Small-time asymptotics under local-stochastic volatility with a jump-to-default: curvature and the heat kernel expansion

John Armstrong*   Martin Forde†   Matthew Lorig‡   Hongzhong Zhang§

November 2, 2016

Abstract

We compute a sharp small-time estimate for implied volatility under a general uncorrelated local-stochastic volatility model. For this we use the Bellaiche [Bel81] heat kernel expansion combined with Laplace's method to integrate over the volatility variable on a compact set, and (after a gauge transformation) we use the Davies [Dav88] upper bound for the heat kernel on a manifold with bounded Ricci curvature to deal with the tail integrals. If the correlation $\rho < 0$, our approach still works if the drift of the volatility takes a specific functional form and there is no local volatility component, and our results include the SABR model for $\beta = 1, \rho \leq 0$. For uncorrelated stochastic volatility models, our results also include a SABR-type model with $\beta = 1$ and an affine mean-reverting drift, and the exponential Ornstein-Uhlenbeck model. We later augment the model with a single jump-to-default with intensity λ, which produces qualitatively different behaviour for the short-maturity smile; in particular, for $\rho = 0$, log-moneyness $x > 0$, the implied volatility increases by $\lambda f(x)t + o(t)$ for some function $f(x)$ which blows up as $x \searrow 0$. Finally, we compare our result with the general asymptotic expansion in Lorig, Pagliarani & Pascucci [LPP15], and we verify our results numerically for the SABR model using Monte Carlo simulation and the exact closed-form solution given in Antonov & Spector [AS12] for the case $\rho = 0$. [1]

1   Introduction

In the physics literature, a very convenient form for the heat kernel was originally given by de Witt [dW65] (see also McAvity & Osborn [MO91]). We can re-write the second order elliptic operator associated with a general diffusion process on $\mathbb{R}^n$ in terms of the Laplace-Beltrami operator plus a first order differential operator (i.e. a vector field); the heat kernel expansion is obtained as the exponential of the work done by the vector field along the geodesic joining the start and the end points, multiplied by the Minakshisundaram-Pleijel [MP49] heat kernel expansion for the usual Laplace-Beltrami operator, which contains the leading order exponential term from large deviations theory multiplied by the square root of the Riemannian volume form element under geodesic spherical coordinates (see also Chavel [Chav84], Hsu [Hsu02], Laurence [Laur10], Neel [Neel07], Molchanov [Mol75]). [Mol75] provides a rigorous probabilistic proof of the de Witt expansion at leading order, using a Girsanov change of measure and conditioning on the end point of the process i.e. considering the bridge process. Bellaiche [Bel81] showed that the Molchanov expansion also holds for non-compact manifolds under certain technical conditions, and it is the latter which we use in this article.

Paulot [Pau10] formally derived a small-time expansion for call options under a general local-stochastic volatility model by applying Laplace's method to integrate the heat kernel over the range of the volatility variable, then using the Tanaka-Meyer formula and some well known asymptotic expansions for the standard normal distribution function. [Pau10] also computes explicit formulae for the well known SABR model. Henry-Labordére [HL08] formally derived a small-time expansion for the implied volatility from a small-time expansion for the *effective local volatility*. However, both authors do not justify integrating over the infinite range of the volatility variable with appropriate tail estimates (which is needed because the aforementioned small-time heat kernel expansions are only known to converge uniformly

on compact sets). This is the main technical issue which we resolve in this article, under suitable conditions on the vol-of-vol coefficient $\alpha(y)$ and drift coefficient $\mu(y)$ as $y \to 0$ and as $y \to \infty$.

[FJ11] characterize the small-time behaviour of the implied volatility (at leading order) for a local-stochastic volatility model with zero correlation, using the Freidlin-Wentzell theory of large deviations for SDEs and then converting to the differential geometry problem of computing the shortest geodesic from a point to a vertical line on a Riemannian manifold. The volatility is assumed to be bounded, which means that the curvature can change sign (unlike the SABR model), and the solution to this variable endpoint problem is obtained using conserved quantities which arise from integrating the geodesic equations and a transversality condition, where the shortest geodesic comes in perpendicular to the vertical line under the aforementioned metric. The small-time behaviour of the price of an out-of-the-money European call option is computed using Hölder's inequality, and this is then translated into a statement about the small-time behaviour of the implied volatility. [FJ11] also derive a series expansion for the implied volatility in the small-maturity limit, in powers of the log-moneyness, and they show how to calibrate such a model to the observed implied volatility smile in the small-maturity limit.

Gatheral et al. [GHLOW12] consider small-time asymptotics for a one-dimensional local volatility model, using Girsanov's theorem and conditioning with a bridge process to derive a small-time expansion for the transition density which holds uniformly in $\mathbb{R}$. They also derive the corresponding expansion for the implied volatility. When the diffusion coefficient is time-dependent, they find that even the leading order term in the expansion requires a small but important modification. For a time-homogenous one-dimension diffusion process $dS_t = S_t \sigma(S_t) dW_t$, they prove the following asymptotic expansion for the implied volatility $\hat{\sigma}(K, t)$ at strike K and time-to-maturity t

$$\hat{\sigma}(K, t) \;=\; \hat{\sigma}_0(K) + \frac{\hat{\sigma}_0(K)^3}{\left(\log \frac{K}{S_0}\right)^2} \log \frac{\sqrt{\sigma(S_0)\sigma(K)}}{\hat{\sigma}_0(K)} t \;+\; O(t^2)$$

as $t \to 0$, where $\hat{\sigma}_0(K) = \left(\frac{1}{\log \frac{K}{S_0}} \int_{S_0}^{K} \frac{du}{u\sigma(u)}\right)^{-1}$ is the well known leading order term (see also Busca et al. [BBF02], [BBF04]).

