Mathematical Finance, Vol. 22, No. 4 (October 2012), 591–620

ASYMPTOTICS OF IMPLIED VOLATILITY IN LOCAL VOLATILITY MODELS

JIM GATHERAL

Baruch College, CUNY

ELTON P. HSU

Northwestern University

PETER LAURENCE

Università di Roma and New York University

CHENG OUYANG

Purdue University

TAI-HO WANG

Baruch College, CUNY

Using an expansion of the transition density function of a one-dimensional time inhomogeneous diffusion, we obtain the first- and second-order terms in the short time asymptotics of European call option prices. The method described can be generalized to any order. We then use these option prices approximations to calculate the first- and second-order deviation of the implied volatility from its leading value and obtain approximations which we numerically demonstrate to be highly accurate.

KEY WORDS: implied volatility, local volatility, asymptotic expansion, heat kernels.

1. INTRODUCTION

Local volatility models are powerful tools for modeling the price evolution of a financial asset consistent with known prices of European options on that asset. In the driftless case applicable to the evolution of the asset price under the forward measure, local volatility models take the form

$$df_t = a(f_t, t)\, dW_t, \tag{1.1}$$

We thank the anonymous referees for their helpful comments and suggestions to improve the paper and the interesting points raised. All errors are our responsibility. THW is partially supported by the grant PSC-CUNY40-60125-39 of The City University of New York.

Manuscript received November 2009; final revision received June 2010.

Address correspondence to Peter Laurence, Dipartimento di Matematica, Università di Roma 1 "La Sapienza," Rome, Italy; e-mail: laurence@courant.nyu.edu.

DOI: 10.1111/j.1467-9965.2010.00472.x

where W_t is a Brownian motion. In general, known European option prices can be matched only by allowing an explicit dependence of the volatility of f_t on t. We thus allow f_t to evolve according to a *time-inhomogeneous* diffusion.

From a historical perspective, one of the earliest examples of a local volatility model is the constant elasticity of variance (CEV) model introduced by Cox (1975), in which $a(f) = f^{\beta}$, $\beta < 1$. Rubinstein showed that the CEV model could be used to match single-maturity smiles in index option markets with fitted values of β typically less than zero. The CEV model and its close relative the CIR model (Cox, Ingersoll, and Ross), as well as their multidimensional generalizations, continue to be very popular to this day. Another special but quite flexible local volatility model that has gained increasing popularity recently is the quadratic model explored by Zuhlsdorff (2001), Lipton (2002), Andersen (2008), and others.

In pioneering contributions to mathematical finance, Dupire (1994) and Derman and Kani (1994) showed that, given a set of European options with a continuum of strikes and expirations, it is possible to back out a local volatility function such that an asset evolving according to equation (1.1) will generate call prices that exactly match the prices of these options. Equivalently, given an implied volatility surface (a function returning implied volatilities for all strikes and expirations), the local volatility function may be computed using a simple formula, see for example (1.1) in Gatheral (2006). Dupire (1994) and Derman and Kani (1994) further showed that local variance (the square of local volatility) may be represented as a conditional expectation of instantaneous variance in a stochastic volatility model.

Shortly thereafter Dupire (2004) pointed out that in the context of *European* option pricing, local volatility models also arise naturally, as a mean to replace a multifactor stochastic volatility model for the asset f_t and a possibly stochastic volatility and interest rate, by a less complex one-factor local volatility model, generating identical European option prices. It turned out that this idea had, outside of mathematical finance, been independently introduced in an even more general form by Gyöngi (1986). Piterbarg (2007) took this one step further by offering a systematic set of tools (dubbed *Markovian projection*) to generate an approximate local volatility model given any specific stochastic volatility model. Piterbarg then showed how Markovian projection combined with his parameter averaging technique can permit the efficient calibration of complex models of asset dynamics to European option prices. In the same paper, Piterbarg subsequently applied these techniques to the efficient approximation of the prices of spread and basket options. Using an alternative and typically more accurate approach, Avellaneda et al. (2002) showed how to generate a one-factor local volatility function that is effective in reproducing the option prices of a multidimensional index governed by a multidimensional local volatility model. In summary, even though local volatility models as such are not considered reasonable models for asset dynamics, they do arise naturally in the context of fast approximations to option prices in more complex (and presumably more reasonable) models of asset dynamics. A central question is then how to efficiently approximate implied volatilities given a local volatility function.

In this paper, we show how to apply the heat kernel expansion to the approximation of implied volatilities in a local volatility model when time to expiration is small. This geometric technique, based on a natural Riemannian metric, was introduced into mathematical finance in a Courant Institute lecture by Lesniewski (2002). Thereafter it has been considerably extended and developed by Henry-Labordère (2005). The contribution of our paper is two-fold. First, we provide a rigorous derivation of this expansion up to first order in the time to expiration. Second, we take the analysis one step further and determine, in addition to the first-order correction, the second-order correction in the case of interest rate $r = 0$, and moreover allowing time-dependent coefficients. The second-order correction in the case of interest rate $r \neq 0$ can then readily be obtained by a procedure we will describe in brief.

In the case when the volatility does not depend explicitly on time (time-homogeneous models), our lowest order approximation to implied volatility agrees with the well-known formula of Berestycki, Busca, and Florent (2002). When the local volatility does depend explicitly on time (time-inhomogeneous models), we find that the formula for the zeroth-order term requires a small but key correction relative to the prior formulae of Hagan, Lesniewski, and Woodward (2004) and Henry-Labordère (2005).

Even in the time-homogeneous case, our first-order correction term is different from and more accurate than the ones in Hagan and Woodward (1999), Hagan et al. (2002), Hagan, Lesniewski, and Woodward (2004), and Henry-Labordère (2005). In all of these earlier contributions the first-order correction involves the first- and second-order derivatives of the local volatility function. Our first-order correction on the other hand does not involve any derivatives of the local volatility function. In addition, we show rigorously that our first-order correction actually corresponds to the first-order derivative of the implied volatility with respect to the time to expiration, and similarly for the second-order correction. This characterization implies that for very small times to maturity the formula is *optimal*, being the unique limit of the ratio of difference quotients, i.e., the ratio of change in implied volatility over a small time interval to the length of the time interval. Our numerical experiments show that there is a clear gain in accuracy even for times that are not small.

After this work was completed, it was brought to our attention that our first-order correction, specialized to the time-homogeneous case and to the case of interest rate $r = 0$, had already been presented by Henry-Labordère (2008). There it is obtained by a heuristic procedure, in which the nonlinear equation of Berestycki, Busca, and Florent (2002) satisfied by the implied volatility is expanded in powers of the time to maturity. Our approach allows us to rigorously justify this formula. More recently still, we learned from Leif Andersen that the same formula, restricted to the time-homogeneous case appears in a joint paper with Brotherton-Ratcliffe (see proposition 1 in Andersen and Brotherton-Ratcliffe 2001).

The form of the first-order correction we give in the case $r \neq 0$ appears to be new, as does the formula for the first-order term σ_1 in the case of time-inhomogeneous diffusions. In the simplest case, when the interest rate $r = 0$ and time-homogeneous diffusions, the first-order correction is given by

$$\hat{\sigma}_1 = \frac{\hat{\sigma}_0^3}{(\ln K - \ln s)^2} \ln \frac{\sqrt{\sigma(s)\sigma(K)}}{\hat{\sigma}_0(T, s)},$$

where

$$\hat{\sigma}_0 = \left[\frac{1}{\ln K - \ln s} \int_s^K \frac{du}{u\sigma(u)} \right]^{-1}$$

is the leading term of the implied volatility.

Besides the results we present here, the only other rigorous results leading to a justification of the zeroth-order approximation in local and stochastic volatility models we are aware of were provided by Berestycki, Busca, and Florent (2002, 2004). Medvedev and Scaillet (2007) and Henry-Labordère (2008) have shown how to obtain similar results by matching the coefficients of the powers of the time to expiration in the nonlinear partial differential equation satisfied by the implied volatility. Also, the work of Kunimoto and Takahashi (2001), beginning with a series of papers, was based on a rigorous perturbation theory of Malliavin–Watanabe calculus. Takahashi, Takehara, and Toda (2009) have recently applied this approach to the λ-Sabr model.

We derive our results up to first order in the time-homogeneous case by two different methods, both ultimately resting on the same heat kernel expansion. Indeed the heat kernel expansion has since its inception admitted both a probabilistic approach, pioneered by Molchanov (1975), and an analytic one going back to the work of Hadamard (1952) and Minakshisundaram and Pleijel (1949). The latter approach was refined in a form more suitable to our purposes a few years later by Yoshida (1953). The probabilistic method is described in Section 2. The second analytic approach is described more briefly in Section 3, can also be made rigorous in the case of nondegenerate diffusions, by coupling the "geometric expansion" with the Levi parametrix method to control the tail of the series. As mentioned above, the PDE approach in this form was first discovered by Yoshida (1953).

In Section 2, we only carry out the expansion to order one for time-homogeneous diffusions. However the second PDE approach of Section 3 is computationally more straightforward, so we use it to compute the first- and second-order corrections for the call prices and for the implied volatilities in the more general setting of time-inhomogeneous diffusions. An Appendix is devoted to the extension of the results in the body of the paper to the case, which is almost universal in mathematical finance, where the volatility is degenerate on the boundary. The probabilistic approach is readily extended to this case using the so-called principle of not feeling the boundary, first formulated by Kac (1951). Such a relatively straightforward extension to the degenerate case is a strong point of the probabilistic approach. Although Yoshida's analytic approach described in Section 3 can easily be made rigorous in the case of a nondegenerate diffusion, it appears quite difficult to extend those proofs to the degenerate setting. However, in the end both approaches must lead to the same formulae, resting as they do on the heat kernel expansion. And in addition to being computationally more convenient, we feel strongly that Yoshida's approach deserves to be more widely known in the mathematical finance community. A numerical Section 4 explores the accuracy of our results, comparing our approximations with those appearing in the existent literature.

2. PROBABILISTIC APPROACH

2.1. Call Price Expansion

Suppose that the dynamics of the stock price S is given by

$$dS_t = S_t \{r\, dt + \sigma(S_t)\, dW_t\}.$$

Then the stochastic differential equation for the logarithmic stock price process $X = \ln S$ is

$$dX_t = \eta(X_t)\, dW_t + \left[r - \frac{1}{2}\eta(X_t)^2\right] dt,$$

where $\eta(x) = \sigma(e^x)$. Denote by $c(t, s)$ the price of the European call option (with expiry T and strike price K understood) at time t and stock price s in the local volatility model.

