$, and $n(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. Then we have the following estimate for $\Phi^c(x) = \int_x^\infty n(z)dz$

$$\left(x + \frac{1}{x}\right)^{-1} n(x) \leq \Phi^c(x) \leq \frac{1}{x} n(x) \tag{C.1}$$

(see section 14.8 in Williams [25]).

Acknowledgments

The work of Forde has been supported by the European Science Foundation, AMaMeF Exchange Grant 2107. The authors would like to thank John Appleby, Jin Feng, Jean-Pierre Fouque, Alan Lewis, Roger Lee and Aleksandar Mijatović for useful discussions.

References

[1] E. Alòs and C. O. Ewald, Malliavin differentiability of the Heston volatility and applications to option pricing, *Advances in Applied Probability* **40**(1) (2008) 144–162.
[2] H. Albrecher, P. Mayer, W. Schoutens and J. Tistaert, The little Heston trap, *Wilmott Magazine*, **January** (2007) 83–92.
[3] L. B. G. Andersen and V. V. Piterbarg, Moment Explosions in Stochastic Volatility Models, *Finance and Stochastics* (2007).
[4] H. Berestycki, J. Busca and I. Florent, Asymptotics and calibration of local volatility models, *Quantitative Finance* Vol. 2 (2002) 61–69.
[5] H. Berestycki, J. Busca and I. Florent, Computing the implied volatility in stochastic volatility Models, *Communications on Pure and Applied Mathematics*, **LVII** (2004) 1352–1373.
[6] M. G. Crandall and P. L. Lions, Viscosity solutions of Hamilton-Jacobi equations, *Trans AMS* **277** (1984) 1–42.
[7] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications* (Jones and Bartlet Publishers, Boston, 1998).

[8] V. Durrleman, From implied to spot volatilities, PhD dissertation, Princeton University (2004)
[9] V. Durrleman, From implied to spot volatilities, working paper (2005).
[10] J. Feng, M. Forde and J. P. Fouque, Short maturity asymptotics for a fast mean-reverting Heston stochastic volatility model, submitted (2009).
[11] W. H. Fleming and H. Mete Soner, *Controlled Markov Processes and Viscosity Solutions* (Springer-Verlag, New York, 1993).
[12] J. Feng and T. G. Kurtz, *Large Deviation for Stochastic Processes*, Mathematical Surveys and Monographs, Vol. 131 (American Mathematical Society, 2006).
[13] P. Hagan, D. Kumar, A. S. Lesniewski and D. E. Woodward, Managing Smile Risk, *Wilmott Magazine* (2002).
[14] P. Henry-Labordère, A general asymptotic implied volatility for stochastic volatility models, working paper (2005).
[15] P. Henry-Labordère, Combining the SABR and LMM models, *Risk* (2007).
[16] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus* (Springer-Verlag, 1991).
[17] M. Keller-Ressel, Moment explosions and long-term behavior of affine stochastic volatility models, forthcoming in *Mathematical Finance* (2008).
[18] P. Laurence, Implied Volatility, Fundamental solutions, asymptotic analysis and symmetry methods, presentation at Linz, Ricam kick-off Workshop (2008).
[19] A. Lewis, *Option Valuation Under Stochastic Volatility* (Finance Press, 2000).
[20] A. Lewis, Geometries and smile asymptotics for a class of stochastic volatility models, www.optioncity.net (2007).
[21] E. Lukacs, *Characteristic Functions*, second ed. (Griffin, London, 1970).
[22] A. Medvedev & O. Scaillet, A simple calibration procedure of stochastic volatility models with jumps by short term asymptotic, FAME Research Paper Series rp93, International Center for Financial Asset Management and Engineering (2004).
[23] S. R. S. Varadhan, On the behavior of the fundamental solution of the heat equation with variable coefficients, *Comm. Pure Appl. Math.* **20** (1967) 431–455.
[24] S. R. S. Varadhan, Diffusion processes in a small time interval, *Comm. Pure Appl. Math.* **20** (1967) 659–685.
[25] D. Williams, *Probability with Martingales* (Cambridge Mathematical Textbooks, 1991).