It satisfies the following Black–Scholes equation

$$c_t + \frac{1}{2}\sigma(s)^2 s^2 c_{ss} + rsc_s - rc = 0, \quad c(T, s) = (s - K)^+.$$

It is more convenient to work with the function

$$v(\tau, x) = c(T - \tau, e^x), \quad (\tau, x) \in [0, T] \times \mathbb{R},$$

where $\tau = T - t$ is the time to expiration. The reason is that for this function the Black–Scholes equation takes the simpler following form

$$v_\tau = \frac{1}{2}\eta(x)^2 v_{xx} + \left[r - \frac{1}{2}\eta(x)^2\right] v_x - rv, \quad v(0, x) = (e^x - K)^+.$$

We now study the asymptotic behavior of the modified call price function $v(\tau, x)$ as $\tau \downarrow 0$. Our basic technical assumption is as follows. There is a positive constant C such that for all $x \in \mathbb{R}$,

$$C^{-1} \le \eta(x), \; |\eta'(x)| \le C, \; |\eta''(x)| \le C. \tag{2.1}$$

Because $\eta(x) = \sigma(e^x)$, the above assumption is equivalent to the assumption that there is a constant C such that for all $s \ge 0$,

$$C^{-1} \le \sigma(s) \le C, \; |s\sigma'(s)| \le C, \; |s^2\sigma''(s)| \le C.$$

For many popular models (e.g., CEV model), these conditions may not be satisfied in a neighborhood of the boundary points $s = 0$ and $s = \infty$. However, because we only consider the situation where the stock price and the strike have fixed values other than these boundary values, or more generally vary in a bounded closed subinterval of $(0, \infty)$, the behavior of the coefficient functions in a neighborhood of the boundary points will not affect the asymptotic expansions of the transition density and the call price. Therefore, we are free to modify the values of σ in a neighborhood of $s = 0$ and $s = \infty$ so that the above conditions are satisfied. The principle of not feeling the boundary (see Appendix A) shows that such a modification only produces an exponentially negligible error which will not show up in the relevant asymptotic expansions.

PROPOSITION 2.1. *Let $X = \ln S$ be the logarithmic stock price. Denote the density function of X_t by $p_X(\tau, x, y)$. Then as $\tau \downarrow 0$,*

$$p_X(\tau, x, y) = \frac{u_0(x, y)}{\sqrt{2\pi\tau}} e^{-\frac{d^2(x,y)}{2\tau}} [1 + O(\tau; x, y)].$$

Here

$$d(x, y) = \int_x^y \frac{du}{\eta(u)}$$

and

$$u_0(x, y) = \eta(x)^{1/2}\eta(y)^{-3/2} \exp\left[-\frac{1}{2}(y - x) + r\int_x^y \frac{du}{\eta(u)^2}\right]. \tag{2.2}$$

Furthermore, the remainder satisfies the inequality $|O(\tau; x, y)| \le C\tau$ for a constant C independent of x and y.

Proof. Recall that $\{X_t\}$ satisfies the following stochastic differential equation

$$dX_t = \eta(X_t)dW_t + \left[\frac{1}{2}\eta(X_t) - r\right] dt,$$

where $\eta(x) = \sigma(e^x)$. Introduce the function

$$f(z) = \int_x^z \frac{du}{\eta(u)}$$

and let $Y_t = f(X_t)$. Then

$$dY_t = dW_t - h(Y_t)\, dt.$$

Here

$$h(y) = \frac{\eta \circ f^{-1}(y) + \eta' \circ f^{-1}(y)}{2} - \frac{r}{\eta \circ f^{-1}(y)}.$$

It is enough to study the transition density function $p_Y(\tau, x, y)$ of the process Y.

Introduce the exponential martingale

$$Z_\tau = \exp\left[\int_0^\tau h(Y_s)\, dW_s - \frac{1}{2}\int_0^\tau h(Y_s)^2 ds\right]$$

and a new probability measure $\tilde{\mathbb{P}}$ by $d\tilde{\mathbb{P}} = Z_T d\mathbb{P}$. By Girsanov's theorem, the process Y is a standard Brownian motion under $\tilde{\mathbb{P}}$. For any bounded positive measurable function φ we have

$$\int_\Omega \varphi(Y_\tau) d\tilde{\mathbb{P}} = \int_\Omega \varphi(Y_\tau) Z_\tau\, d\mathbb{P}.$$

Hence, denoting $\mathbb{E}_{x,y}\{\cdot\} = \mathbb{E}_x\{\cdot \,|\, Y_\tau = y\}$, we have

$$\int \varphi(y) p(\tau, x, y)\, dy = \int \varphi(y) \mathbb{E}_{x,y}(Z_\tau) p_Y(\tau, x, y)\, dy,$$

where

$$p(\tau, x, y) = \frac{1}{\sqrt{2\pi\tau}} \exp\left[-\frac{(y-x)^2}{2\tau}\right]$$

is the transition density function of a standard one-dimensional Brownian motion. It follows that

$$\frac{p(\tau, x, y)}{p_Y(\tau, x, y)} = \mathbb{E}_{x,y}(Z_\tau). \tag{2.3}$$

For the conditional expectation of Z_τ, we have

$$\begin{aligned}
\mathbb{E}_{x,y} Z_\tau &= \mathbb{E}_{x,y} \exp\left[\int_0^\tau h(Y_s) \circ dY_s - \frac{1}{2}\int_0^\tau \left\{h'(Y_s) - h(Y_s)^2\right\} ds\right] \\
&= e^{H(y)-H(x)} \mathbb{E}_{x,y} \exp\left[-\frac{1}{2}\int_0^\tau \left\{h'(Y_s) - h(Y_s)^2\right\} ds\right],
\end{aligned}$$

where $H'(y) = h(y)$. The relation (2.3) now reads as

$$\frac{p(\tau, x, y)e^{H(y)}}{p_Y(\tau, x, y)e^{H(x)}} = \mathbb{E}_{x,y} \exp\left[-\frac{1}{2}\int_0^\tau \left\{h'(Y_s) - h(Y_s)^2\right\} ds\right]. \tag{2.4}$$

From the assumption (2.1) it is easy to see that the conditional expectation above is of the form $1 + O(\tau; x, y)$ and $|O(\tau; x, y)| \leq C\tau$ with a constant C independent of x and y. From

$$h(y) = \frac{\eta \circ f^{-1}(y) + \eta_x \circ f^{-1}(y)}{2} - \frac{r}{\eta \circ f^{-1}(y)}$$

we have

$$H(y) - H(x) = \frac{f^{-1}(y) - f^{-1}(x)}{2} + \frac{1}{2}\ln\frac{\eta(f^{-1}(y))}{\eta(f^{-1}(x))} - \int_{f^{-1}(x)}^{f^{-1}(y)} \frac{r}{\eta^2(v)} dv.$$

This together with (2.4) gives us the asymptotics $p_Y(\tau, x, y)$ of Y_t. Once $p_Y(\tau, x, y)$ is found, it is easy to convert it into the density of $X_t = f^{-1}(Y_t)$ using the relation

$$p_X(\tau, x, y) = p_Y(\tau, f(x), f(y))f'(y). \qquad \square$$

Now we compute the leading term of the call price $v(\tau, x) = c(T - \tau, e^x)$ as $\tau \downarrow 0$. For the heat kernel $p_X(\tau, x, y)$ itself, the leading term is

$$\frac{u_0(x, y)}{\sqrt{2\pi\tau}} \exp\left[-\frac{d(x, y)^2}{2\tau}\right].$$

The in-the-money case $s > K$ yielding nothing extra (see Section 2.3), we only consider the out-of-the-money case $s < K$, or equivalently, $x < \ln K$. We express the call price function $v(\tau, x)$ in terms of the density function $p_X(\tau, x, y)$ of X_t. We have

$$\begin{aligned} v(\tau, x) &= c(T - \tau, e^x) \\ &= e^{-r\tau}\mathbb{E}\left[(S_T - K)^+ \mid S_{T-\tau} = e^x\right] \\ &= e^{-r\tau}\mathbb{E}\left[(S_\tau - K)^+ \mid S_0 = e^x\right] \\ &= e^{-r\tau}\mathbb{E}_x[(e^{X_\tau} - K)^+]. \end{aligned}$$

In the third step we used the Markov property of S. Therefore,

$$\begin{aligned} v(\tau, x) &= \frac{1}{\sqrt{2\pi\tau}}\int_{\ln K}^{\infty} (e^y - K)e^{-r\tau} p_X(\tau, x, y)\, dy \\ &= \frac{1}{\sqrt{2\pi\tau}}\int_{\ln K}^{\infty} (e^y - K)e^{-r\tau} u_0(x, y)\cdot \\ &\qquad \exp\left[-\frac{d(x, y)^2}{2\tau}\right](1 + O(\tau; x, y))\, dy. \end{aligned} \tag{2.5}$$

From the inequality $|O(\tau; x, y)| \leq C\tau$ for some constant C, it is clear from (2.5) that remainder term will not contribute to the leading term of $v(\tau, x)$. For this reason we will ignore it completely in the subsequent calculations. Similarly, because $|e^{-r\tau} - 1| \leq r\tau$, we can replace $e^{-r\tau}$ in the integrand by 1. A quick inspection of (2.5) reveals that the leading term is determined by the values of the integrand near the point $y = \ln K$. Introducing the new variable $z = y - \ln K$, we conclude that $v(x, \tau)$ have the same leading term as the function

$$v^\sharp(\tau, x) = \frac{K}{\sqrt{2\pi\tau}} \int_0^\infty (e^z - 1) u_0(x, z + \ln K) \exp\left[-\frac{d(x, z + \ln K)^2}{2\tau}\right] dz. \tag{2.6}$$

The key calculation is contained in the proof of the following result.

LEMMA 2.2. *We have as* $\tau \downarrow 0$

$$\int_0^\infty z^k \exp\left[-\frac{d(x, z + \ln K)^2}{2\tau}\right] dz \sim k! \left[\frac{\sigma(K)\tau}{d(x, \ln K)}\right]^{k+1} \exp\left[-\frac{d(x, \ln K)^2}{2\tau}\right].$$

Proof. We follow the method in De Bruin (1999). Recall that

$$d(x, y) = \int_x^y \frac{du}{\eta(u)}.$$

Let

$$f(z) = d(x, z + \ln K)^2 - d(x, \ln K)^2.$$

The essential part of the exponential factor is $e^{-f(z)/2\tau}$. For any $\epsilon > 0$, there is $\lambda > 0$ such that $f(z) \geq \lambda$ for all $z \geq \epsilon$, hence

$$\frac{f(z)}{\tau} \geq \left(\frac{1}{\tau} - 1\right)\lambda + f(z).$$

From our basic assumptions (2.1) we see that there is a positive constant C such that $f(z) \geq Cz^2$ for sufficiently large z. Because the integral

$$\int_0^\infty z^k e^{-Cz^2}\, dz$$

is finite, the part of the original integral in the range $[\epsilon, \infty)$ does not contribute to the leading term of the integral. On the other hand, near $z = 0$ we have $f(z) \sim f'(0)z$ with $f'(0) = 2d(x, \ln K)/\sigma(K)$. It follows that the integral has the same leading term as

$$\exp\left[-\frac{d(x, \ln K)^2}{2\tau}\right] \int_0^\infty z^k \exp\left[-\frac{d(x, \ln K)}{\sigma(K)\tau} z\right] dz.$$

The last integral can be computed easily and we obtain the desired result. □

The main result of this section is the following.

THEOREM 2.3. *Suppose that the volatility function σ satisfies the basic assumption (2.1). If $x < \ln K$ we have as $\tau \downarrow 0$,*

$$v(\tau, x) \sim \frac{Ku_0(x, \ln K)}{\sqrt{2\pi}} \left[\frac{\sigma(K)}{d(x, \ln K)}\right]^2 \tau^{3/2}\, e^{-d(x,\ln K)^2/2\tau}.$$

Proof. We have shown that $v(x, \tau)$ and $v^\sharp(x, \tau)$ defined in (2.6) have the same leading term. We need to replace the function before the exponential factor by the first-order approximation of its value near the boundary point $z = 0$. First of all, we have

$$|e^z - 1 - z| \le z^2 e^z.$$

From the explicit expression (2.2) and the basic assumption (2.1) it is easy to verify that

$$|u_0(x, z + \ln K) - u_0(x, \ln K)| \le ze^{Cz}$$

for some positive constant C. By the estimate in Lemma 2.2 we obtain

$$\begin{aligned} v(\tau, x) &\sim v^\sharp(\tau, x) \\ &\sim \frac{Ku_0(x, \ln K)}{\sqrt{2\pi\tau}} \int_0^\infty z \exp\left[-\frac{d(x, z + \ln K)^2}{2\tau}\right] dz \\ &\sim \frac{Ku_0(x, \ln K)}{\sqrt{2\pi\tau}} \left[\frac{\sigma(K)\tau}{d(x, \ln K)}\right]^2 e^{-d(x,\ln K)^2/2\tau}. \end{aligned}$$

□

2.2. Implied Volatility Expansion

Using the leading term of the call price function calculated in the previous section, we are now in a position to prove the main theorem on the asymptotic behavior of the implied volatility $\hat{\sigma}(t, s)$ near expiry T. We will obtain this by comparing the leading terms of the relation

$$c(t, s) = C(t, s; \hat{\sigma}(t, s), r).$$

Here $C(t, s; \sigma, r)$ is the classical Black–Scholes pricing function. For this purpose, we need to calculate the leading term of the classical Black–Scholes call price function. Our main result is the following.

THEOREM 2.4. *Let $\hat{\sigma}(t, s)$ be the implied volatility when the stock price is s at time t. Then as $\tau \downarrow 0$,*

$$\hat{\sigma}(t, s) = \hat{\sigma}_0 + \hat{\sigma}_1 \tau + O(\tau^2),$$

where

$$\hat{\sigma}_0 = \left[\frac{1}{\ln K - \ln s} \int_s^K \frac{du}{u\sigma(u)}\right]^{-1}$$

and

$$\hat{\sigma}_1 = \frac{\hat{\sigma}_0^3}{(\ln K - \ln s)^2} \cdot \left[\ln \frac{\sqrt{\sigma(s)\sigma(K)}}{\hat{\sigma}(T,s)} + r \int_s^K \left(\frac{1}{\sigma^2(u)} - \frac{1}{\hat{\sigma}^2(T,s)} \right) \frac{du}{u} \right]. \tag{2.7}$$

As we have mentioned above, the leading term $\hat{\sigma}(T, s)$ was obtained in Berestycki, Busca, and Florent (2002). They first derived a quasi-linear partial differential equation for the implied volatility and used a comparison argument. When the interest rate $r = 0$, the first-order approximation of the implied volatility is given by

$$\hat{\sigma}_1(T,s) = \frac{\hat{\sigma}_0^3}{(\ln K - \ln s)^2} \ln \frac{\sqrt{\sigma(s)\sigma(K)}}{\hat{\sigma}(T,s)}.$$

This case has already appeared in Henry-Labordère (2008).

For the proof of Theorem 2.4 we first establish the following asymptotic result for the classical Black–Scholes call price function.

LEMMA 2.5. *Let $V(\tau, x; \sigma, r) = C(T - \tau, e^x; \sigma, r)$ be the classical Black–Scholes call price function. Then as $\tau \downarrow 0$,*

$$V(\tau, x; \sigma, r) \sim \frac{1}{\sqrt{2\pi}} \frac{K\sigma^3\tau^{3/2}}{(\ln K - x)^2} \exp\left[-\frac{\ln K - x}{2} + \frac{r(\ln K - x)}{\sigma^2} \right] \exp\left[-\frac{(\ln K - x)^2}{2\tau\sigma^2} \right] + R(\tau, x; \sigma, r).$$

The remainder satisfies

$$|R(\tau, x; \sigma, r)| \le C\tau^{5/2} \exp\left[-\frac{(\ln K - x)^2}{2\tau\sigma^2} \right],$$

where $C = C(x, \sigma, r, K)$ is uniformly bounded if all the indicated parameters vary in a bounded region.

Proof. The result can be proved by the classical Black–Scholes formula for $V(\tau, x; \sigma, r)$. We omit the detail. Note that the leading term can also be obtained directly from Theorem 2.3 by assuming that $\sigma(x)$ is independent of x. □

We are now in a position to complete the proof of the main Theorem 2.4. Set $\tilde{\sigma}(\tau, x) = \hat{\sigma}(T - \tau, s)$. From the relation

$$v(\tau, x) = V(\tau, x; \tilde{\sigma}(\tau, x), r) \tag{2.8}$$

and their expansions we see that the limit

$$\hat{\sigma}_0 = \lim_{\tau \downarrow 0} \tilde{\sigma}(\tau, x)$$

exists and is given by the given expression in the statement of the theorem. Indeed, by comparing only the exponential factors on the two sides we obtain

$$d(x, K) = \frac{\ln K - x}{\hat{\sigma}_0},$$

from which the desired expression for $\hat{\sigma}_0$ follows immediately.

Next, let

$$\tilde{\sigma}_1(\tau, x) = \frac{\tilde{\sigma}(\tau, x) - \tilde{\sigma}(0, x)}{\tau}.$$

We have obviously

$$\tilde{\sigma}(\tau, x) = \tilde{\sigma}(0, x) + \tilde{\sigma}_1(\tau, x)\tau. \tag{2.9}$$

From this a simple computation shows that

$$\exp\left[-\frac{(\ln K - x)^2}{2\tau\tilde{\sigma}(\tau, x)^2}\right] = \exp\left[-\frac{d^2}{2\tau} + \frac{\rho^2}{\sigma_0^3}\cdot\sigma_1(\tau, x)\right][1 + O(\tau)], \tag{2.10}$$

where for simplicity we have set

$$\rho = \ln K - x,\ d = d(x, K),\ \sigma_0 = \tilde{\sigma}(0, x)$$

on the right-hand side. From (2.8) and (2.9) we have

$$v(\tau, x) = V(\tau, x; \tilde{\sigma}(0, x) + \sigma_1(\tau, x)\tau, r).$$

We now use the asymptotic expansions for $v(\tau, x)$ and $V(\tau, x; \sigma, r)$ given by Theorem 2.3 and Lemma 2.5, respectively, and then apply (2.10) to the second exponential factor in the equivalent expression for $V(\tau, x; \tilde{\sigma}, r)$. After some simplification we obtain

$$u_0\,\sigma(K)^2(1 + O(\tau)) = \sigma_0 \exp\left[-\frac{\rho}{2} + \frac{r\rho}{\sigma_0^2}\right]\exp\left[-\frac{\rho^2}{\sigma_0^3}\cdot\tilde{\sigma}_1(\tau, x)\right],$$

where $u_0 = u_0(x, \ln K)$. Letting $\tau \downarrow 0$, we see that the limit

$$\hat{\sigma}_1 = \lim_{\tau\downarrow 0}\frac{\tilde{\sigma}(\tau, x) - \tilde{\sigma}(0, x)}{\tau}$$

exists and is given by

$$\hat{\sigma}_1 = \frac{r}{\rho} - \frac{\sigma_0^3}{2\rho^2} - \frac{\sigma_0^3}{\rho^2}\ln\frac{u_0\sigma(K)^2}{\sigma_0}.$$

Using the expression of $u_0 = u_0(x, \ln K)$ in (2.2) we immediately obtain the formula for $\hat{\sigma}_1$.

2.3. In-the-Money Case

For the in-the-money case $s > K$, from

$$(s - K)^+ = (s - K) + (s - K)^-,$$

we have

$$\begin{aligned} &u(\tau, x) - e^x \\ &= \mathbb{E}_x \left[e^{X_\tau} - K\right]^+ e^{-r\tau} - e^x \\ &= \mathbb{E}_x \left[e^{X_\tau} e^{-r\tau} - e^x\right] + e^{-r\tau}\mathbb{E}_x \left[e^{X_\tau} - K\right]^- Ke^{-r\tau}. \end{aligned}$$

The process $e^{X_\tau} e^{-r\tau} = S_\tau e^{-r\tau}$ is a martingale in τ starting from e^x, hence the first term immediately after the second equal sign vanishes. The calculation of the leading term of the second term is similar to that of the case when $s < K$. Therefore, for $s > K$ we have the following asymptotic expansion for $u(\tau, x)$

$$e^x - Ke^{-r\tau} - \frac{Ku_0(x, \ln K)}{\sqrt{2\pi}} \left[\frac{\sigma(K)}{d(x, \ln K)}\right]^2 \tau^{3/2} \exp\left[-\frac{d(x, \ln K)^2}{2\tau}\right].$$

It is easy to see that this case produces nothing new.

3. YOSHIDA'S APPROACH TO HEAT KERNEL EXPANSION

3.1. Time-Inhomogeneous Equations in One Dimension

In this section we review an expression of the heat kernel for a general nondegenerate linear parabolic differential equation due to Yoshida (1953). We will only work out the one-dimensional case but in a form that is more general than we actually need in view of possible future use. In our opinion, this form of heat kernel expansion is more efficient for applications at hand than the covariant form pioneered by Avramidi and adapted by Henry-Labordère. The latter, being intrinsic, is preferable for higher order corrections. However, the Yoshida approach is completely self-contained and, especially when the coefficients in the diffusions depend explicitly on time, introduces some clear simplifications of the necessary computations. Moreover, as mentioned in the introduction, Yoshida's heat kernel expansion is fully rigorous and leads to a uniformly convergent series on any compact subdomain of $\mathbb{R}$. Thus, in the nondegenerate case, this method will yield an asymptotic expansion of the implied volatility in the form

$$\sigma_{\mathrm{BS}}(t, T) \sim \sum_{i=0}^{n} \sigma_{\mathrm{BS},i} \tau^i, \tag{3.1}$$

as the time to expiration $\tau = T - t$ goes to zero. In this paper, we will limit ourselves to describing in detail only the zeroth, first, and second coefficient. It should be pointed out that Yoshida's approach does not easily extend to the degenerate (on the boundary) case.

Consider the following one-dimensional parabolic differential equation

$$u_t + \mathscr{L}u = u_t + \frac{1}{2}a(s, t)^2 u_{ss} + b(s, t)u_s + c(s, t)u = 0, \tag{3.2}$$

where subscripts refer to corresponding partial differentiations. In our case, $a(s, t) = s\sigma(s, t)$, where $\sigma(s, t)$ is the local volatility function. Note that $a(s, t)$ vanishes at $s = 0$ so it is not nondegenerate at this point. For the applicability of Yoshida's method in this case see Remark 3.5. We seek an expansion for the kth-order approximation to the fundamental solution $p(s, t, K, T)$ in the form

$$p(s, t, K, T) \sim \frac{e^{-d(K,s,t)^2/2\tau}}{\sqrt{2\pi\tau}a(K, T)} \sum_{i=0}^{k} u_i(s, K, t)\tau^i, \tag{3.3}$$

where

$$d(K, s, t) = \int_K^s \frac{d\eta}{a(\eta, t)} \tag{3.4}$$

is the Riemannian distance between the points K and s with respect to the time dependent Riemannian metric $ds^2/a(s, t)^2$, and $\tau = T - t$. Yoshida (1953) established that the coefficients u_i have the following form:

$$u_0(s, K, t) = \sqrt{\frac{a(s, t)}{a(K, t)}} \exp\left[-\int_K^s \frac{b(\eta, t)}{a(\eta, t)^2} d\eta - \int_K^s \frac{d_t(K, \eta, t)}{a(\eta, t)} d\eta\right]$$

and

$$u_i(s, K, t) = \frac{u_0(s, K, t)}{d(K, s, t)^i} \int_K^s \frac{d(K, \eta, t)^{i-1}}{u_0(\eta, K, t)} \left[\mathscr{L}u_{i-1} + \frac{\partial u_{i-1}}{\partial t}\right] \frac{d\eta}{a(\eta, t)}. \tag{3.5}$$

The function u_0 is given explicitly and u_i can be calculated recursively using mathematical software packages such as *Mathematica* or *Maple*.

The case $b = c = 0$ in the equation (3.2) is of special interest to us. There

$$u_0(s, K, t) = \sqrt{\frac{a(s, t)}{a(K, t)}} \exp\left[-\int_K^s \frac{d_t(K, \eta, t)}{a(\eta, t)} d\eta\right] \tag{3.6}$$

and

$$u_1(s, K, t) = \frac{u_0(s, K, t)}{d(K, s, t)} \int_K^s \frac{1}{u_0(\eta, K, t)} \left[\frac{a^2}{2} \frac{\partial^2 u_0}{\partial s^2} + \frac{\partial u_0}{\partial t}\right] \frac{d\eta}{a(\eta, t)}. \tag{3.7}$$

In particular, when in addition a is independent of time, we have

$$u_0(s, K) = \sqrt{\frac{a(s)}{a(K)}},$$

$$u_1(s, K) = \frac{1}{4d(K, s)} \sqrt{\frac{a(s)}{a(K)}} \left[a'(s) - a'(K) - \frac{1}{2} \int_K^s \frac{a'(\eta)^2}{a(\eta)} d\eta\right].$$

REMARK 3.1. In the Black–Scholes setting, $a(s, t) = \sigma_{\mathrm{BS}} s$ and u_0^{BS} and u_1^{BS} are given explicitly as

$$u_0^{\mathrm{BS}}(s, K) = \sqrt{\frac{s}{K}}, \qquad u_1^{\mathrm{BS}}(s, K) = -\frac{\sigma_{\mathrm{BS}}^2}{8} \sqrt{\frac{s}{K}}.$$

In fact, every u_k^{BS} can be calculated explicitly

$$u_k^{\mathrm{BS}}(s, K) = \frac{(-1)^k}{k!}\left(\frac{\sigma_{\mathrm{BS}}^2}{8}\right)^k \sqrt{\frac{s}{K}},$$

which in turn yields the following heat kernel expansion for Black–Scholes' transition probability density

$$p_{\mathrm{BS}}(s, K, t) = \frac{1}{\sqrt{2\pi t}\sigma_{\mathrm{BS}} K}\sqrt{\frac{s}{K}} \exp\left[-\frac{(\ln s - \ln K)^2}{2\sigma_{\mathrm{BS}}^2 t} - \frac{\sigma_{\mathrm{BS}}^2 t}{8}\right]. \tag{3.8}$$

This formula can also be verified directly.

3.2. Option Price Calculations

We follow the approach adopted by Henry-Labordère (2005) based on the earlier work of Dupire (2004) and Derman and Kani (1994), who used the following method to obtain the call prices directly from the probability density function without requiring a space integration. Unlike the method described in Section 2, the result can be obtained without using Laplace's method. Thus an additional approximation is avoided at this stage.

Suppose that $b = c = 0$ in 3.2. The Carr–Jarrow formula in Carr and Jarrow (1990) for the call prices $C(s, K, t, T)$ reads

$$C(s, K, t, T) = (s - K)^+ + \frac{1}{2}\int_t^T a(K, u)^2 p(s, t, K, u)\, du.$$

We use the Yoshida expansion (3.3) for the heat kernel $p(s, t, K, u)$, which gives

$$\begin{aligned} &C(s, K, t, T) - (s - K)^+ \\ &\sim \frac{1}{2\sqrt{2\pi}} \sum_{i=0}^{k} \left[\int_t^T a(K, u) e^{-d(K,s,t)^2/2(u-t)} (u - t)^{i-\frac{1}{2}} du\right] u_i(s, K, t). \end{aligned}$$

When we calculate the coefficient $\sigma_{\mathrm{BS},2}$ in the expansion (3.1) for the implied volatility σ_{BS}, we need an expansion for the call price function in the following form:

$$\begin{aligned} &[C(s, t, K, T) - (s - K)^+]e^{d(K,s,t)^2/2\tau} \\ &= C^{(1)}(s, K, t)\tau^{3/2} + C^{(2)}(s, K, t)\tau^{5/2} + o(\tau^{5/2}). \end{aligned} \tag{3.9}$$

The key step in making the approach rigorous is to show that the remainder in (3.9) truly is $o(\tau^{5/2})$. This in turn will follow from the fact that the first few terms in the "geometric series expansion" can be complemented by the Levi parametrix method to ensure that we have

$$p(s, t, K, T) = \frac{e^{-d(K,s,t)^2/2\tau}}{\sqrt{2\pi} a(K, T)}\left[\sum_{i=0}^{2} u_i(s, K, t)\tau^{i-\frac{1}{2}} + o\left(\tau^{\frac{3}{2}}\right)\right],$$

i.e., that the suitably modified preliminary approximation of the heat kernel can actually give a convergent series with a tail of order smaller than the last term in the geometric series. Note that the theory requires us to proceed until order $k = n - 1$ (or $k = n$ in the at-the-money case, see below) in the series in order to calculate $\sigma_{\mathrm{BS},n}$ in the expansion for the implied volatility. If we wish to use the series to calculate a first order Greek, like the Delta, we would need to expand up to order $k = 2$ before using the Levi parametrix. See Section 3.4 for some additional details.

We now proceed with some additional approximations that will be necessary to obtain the expansion for the implied volatility. We will set $\tau = T - t$ as before and use $d = d(K, s, t)$ for simplicity. We have

$$\begin{aligned}
&\int_t^T a(K,u)e^{-d^2/2(u-t)}(u-t)^{i-\frac{1}{2}}\,du \\
&\quad\sim \int_t^T [a(K,t) + a_t(K,t)(u-t) + \frac{1}{2}a_{tt}(K,t)(u-t)^2]e^{-d^2/2(u-t)}(u-t)^{i-\frac{1}{2}}\,du \\
&\quad= a(K,t)\int_t^T e^{-d^2/2(u-t)}(u-t)^{i-\frac{1}{2}}\,du \\
&\qquad+ a_t(K,t)\int_t^T e^{-d^2/2(u-t)}(u-t)^{i+\frac{1}{2}}\,du \\
&\qquad+ \frac{1}{2}a_{tt}(K,t)\int_t^T e^{-d^2/2(u-t)}(u-t)^{i+\frac{3}{2}}\,du \\
&\quad= a(K,t)U_i(d,\tau) + a_t(K,t)U_{i+1}(d,\tau) + \frac{1}{2}a_{tt}(K,t)U_{i+2}(d,\tau),
\end{aligned}$$

where

$$U_i(\omega,\tau) = \int_0^\tau u^{i-\frac{1}{2}}e^{-\omega^2/2u}\,du. \tag{3.10}$$

Expanding $a(K, u)$ to the second order is sufficient for computing the implied volatility expansion up to the second order of τ. Inserting this into (3.9) we get

$$\begin{aligned}
&C(s,K,t,T) - (s-K)^+ \\
&\quad\sim \frac{1}{2\sqrt{2\pi}}\sum_{i=0}^{k}\left[a(K,t)U_i(d,\tau) + a_t(K,t)U_{i+1}(d,\tau) + \frac{1}{2}a_{tt}(K,t)U_{i+2}\right]u_i(s,K,t).
\end{aligned} \tag{3.11}$$

We need to take $k = 1$ for the case $s \neq K$ and $k = 2$ for the case $s = K$ (at the money).

REMARK 3.2. In the Black–Scholes setting the above asymptotic relation reads

$$C_{\mathrm{BS}}(s,K,t,T) - (s-K)^+ \sim \frac{\sqrt{sK}}{2\sqrt{2\pi}}\left[\sigma_{\mathrm{BS}}U_0(d_{\mathrm{BS}},\tau) - \frac{\sigma_{\mathrm{BS}}^3}{8}U_1(d_{\mathrm{BS}},\tau)\right],$$

where

$$d_{\mathrm{BS}} = d_{\mathrm{BS}}(K,s) = \frac{1}{\sigma_{\mathrm{BS}}}\ln\frac{s}{K}$$

is the distance between K and s in the Black–Scholes' setting. In fact, the complete series can be obtained by using the general formula (3.8) in Remark 3.1

$$C_{\mathrm{BS}}(s,K,t,T)-(s-K)^{+}=\frac{\sqrt{sK}\sigma_{\mathrm{BS}}}{2\sqrt{2\pi}}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\left(\frac{\sigma_{\mathrm{BS}}^2}{8}\right)^k U_k(d_{\mathrm{BS}},\tau).$$

The leading term of the call price away from the money is $\tau^{3/2}e^{-d^2/2\tau}$. Besides the leading term, we also need the next term, which has the order $\tau^{5/2}e^{-d^2/2\tau}$. After canceling some common factors and dropping higher order terms we have arrived at the following balance relation between the call prices from the local volatility model and from the Black–Scholes setting

$$\begin{aligned}&\sqrt{sK}\left[\sigma_{\mathrm{BS}}U_0(d_{\mathrm{BS}},\tau)-\frac{\sigma_{\mathrm{BS}}^3}{8}U_1(d_{\mathrm{BS}},\tau)\right]\\ &\quad\sim\left[a(K,t)U_0(d,\tau)+a_t(K,t)U_1(d,\tau)+\frac{1}{2}a_{tt}(K,t)U_2(d,\tau)\right]u_0(s,K,t)\\ &\qquad+[a(K,t)U_1(d,\tau)+a_t(K,t)U_2(d,\tau)]\,u_1(s,K,t).\end{aligned}\tag{3.12}$$

We now consider two regimes separately.

REGIME 1: $s\neq K$ fixed (away from the money). We shall use the following asymptotic formulas for U_0 and U_1 that are easily obtained from the well-known asymptotic formula for the complementary error function. As $\tau\downarrow 0$, we have

$$\begin{aligned}U_0(\omega,\tau)&=2\sqrt{\tau}e^{-\omega^2/2\tau}-2\omega\int_{\frac{\omega}{\sqrt{\tau}}}^{\infty}e^{-x^2/2}dx\\ &=2\left[\frac{\tau^{3/2}}{\omega^2}-\frac{3\tau^{5/2}}{\omega^4}+O\left(\tau^{7/2}\right)\right]e^{-\omega^2/2\tau},\\ U_1(\omega,\tau)&=\frac{2\tau^{3/2}}{3}e^{-\omega^2/2\tau}-\frac{\omega^2}{3}U_0(\omega,\tau)\\ &=\left[\frac{2\tau^{5/2}}{\omega^2}+O\left(\tau^{7/2}\right)\right]e^{-\omega^2/2\tau}.\end{aligned}$$

These expansions will be applied to $\omega=d(K,s,t)$ and $\omega=d_{\mathrm{BS}}(K,s)$, respectively. We now let

$$\xi=\ln\frac{s}{K}.$$

The relation between d_{BS} and σ_{BS} is

$$d_{\mathrm{BS}}=\frac{\xi}{\sigma_{\mathrm{BS}}}.\tag{3.13}$$

In the time-inhomogeneous case we seek an expansion for σ_{BS} in the form (3.1). Note the dependence of the coefficients in the expansion on the *spot variable t*. This dependence is absent in the time-homogeneous case, as will be clear from the result of the expansion. Its presence in the case of time-inhomogeneous diffusions is natural because already the transition probability density depends jointly on t and T and not only on their difference. We seek natural expressions for the coefficients which do not depend on the expiry explicitly. Mathematically it is of course also possible to expand around $t = T$, but in this case more terms are needed to recover the same accuracy.

We now use the above expansions for $U_0(\omega, \tau)$ and $U_1(\omega, \tau)$ and (3.13) in (3.12). After removing the factor $\tau^{3/2}/2\sqrt{2\pi}$, we see that the left-hand side of (3.12) becomes

$$\begin{aligned}&\frac{\sqrt{sK}}{\xi^2}\exp\left[-\frac{\xi^2}{2\sigma_{\mathrm{BS},0}^2\tau}+\frac{\xi^2\sigma_{\mathrm{BS},1}}{\sigma_{\mathrm{BS},0}^3}\right]\\ &\times\left\{\sigma_{\mathrm{BS},0}^3+\left[3\sigma_{\mathrm{BS},0}^2\sigma_{\mathrm{BS},1}-\frac{\xi^2\sigma_{\mathrm{BS},0}}{2}\left(3\left(\frac{\sigma_{\mathrm{BS},1}}{\sigma_{\mathrm{BS},0}}\right)^2-\frac{2\sigma_{\mathrm{BS},2}}{\sigma_{\mathrm{BS},0}}\right)\right.\right.\\ &\left.\left.-\sigma_{\mathrm{BS},0}^5\left(\frac{3}{\xi^2}+\frac{1}{8}\right)\right]\tau\right\}.\end{aligned} \tag{3.14}$$

Here we have used the following expansion for expanding the exponent in the exponential term:

$$\frac{1}{\sigma_{\mathrm{BS}}^2}\sim\frac{1}{\sigma_{\mathrm{BS},0}^2}\left[1-\frac{2\sigma_{\mathrm{BS},1}}{\sigma_{\mathrm{BS},0}}\tau+\left(\frac{3\sigma_{\mathrm{BS},1}^2}{\sigma_{\mathrm{BS},0}^2}-\frac{2\sigma_{\mathrm{BS},2}}{\sigma_{\mathrm{BS},0}}\right)\tau^2\right].$$

On the local volatility side, in the present regime (away from the money) the terms involving a_{tt} in (3.12) can be ignored because their orders in τ is at least 7/2. After again removing the factor $\tau^{3/2}/2\sqrt{2\pi}$, we see that the terms of up to order τ are

$$\exp\left[-\frac{d^2}{2\tau}\right]\left[\frac{a(K,t)u_0}{d^2}+\left(\frac{a_t(K,t)u_0}{d^2}-\frac{3a(K,t)u_0}{d^4}+\frac{a(K,t)u_1}{d^2}\right)\tau\right]. \tag{3.15}$$

We are ready to identify the corresponding terms in (3.14) and (3.15).

- By matching the exponents, we obtain

$$\sigma_{\mathrm{BS},0}=\frac{\xi}{d(K,s,t)}=\left[\frac{1}{\ln s-\ln K}\int_K^s\frac{d\eta}{a(\eta,t)}\right]^{-1}. \tag{3.16}$$

- By matching the constant terms, we obtain

$$\begin{aligned}\sigma_{\mathrm{BS},1}&=\frac{\sigma_{\mathrm{BS},0}^3}{\xi^2}\ln\left[\frac{u_0(s,K,t)a(K,t)\xi^2}{\sqrt{sK}d^2\sigma_{\mathrm{BS},0}^3}\right]\\ &=\frac{\xi}{d(K,s,t)^3}\ln\left[\frac{u_0(s,K,t)a(K,t)d(K,s,t)}{\xi\sqrt{sK}}\right],\end{aligned} \tag{3.17}$$

- By matching the first-order term, we obtain

$$\begin{aligned}\sigma_{\mathrm{BS},2}&=-\frac{3\sigma_{\mathrm{BS},1}\sigma_{\mathrm{BS},0}^2}{\xi^2}+\frac{3\sigma_{\mathrm{BS},1}^2}{2\sigma_{\mathrm{BS},0}}+\frac{\sigma_{\mathrm{BS},0}^3}{\xi^2}\\ &\quad\times\left[\frac{3\sigma_{\mathrm{BS},0}^2}{\xi^2}+\frac{\sigma_{\mathrm{BS},0}^2}{8}+\frac{a_t(K,t)}{a(K,t)}-\frac{3}{d^2(K,s,t)}+\frac{u_1(s,K,t)}{u_0(s,K,t)}\right]\\ &=-\frac{3\sigma_{\mathrm{BS},1}\sigma_{\mathrm{BS},0}}{\xi^2}+\frac{3\sigma_{\mathrm{BS},1}^2}{2\sigma_{\mathrm{BS},0}}+\frac{\sigma_{\mathrm{BS},0}^5}{8\xi^2}+\frac{\sigma_{\mathrm{BS},0}^3}{\xi^2}\left[\frac{a_t(K,t)}{a(K,t)}+\frac{u_1(s,K,t)}{u_0(s,K,t)}\right].\end{aligned} \tag{3.18}$$

In the above formulas we need the expressions for u_0 and u_1 obtained earlier in (3.6) and (3.7).

REGIME 2: $s = K > 0$ (at the money). We use again the expansion (3.11) for the call price. After setting $s = K$ there we find that up to the order $\tau^{5/2}$ the call price $C(K, K, t, T)$ is equivalent to

$$\frac{1}{2\sqrt{2\pi}}\sum_{i=0}^{2}\left[aU_i(0,\tau)+a_tU_{i+1}(0,\tau)+\frac{1}{2}a_{tt}U_{i+2}(0,\tau)\right]u_i(K,K,t). \tag{3.19}$$

With $\omega = 0$ in (3.10) we have trivially

$$U_0(0,\tau)=2\sqrt{\tau},\quad U_1(0,\tau)=\frac{2}{3}\tau^{3/2},\quad U_2(0,\tau)=\frac{2}{5}\tau^{5/2}.$$

Therefore in the expansion (3.19) only for the case $i = 0$ the term involving the second derivative a_{tt} needs to be kept. On the Black–Scholes side we then have, after dropping the factor $1/2\sqrt{2\pi}$,

$$K\Big[2\left(\sigma_{\mathrm{BS},0}+\tau\sigma_{\mathrm{BS},1}+\tau^2\sigma_{\mathrm{BS},2}\right)\sqrt{\tau}-\frac{1}{12}\left(\sigma_{\mathrm{BS},0}+\sigma_{\mathrm{BS},1}\tau+\sigma_{\mathrm{BS},2}\tau^2\right)^3\tau^{3/2}$$
$$+\frac{1}{320}\left(\sigma_{\mathrm{BS},0}+\sigma_{\mathrm{BS},1}\tau+\sigma_{\mathrm{BS},2}\tau^2\right)^5\tau^{\frac{5}{2}}+O(\tau^{7/2})\Big]$$

Grouping the terms according to powers of τ, we obtain

$$2K\sigma_{\mathrm{BS},0}\sqrt{\tau}+K\left(\sigma_{\mathrm{BS},1}-\frac{1}{12}\sigma_{\mathrm{BS},0}^3\right)\tau^{3/2}$$
$$+\left[2K\sigma_{\mathrm{BS},2}-\frac{K}{4}\sigma_{\mathrm{BS},0}^2\sigma_{\mathrm{BS},1}+K\frac{\sigma_{\mathrm{BS},0}^5}{320}\right]\tau^{5/2}+O(\tau^{7/2}).$$

On the local volatility side we have

$$2au_0\sqrt{\tau}+\frac{2}{3}(a_tu_0+au_1)\tau^{3/2}+\frac{2}{5}(a_{tt}u_0+a_tu_1+au_2)\tau^{5/2}+O(\tau^{7/2}),$$

where we have omitted the dependence on the independent variables, in all of which we replace s by K. Matching the coefficients of the powers of τ and using $u_0(K, K, t) = 1$, we obtain

$$K\sigma_{\mathrm{BS},0}=a,$$
$$K\left(2\sigma_{\mathrm{BS},1}-\frac{1}{12}\sigma_{\mathrm{BS},0}^3\right)=\frac{2}{3}(a_t+au_1),$$
$$2K\sigma_{\mathrm{BS},2}-\frac{K}{4}\sigma_{\mathrm{BS},0}^2\sigma_{\mathrm{BS},1}+\frac{K}{320}\sigma_{\mathrm{BS},0}^5=\frac{2}{5}(a_{tt}u_0+a_tu_1+au_2).$$

They yield consecutively,

$$
\begin{aligned}
\sigma_{\mathrm{BS},0} &= \frac{a(K,t)}{K}, \\
\sigma_{\mathrm{BS},1} &= \frac{a_t + a u_1}{3K} + \frac{a(K,t)^3}{24K^3}, \\
\sigma_{\mathrm{BS},2} &= \frac{a_{tt} u_0 + a_t u_1 + a u_2}{5K} + \frac{\sigma_{\mathrm{BS},0}^2 \sigma_{\mathrm{BS},1}}{8} - \frac{\sigma_{\mathrm{BS},0}^5}{640}.
\end{aligned}
\tag{3.20}
$$

REMARK 3.3. We have checked by explicit computation that in the time-homogeneous case, these at-the-money expressions for the $\sigma_{\mathrm{BS},i}$, $i = 0, 1, 2$, coincide with limits as $s \to K$ of the out-of-the-money expressions (3.16), (3.17), and (3.18).

REMARK 3.4. One possible application of our approximate formula (3.20) for Black–Scholes implied volatility is to generate an implied volatility smile when good prices for out-of-the-money options are not readily available, as is often the case with less liquid underlyings. Supposing that the prices (and therefore implied volatilities) of at-the-money options are available, we may use (3.20) to build an implied volatility surface that matches the empirically observed implied volatility surface at the money (ATM).

Specifically, suppose we posit a functional form for the local volatility function $a(\eta, t)$, motivated for example either by a stochastic volatility model or from a fit to another (better resolved) volatility surface. Denote observed ATM implied volatilities by $\sigma_{\mathrm{BS}}^{\mathrm{emp}}(K, T)$ and as before, the approximate implied volatilities obtained from (3.20) by $\sigma_{\mathrm{BS}}(K, T)$. Then we may define a time-change $T \mapsto \theta(T)$ through the relation

$$\sigma_{\mathrm{BS}}(K, \theta(T))^2\, \theta(T) = \sigma_{\mathrm{BS}}^{\mathrm{emp}}(K, T)^2\, T.$$

The generated volatility surface given by

$$\tilde{\sigma}_{\mathrm{BS}}(x, T) = \sigma_{\mathrm{BS}}(x, \theta(T))$$

would then by construction match perfectly ATM with a smile consistent with the posited functional form for the local volatility function $a(\eta, t)$. Moreover, with the above specification of the time change, if $\sigma_{\mathrm{BS}}(\cdot)$ is arbitrage free, $\tilde{\sigma}_{\mathrm{BS}}(\cdot)$ will be as well.

3.3. Case of Nonzero Interest Rate

It is straightforward to combine the Yoshida approach with nonzero interest rates and dividends to account for the presence of the r dependent term in (2.7). Note that if the stock satisfies the time-homogeneous equation

$$dS_t = S_t \left\{ r\, dt + \sigma(S_t)\, dW_t \right\},$$

call it Problem (I), then the forward price $f_t = e^{r\tau} S_t$ satisfies the driftless but time-inhomogeneous equation

$$df_t = f_t \sigma\left(e^{-r\tau} f_t\right) dW_t = f_t \tilde{\sigma}(f_t, t)\, dW_t,$$

call it Problem (II). The relationship between the implied volatilities for these two problems is easily seen to be

$$\sigma^r_{\mathrm{BS}}(s, t, K, T) = \sigma^f_{\mathrm{BS}}\left(s, t, Ke^{-r(\tau)}, T\right).$$

From this it follows that

$$\sigma^r_{\mathrm{BS},0}(s, K) = \sigma^f_{\mathrm{BS},0}(s, K) = \left[\frac{1}{\ln s - \ln K}\int_K^s \frac{du}{u\sigma(u)}\right]^{-1}, \tag{3.21}$$

$$\begin{aligned}\sigma^r_{\mathrm{BS},1}(s, K) &= \sigma^f_{\mathrm{BS},1}(s, K) + \frac{\partial}{\partial t}\left[\sigma^f_{\mathrm{BS},0}(s, Ke^{r(\tau)})\right]\Big|_{t=T} \\ &= \sigma^f_{\mathrm{BS},1}(s, K) + r\left[\int_K^s \frac{du}{u\sigma(u)}\right]^{-1} - \frac{r\log(s/K)}{\sigma(K)}\left[\int_K^s \frac{du}{u\sigma(u)}\right]^{-2},\end{aligned}$$

where in the first term above we need to determine $\sigma^f_{\mathrm{BS},1}(y, x)$ for what is now a time-inhomogeneous problem (II). Because the volatility $\tilde{\sigma}$ now depends explicitly on time, in contrast (3.17), now there is an additional term linear in r with the coefficient

$$-\ln\frac{s}{K}\left[\int_K^s \frac{du}{u\sigma(u)^2}\right]\left[\int_K^s \frac{du}{u\sigma(u)}\right]^{-3} + \frac{1}{\sigma(K)}\ln\frac{s}{K}\left[\int_K^s \frac{du}{u\sigma(u)}\right]^{-2}.$$

The second term in the above expression cancels the third term in (3.21) to produce exactly the expression (2.7).

The procedure we have outlined above that allows us to pass from the zero to the nonzero interest case in the case of the first-order coefficient can also be repeated to determine the second-order correction.

3.4. Yoshida's Construction of the Heat Kernel

As mentioned earlier, if the diffusion coefficient is nondegenerate and the domain is compact, Yoshida (1953) has shown how to use the so-called geometric series in (3.3) to obtain the exact fundamental solution to the backward Kolmogorov equation. In this section, we give an outline of Yoshida's method.

Consider the parabolic equation

$$u_t + \mathscr{L}u = 0,$$

where

$$\mathscr{L}u = \frac{1}{2}a^2 u_{yy} + bu_y + cu.$$

Denote the geometric approximation of $k+1$ terms in (3.3) by p_k. Choose a function of one variable $\eta(x) : [0, \infty) \mapsto [0, \infty)$, which equals 1 on $[0, \epsilon]$, and equals zero for $x > 2\epsilon$. Define the truncated geometric approximation

$$\hat{p}_k(y, s, x, t) = \eta(d(x, y, s))p_k(y, s, x, T),$$

where $d(x, y, s)$ is the distance function (3.4). Define the space–time convolution operator

$$F \otimes G(y, s, x, t) = \int_0^\tau d\tau \int_{\mathbb{R}} F(y, s, z, \tau) G(z, \tau, x, t)\, dz.$$

Starting from

$$K^{(k)} = -\frac{\partial \hat{p}_k}{\partial s} - \mathscr{L}\hat{p}_k,$$

we have a series of convolutions

$$K_n^{(k)} = K^{(k)} \otimes K_{n-1}^{(k)}.$$

Let

$$Q(y, s, x, t) = \sum_{n=1}^{\infty} (-1)^{n+1} K_n^{(k)}(y, s, x, t).$$

Then the fundamental solution is given by

$$P(y, s, x, t) = \hat{p}_k(y, s, x, t) - (\hat{p}_k \otimes Q)(y, s, x, t).$$

Yoshida's proof of convergence of the series representing P assumes the underlying domain is compact (actually a mild generalization thereof). The key step is to show by induction that

$$\left| K_n^{(k)}(y, s, x, t) \right| \leq D C^n (t - s)^{k+n-\frac{3}{2}},$$

where C and D are constants. The proof can easily be extended to unbounded domains if proper conditions are imposed on the distance function $d(x, y, s)$.

REMARK 3.5. Yoshida's approach requires the equation to be nondegenerate

$$\min_{t \in [0, T],\, S \in \mathbb{R}} a(S, t) = c > 0.$$

Models like the CEV model, which will be considered in the numerical section, and even the Black–Scholes model itself, do not satisfy this condition because they are degenerate when $S = 0$. There may also be problems at $S = \infty$. We have encountered similar problems in the probabilistic approach in Section 2. However, as we have explained in Section 2, as long as we keep away from these two boundary points, the behavior of the coefficient functions in a neighborhood of the boundary points are irrelevant as long as we also impose some moderate conditions on the growth of $a(S, t)$. In particular the call price expansion will not be affected. This principle of not feeling the boundary is explained in detail in Appendix A.

4. NUMERICAL RESULTS

We have derived an expansion formula for implied volatility up to second order in time-to-expiration in the form

$$\sigma_{\mathrm{BS}}(t, T) = \sigma_{\mathrm{BS},0}(t) + \sigma_{\mathrm{BS},1}(t)\tau + \sigma_{\mathrm{BS},2}(t)\tau^2 + O\left(\tau^3\right),$$

where the coefficients $\sigma_{\mathrm{BS},i}(t)$ are given by (3.16), (3.17), and (3.18).

To test this expansion formula numerically, we use well-known exact formulas for option prices in two specific time-homogeneous local volatility models: the CEV model and the quadratic local volatility model, as developed by Zuhlsdorff (2001), Lipton (2002), Andersen (2008), and others.

Time dependence is modeled as a simple time-change so that these exact time-independent solutions may be reused. Specifically, the time change is

$$\tau(T) = \int_0^T e^{-2\lambda t}\, dt = \frac{1}{2\lambda}\left(1 - e^{-2\lambda T}\right).$$

Throughout, for simplicity, we assume zero interest rates and dividends so that $b(s, t) = 0$ in equation (3.2).

4.1. Henry-Labordère's Approximation

Henry-Labordère (2008) presents a heat kernel expansion-based approximation to implied volatility in equation (5.40) on p. 140 of his book

$$\sigma_{BS}(K, T) \approx \sigma_0(K, t)\left\{1 + \frac{T}{3}\left[\frac{1}{8}\sigma_0(K, t)^2 + \mathcal{Q}(f_{av}) + \frac{3}{4}\mathcal{G}(f_{av})\right]\right\} \tag{4.1}$$

with

$$\mathcal{Q}(f) = \frac{C(f)^2}{4}\left[\frac{C''(f)}{C(f)} - \frac{1}{2}\left(\frac{C'(f)}{C(f)}\right)^2\right]$$

and

$$\mathcal{G}(f) = 2\,\partial_t \log C(f) = 2\,\frac{\partial_t a(f, t)}{a(f, t)},$$

where $C(f) = a(f, t)$ in our notation, $f_{av} = (s + K)/2$ and the term $\sigma_0(K, t)$ is our lowest order coefficient (3.16) originally derived in Berestycki et al. (2002)

$$\sigma_0(K, t) = \left[\frac{1}{\ln s - \ln K}\int_K^s \frac{d\eta}{a(\eta, t)}\right]^{-1}.$$

On p. 145 of his book, Henry-Labordère (2008) presents an alternative approximation to first order in τ, matching ours exactly in the time-homogeneous case and differing only slightly in the time-inhomogeneous case. In Section 2, we have shown that our approximation is the optimal one to first order τ.

4.2. Model Definitions and Parameters

4.2.1. CEV Model. The SDE is

$$df_t = e^{-\lambda t}\,\sigma\,\sqrt{f_t}\,dW_t$$

TABLE 4.1
CEV Model Implied Volatility Errors for Various Strike Prices in the Henry-Labordère (HL) Approximation and Our First- and Second-Order Approximations, Respectively. The Exact Volatility in the Last Column is Obtained by Inverting the Closed-Form Expression for the Option Price in the CEV Model

Strikes	$\Delta\sigma_{\mathrm{HL}}$	$\Delta\sigma_1$	$\Delta\sigma_2$	σ_{exact}
0.50	2.12e-05	1.31e-06	1.98e-08	0.2368
0.75	3.46e-06	7.98e-07	9.87e-09	0.2148
1.00	5.68e-07	5.68e-07	6.03e-09	0.2001
1.25	1.52e-06	4.21e-07	4.08e-09	0.1891
1.50	3.45e-06	3.33e-07	2.96e-09	0.1805
1.75	5.45e-06	2.73e-07	2.18e-09	0.1734
2.00	7.27e-06	2.29e-07	1.70e-09	0.1674

with $\sigma = 0.2$. In the time-independent version, $\lambda = 0$ and in the time-dependent version, $\lambda = 1$. For the CEV model therefore,

$$\mathcal{Q}(f) = -\frac{3}{32}\frac{\sigma^2}{f} \quad \text{and} \quad \mathcal{G}(f) = -2\,\lambda$$

so the Henry-Labordère approximation (4.1) becomes

$$\sigma_{BS}(K, T) \approx \sigma_0(K, t)\left\{1 + \frac{T}{3}\left[\frac{1}{8}\sigma_0(K, t)^2 - \frac{3}{32}\frac{\sigma^2}{f_{av}} - \frac{3}{2}\lambda\right]\right\}.$$

The closed-form solution for the square-root CEV model is well known and can be found, for example, in Shaw (1998).

4.2.2. Quadratic Model. The SDE is

$$df_t = e^{-\lambda t}\,\sigma\,\left\{\psi\, f_t + (1-\psi) + \frac{\gamma}{2}\,(f_t - 1)^2\right\}\,dW_t$$

with $\sigma = 0.2$, $\psi = -0.5$, and $\gamma = 0.1$. Again in the time-independent version, $\lambda = 0$ and in the time-dependent version, $\lambda = 1$. Then for the quadratic model,

$$\mathcal{Q}(f) = \frac{1}{32}\sigma^2\,\{(f-1)^3\,(3\,f+1)\,\gamma^2 + 24\,(1-\psi)\,\gamma\,f \\ + 12\,\psi\,\gamma\,f^2 - 4\,[(4-3\,\psi)\,\gamma + \psi^2]\}$$

and again

$$\mathcal{G}(f) = -2\,\lambda.$$

The closed-form solution for the quadratic model with these parameters[1] is given in Andersen (2008).

[1] The solution is more complicated for certain other parameter choices.

TABLE 4.2
Quadratic Model Implied Volatility Errors for Various Strike Prices in the Henry-Labordère (HL) Approximation and our First- and Second-Order Approximations, Respectively. The Exact Volatility in the Last Column is Obtained by Inverting the Closed-Form Expression for the Option Price in the Quadratic Model

Strikes	$\Delta\sigma_{\text{HL}}$	$\Delta\sigma_1$	$\Delta\sigma_2$	σ_{exact}
0.50	−8.83e-05	−1.04e-05	−1.08e-07	0.3129
0.75	−3.42e-05	−3.05e-06	−1.94e-08	0.2451
1.00	−2.14e-06	−1.09e-06	−4.58e-09	0.2003
1.25	1.99e-05	−4.31e-07	−1.30e-09	0.1675
1.50	3.32e-05	−1.80e-07	−3.92e-10	0.1418
1.75	4.13e-05	−7.59e-08	−5.28e-11	0.1209
2.00	4.56e-05	−3.16e-08	9.57e-12	0.1032

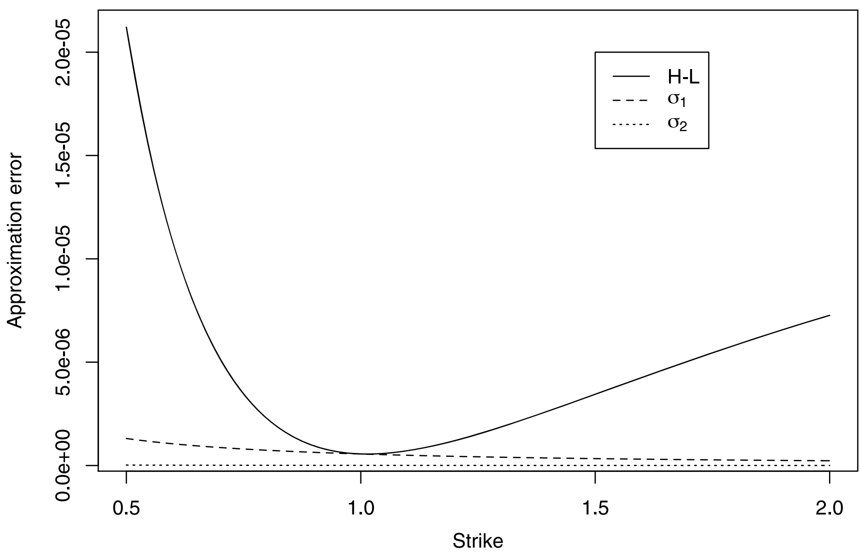

FIGURE 4.1. Approximation errors in implied volatility terms as a function of strike price for the square-root CEV model with the time-homogeneous ($\lambda = 0$) version of the local volatility function of Section 4.2.1. The solid line corresponds to the error in Henry-Labordère's approximation (5.40), and the dashed and dotted lines to our first- and second-order approximations, respectively. Note that the error in our second-order approximation is zero on this scale.

4.3. Results

In Tables 4.1 and 4.2, respectively, we present the errors in the above approximations in the case of time-independent CEV and quadratic local volatility functions (with $\lambda = 0$ and $T = 1$). We note that our approximation does slightly better than Henry-Labordère's, although the errors in both approximations are negligible. In Figures 4.1 and 4.2, respectively, these errors are plotted.

In Figures 4.3 and 4.4, we plot results for the time-dependent cases $\lambda = 1$ with $T = 0.25$ and $T = 1.0$, respectively, comparing our approximation to implied volatility with

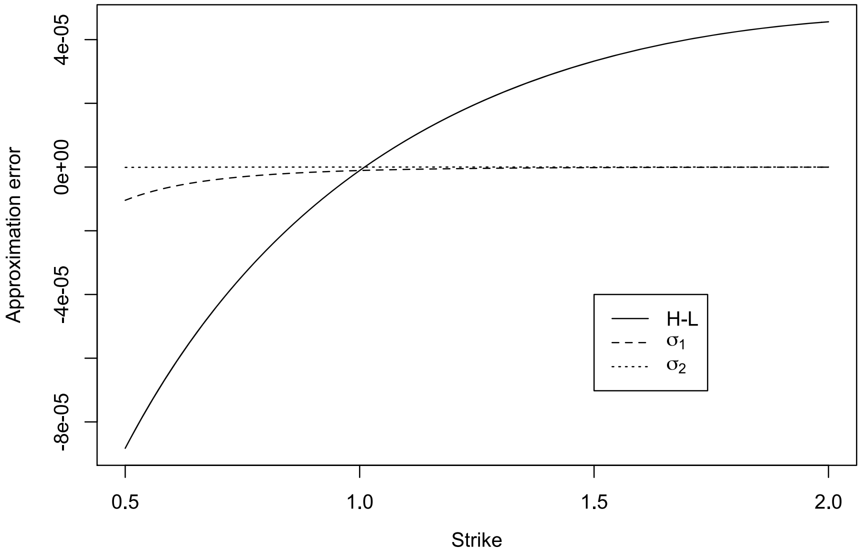

FIGURE 4.2. Approximation errors in implied volatility terms as a function of strike price for the quadratic model with the time-homogeneous ($\lambda = 0$) version of the local volatility function of Section 4.2.2. The solid line corresponds to the error in Henry-Labordère's approximation (5.40), and the dashed and dotted lines to our first- and second-order approximations, respectively. Note that the error in our second-order approximation is zero on this scale.

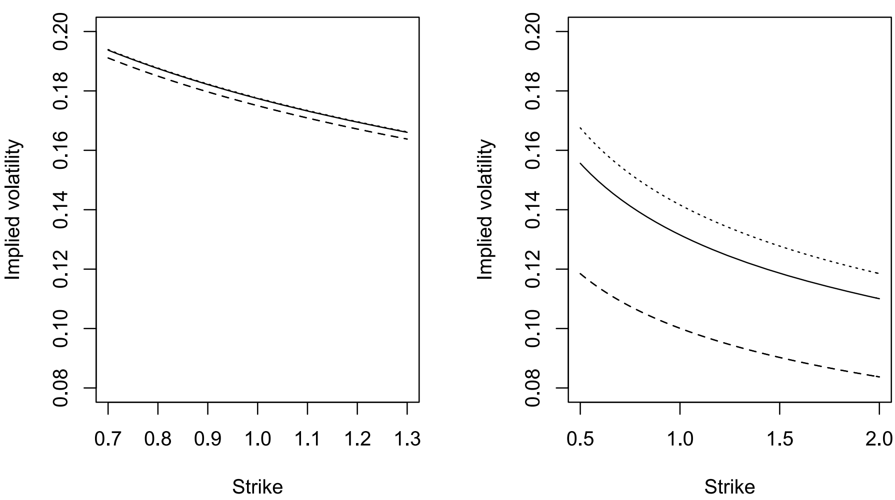

FIGURE 4.3. Implied volatility approximations in the CEV model with the strongly time-inhomogeneous ($\lambda = 1$) version of the local volatility function of Section 4.2.1 for two expirations: $\tau = 0.25$ on the left and $\tau = 1.0$ on the right. The solid line is exact implied volatility, the dashed line is our approximation to first order in $\tau = \tau$ (with only σ_1 and not σ_2), the dotted line is our approximation to second order in τ (including σ_2 and almost invisible in the left panel). On this scale, Henry-Labordère's first-order approximation is indistinguishable from our first-order approximation (the dashed line).

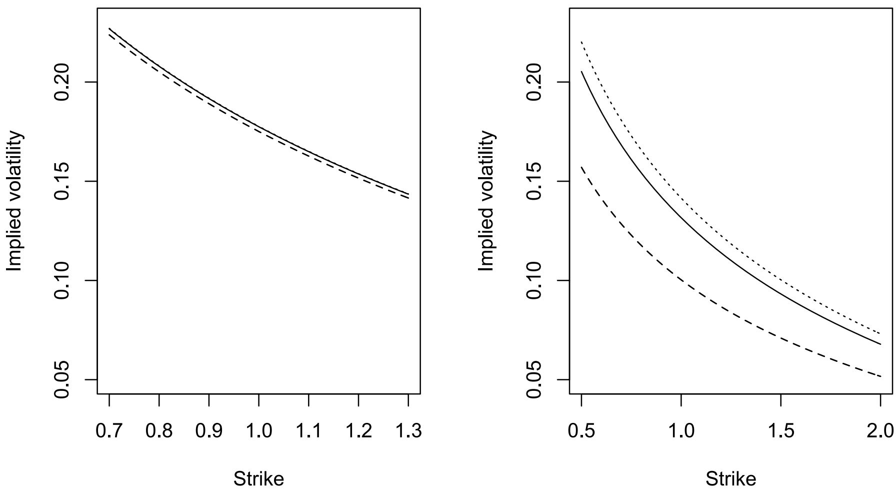

FIGURE 4.4. Implied volatility approximations in the quadratic model with the strongly time-inhomogeneous ($\lambda = 1$) version of the local volatility function of Section 4.2.2 for two expirations: $\tau = 0.25$ on the left and $\tau = 1.0$ on the right. The solid line is exact implied volatility, the dashed line is our approximation to first order in $\tau = \tau$ (with only σ_1 and not σ_2), the dotted line is our approximation to second order in τ (including σ_2 and almost invisible in the left panel). On this scale, Henry-Labordère's first-order approximation is indistinguishable from our first-order approximation (the dashed line).

the exact result. To first order in $\tau = \tau$ (with only σ_1 and not σ_2), we see that our approximation is reasonably good for short expirations ($\lambda\, T \ll 1$) but far off for longer expirations ($\lambda\, T > 1$). The approximation including σ_2 up to order τ^2 is almost exact for the shorter expiration $T = 0.25$ and much closer to the true implied volatility for the longer expiration $T = 1.0$. On this scale, Henry-Labordère's first-order approximation is indistinguishable from our first-order approximation (including σ_1 but excluding σ_2), consistent with the tiny errors shown in Figures 4.1 and 4.2.

APPENDIX A: PRINCIPLE OF NOT FEELING THE BOUNDARY

Consider a one-dimensional diffusion process

$$dX_t = a(X_t)\, dW_t + b(X_t)\, dt$$

on $[0, \infty)$, where the continuous function $a(x) > 0$ for $x > 0$. We assume that b is continuous on $\mathbb{R}_+$. We do not make any assumption about the behavior of $a(y)$ as $y \downarrow 0$. Let $d(a, b)$ be the distance between two points $a, b \in \mathbb{R}_+$ determined by $1/a$. If say $a < b$, then

$$d(a, b) = \int_a^b \frac{dx}{a(x)}.$$

Let

$$\tau_c = \inf\{t \geq 0 :\ X_t = c\}.$$

LEMMA A.1. *Suppose that* $x > 0$ *and* $c > 0$. *Then*

$$\lim_{\tau \downarrow 0} \tau \ln \mathbb{P}_x \{\tau_c \leq \tau\} \leq -\frac{d(x, c)^2}{2}.$$

Proof. Let $Y_t = d(X_t, x)$. By Itô's formula we have

$$dY_t = dW_t + \theta(Y_t)\,dt,$$

where

$$\theta(y) = \frac{b(z)}{a(z)} - \frac{1}{2}a'(z), \qquad y = d(z, x).$$

Without loss of generality we assume that $c > x$. It is clear that

$$\mathbb{P}_x \left\{\tau_c^X \leq \tau\right\} = \mathbb{P}_0 \left\{\tau_D^Y \leq \tau\right\}, \qquad D = d(x, c).$$

Let θ be the lower bound of the function $\theta(z)$ on the interval $[0, D]$. Then $Y_t \leq D$ for all $0 \leq t \leq \tau$ implies that $W_t \leq D - \theta\tau$ for all $0 \leq t \leq \tau$. It follows that

$$\mathbb{P}_0 \left\{\tau_D^Y \leq \tau\right\} \leq \mathbb{P}_0 \left\{\tau_{D-\theta\tau}^W \leq \tau\right\}.$$

The last probability is explicitly known

$$\mathbb{P}_0 \left\{\tau_\lambda^W \leq \tau\right\} = \frac{\lambda}{\sqrt{2\pi}} \int_0^\tau t^{-3/2} e^{-\lambda^2/2t}\,dt.$$

Using this we have after some routine manipulations

$$\lim_{\tau \downarrow 0} \tau \ln \mathbb{P}_0 \{\tau_{D-\theta\tau} \leq \tau\} \leq -\frac{D^2}{2}.$$

The desired result follows immediately.

Note that we do not need to assume that X does not explode. By convention $\tau_c = \infty$ if X explodes before reaching c, thus making the inequality more likely to be true. □

Let $0 < a < x < b < \infty$. Let f be a nonnegative function on $\mathbb{R}_+$ and suppose that f is supported on $x \geq b$, i.e., $f(y) = 0$ for $y \leq b$. This corresponds to the case of an out-of-the-money call option. Consider the call price

$$v(x, \tau) = \mathbb{E}_x f(X_\tau).$$

We compare this with

$$v_1(x, \tau) = \mathbb{E}_x \{f(X_\tau); \tau < \tau_a\}.$$

Note that v_1 only depends on the values of a on $[a, \infty)$, thus the behavior of a near $y = 0$ is excluded from consideration. We have

$$v(x, \tau) - v_1(x, \tau) = \mathbb{E}_x \{f(X\tau); \tau_a \leq \tau\} \stackrel{\text{def}}{=} v_2(x, \tau).$$

By the Markov property we have

$$v_2(x,\tau) = \mathbb{E}_x \left\{ \mathbb{E}_a f(X_s)|_{s=\tau-\tau_a}; \tau_a \le \tau \right\}.$$

Now because $f(y) = 0$ for $y \le b$, we have

$$\mathbb{E}_a f(X_s) = \mathbb{E}_a \{ f(X_s); \tau_b \le s \}.$$

Using the Markov property again we have

$$\mathbb{E}_a f(X_s) = \mathbb{E}_a \left\{ \mathbb{E}_b f(X_t)|_{t=s-\tau_b}; \tau_b \le s \right\}.$$

We assume that

$$\sup_{0\le t\le 1} \mathbb{E}_b f(X_t) \le C.$$

This assumption is satisfied if we bound the growth rates of f, a, and b at infinity appropriately. A typical case is when f grows exponentially (call option), and a and b grow at most linearly. These conditions are satisfied by all the popular models we deal with. It is clear that we have to make some assumption about the behavior of the data at infinity, otherwise the problem may not even make any sense. Under this hypothesis we have

$$\mathbb{E}_a f(X_s) \le C\mathbb{P}_a \{ \tau_b \le s \}.$$

Now we have

$$v_2(x,\tau) \le C\mathbb{P}_a \{ \tau_b \le \tau \}.$$

It follows from Lemma A.1 that

$$\lim_{\tau\downarrow 0} \tau \ln v_2(x,\tau) \le -\frac{d(a,b)^2}{2}.$$

Recall that

$$v(x,\tau) = v_1(x,\tau) + v_2(x,\tau).$$

The function $v_1(x,\tau)$ does not depend on the values of a near $y = 0$. We can alter the values of a near $y = 0$ and the resulting error is bounded asymptotically by $\exp\left[-d(a,b)^2/2\tau\right]$. Now if the support of f (as a closed set) contains $y = b$, then we can prove, assuming a behaves nicely near $y = 0$ if necessary, that

$$\lim_{\tau\downarrow 0} \tau \ln \mathbb{E}_x f(X_\tau) = -\frac{d(x,b)^2}{2}.$$

Because $d(x,b) < d(a,b)$, we have proved the following principle of not feeling the boundary.

THEOREM A.2. *Let X^1 and X^2 be (a_1,b_1)- and (a_2,b_2)-diffusion processes on $\mathbb{R}_+$, respectively, f a nonnegative function on $\mathbb{R}_+$, and $0 < a < x < b$. Suppose that a_i, b_i, f*

satisfy the conditions stated above. Suppose further that $a_1(y) = a_2(y)$ *for* $y \geq a$. *Then*

$$\limsup_{\tau \downarrow 0} \tau \ln \left| \mathbb{E}_x f(X^1_\tau) - \mathbb{E}_x f(X^2_\tau) \right| \leq -\frac{d(a,b)^2}{2}$$

and

$$\lim_{\tau \downarrow 0} \tau \ln \mathbb{E}_x f(X^i_\tau) = -\frac{d(x,b)^2}{2}.$$

COROLLARY A.3. *Under the same conditions, we have*

$$\lim_{\tau \downarrow 0} \frac{\mathbb{E}_x f(X^1_\tau)}{\mathbb{E}_x f(X^2_\tau)} = 1.$$

See Hsu (1990) for a more general principle of not feeling the boundary for higher dimensional diffusions.

REFERENCES

ANDERSEN, L., and R. BROTHERTON-RATCLIFFE (2001): Extended LIBOR Market Models with Stochastic Volatility, Working Paper, Gen Re Securities.

ANDERSEN, L. B. G. (2008): Option Pricing with Quadratic Volatility: A Revisit, Discussion Paper, Bank of America Securities.

AVELLANEDA, M., D. BOYER-OLSON, J. BUSCA, and P. FRIZ (2002): Reconstructing Volatility, *Risk Mag.* 15, 87–91.

BERESTYCKI, H., J. BUSCA, and I. FLORENT (2002): Asymptotics and Calibration of Local Volatility Models, *Quantit. Finance* 2, 61–69.

BERESTYCKI, H., J. BUSCA, and I. FLORENT (2004): Computing the Implied Volatility in Stochastic Volatility Models, *Commun. Pure Appl. Math.* 57(10), 1352–1373.

DE BRUIN, N. G. (1999): *Asymptotic Methods in Analysis*, Amsterdam: Dover Publications.

CARR, P., and R. JARROW (1990): The Stop-Loss Start-Gain Paradox and Option Valuation: A New Decomposition into Intrinsic and Time Value, *Rev. Finan. Stud.* 3(3), 469–492.

COX, J. (1975): Notes on Option Pricing I: Constant Elasticity of Diffusions, unpublished draft. Palo Alto, CA: Stanford University, September 1975.

DERMAN, E., and I. KANI (1994): Riding on a Smile, *Risk Mag.* 7, 32–39.

DUPIRE, B. (1994): Pricing with a Smile, *Risk Mag.* 7, 18–20.

DUPIRE, B. (2004): A Unified Theory of Volatility, Discussion Paper, Paribas Capital Management. [Reprinted in *Derivative Pricing: The Classic Collection*, edited by Peter Carr, *Risk Books*.]

GATHERAL, J. (2006): *The Volatility Surface: A Practitioner's Guide*, Hoboken, NJ: Wiley Finance Series.

GYÖNGI, I. (1986): Mimicking the One Dimensional Marginal Distributions of Processes Having an Itô Differential, *Probab. Theory Relat. Fields* 71, 501–516.

HADAMARD, J. (1952): *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Yale, New Haven, CT: Dover Publications.

HAGAN, P., and D. WOODWARD (1999): Equivalent Black Volatilities, *Appl. Math. Finance* 6, 147–157.

HAGAN, P., A. LESNIEWSKI, and D. WOODWARD (2004): Probability Distribution in the SABR Model of Stochastic Volatility, Working Paper.

HAGAN, P., D. KUMAR, A. LESNIEWSKI, and D. WOODWARD (2002): Managing Smile Risk, *Wilmott Mag.*, September, 84–108.

HENRY-LABORDÈRE, P. (2005): A General Asymptotic Implied Volatility for Stochastic Volatility Models, Preprint.

HENRY-LABORDÈRE, P. (2008): *Analysis, Geometry, and Modeling in Finance*, Financial Mathematics Series, Boca Raton, FL, London, New York: Chapman & Hall/CRC.

HSU, E. P. (1990): Principle of Not Feeling the Boundary, *Indiana Univ. Math. J.* 39(2), 431–442.

KAC, M. (1951): On Some Connections between Probability Theory and Differential and Integral Equations, in *Proceedings of the Second Berkeley Symposium on Math. Stat. and Prob.*, Berkeley and Los Angeles: University of California Press, pp. 189–215.

KUNIMOTO, N., and A. TAKAHASHI (2001): The Asymptotic Expansion Approach to the Valuation of Interest Rate Contingent Claims, *Math. Finance* 11(1), 117–151.

LESNIEWSKI, A. (2002): WKB Method for Swaption Smile, Courant Institute Lecture.

LIPTON, A. (2002): The Volatility Smile Problem, *Risk Mag.*, February 61–65.

MEDVEDEV, A., and O. SCAILLET (2007): Approximation and Calibration of Short-term Implied Volatilities under Jump-diffusion Stochastic Volatility, *Rev. Finan. Stud.* 20, 427–459.

MINAKSHISUNDARAM, S., and A. PLEIJEL (1949): Some Properties of Eigenfunctions of the Laplace Operatoron Riemannnian Manifolds, *Can. J. Math.* 1, 242–256.

MOLCHANOV, S. (1975): Diffusion Processes and Riemannian Geometry, *Russ. Math. Surv.* 30, 11–63.

PITERBARG, V. (2007): Markovian Projection for Volatility Calibration, *Risk Mag.* 4, 84–89.

SHAW, W.T. (1998): *Modelling Financial Derivatives with Mathematica*, Cambridge, UK: Cambridge University Press.

TAKAHASHI, A., K. TAKEHARA, and M. TODA (2009): Computation in an Asymptotic Expansion Method, University of Tokyo, Working Paper CIRJE-F-621.

YOSHIDA, K. (1953): On the Fundamental Solution of the Parabolic Equation in a Riemannian Space, *Osaka Math. J.* 1(1), 65–74.

ZUHLSDORFF, C. (2001): The Pricing of Derivatives on Assets with Quadratic Volatility, *Appl. Math. Finance* 8, 235–262.