Deuschel et al. [DFJV11], [DFJV11b] use Laplace's method on Wiener space in the same spirit as Azencott, Bismut and Ben Arous [BA88] to compute a small-noise expansion for the density of the canonical projection into $\mathbb{R}^m$ of an n-dimensional hypoelliptic diffusion process X_t^ε. This is accomplished using the Ben Arous expansion applied to the characteristic function of X_t^ε combined with a Fourier inversion. This has the advantage over the heat kernel expansion that the diffusion coefficient need not be uniformly elliptic. However, they do not compute the pre-factor that goes in front of the expansion, which is needed to compute the correction term for implied volatility in the small-time limit, which is computed in this article. More recently, Friz & deMarco [FdM13] consider a stochastic volatility model governed by a hypoelliptic diffusion satisfying a strong Hormander condition, and in this setting they compute a Varadhan-type formula for the small-time behaviour of the stock price density and characterize the small-time behaviour of the law of the terminal stock price conditioned on the terminal volatility, from which we can then compute the effective local volatility $\mathbb{E}(\sigma_t^2 | S_t = K)$ which is a fundamental quantity in the mimicking theorems of Gyöngy [Gyö86] and the more recent work of Brunick & Shreve [BS13].

Lorig et al. [LPP15] derive a full asymptotic expansion for the price of a call option and the associated implied volatility under a general class of local-stochastic volatility models, and provide a rigorous error bound under a uniform ellipticity condition on the diffusion coefficient for the model. Their error bound is obtained using Duhamel's principle and classical estimates by Friedman on partial derivatives of the fundamental solution to the inhomogeneous heat equation $[\partial_t + a(x)\partial_x^2]u = 0$ in terms of the fundamental solution to the standard (homogenous) heat equation $[\partial_t + \frac{1}{2}\sigma^2\partial_x^2]u = 0$ for σ constant. The uniform ellipticity condition is relaxed in Pagliarani & Pascucci [PP14], who consider a general class of degenerate (i.e. non-uniformly parabolic) PDEs, which includes the CEV, Heston and SABR models, and hybrid credit-equity models such as the JDCEV model, and they again derive a rigorous error bound for small-times.

1.1 Outline of article

In Theorem 2.2, we recall the Bellaiche [Bel81] small-time heat kernel expansion. This is the main result that is used for proving the main result (Theorem 4.6) where we compute a small-time expansion for non at-the-money call options under a general local-stochastic volatility model with zero correlation. The pre-factor in the heat kernel expansion is expressed in terms of the Jacobian of the exponential map, which is a ratio of the pullback of two volume forms at

the start and the end point of the geodesic. We show how this calculation is simplified by working in geodesic normal coordinates, and explain the geometric meaning of curvature as the first order deviation from the Euclidean metric in geodesic normal coordinates.

In section 3 we introduce our local-stochastic volatility model, and we compute the Laplace-Beltrami operator, the metric and the curvature associated with the model (the metric is induced by the inverse of the diffusion coefficient matrix). We then state the technical assumptions on the coefficients in the SDEs, the most important of which is that $\alpha(y) \sim A_1 y$ as $y \to 0$ and $\alpha(y) \sim B_1 y^p$ as $y \to \infty$ for some constants $A_1, B_1 > 0$ and $p \in (0, 1]$, where $\alpha(y)$ is the vol-of-vol coefficient. Here $\sim$ means that the ratio between the two sides tends to 1 in the limit. This ensures that the associated Riemannian manifold on the upper half plane is complete: the distance to $y = 0$ and $y = \infty$ under the metric g_{ij} is infinite, which ensures that $y = 0$ and $y = \infty$ are unattainable boundaries for the volatility process Y_t. We then impose that the manifold M has negative curvature (recall that in two dimensions the sectional and the Gaussian curvature are the same, which by Hadamard's theorem, implies that the cut locus of M is empty), and we discuss some simple well known examples of parametric stochastic volatility models.

In Theorem 4.6, we give the main result of the article - a small-time expansion for non at-the-money call options under the aforementioned model. This effectively sharpens the result obtained in [FJ11] and relaxes the assumption of bounded volatility to allow for more realistic tail behaviour (e.g. moment explosions). Our model setup can include an extended SABR-type model with $\rho = 0$ and an affine mean-reverting drift coefficient and β parameter equal to 1, and the exponential Ornstein-Uhlenbeck model (see Table 1). The proof follows the steps of Paulot [Pau10] but with full rigour. We use Laplace's method to integrate the two-dimensional heat kernel with respect to Lebesgue measure multiplied by the local volatility squared to compute a small-time expansion for $\frac{1}{dx}\mathbb{E}(\sigma(X_t)^2 Y_t^2 1_{X_t \in dx})$ where Y_t is the volatility and X_t is the log forward price. The tail integrals are dealt with using the Davies [Dav88] upper bound for the heat kernel on a Riemannian manifold with Ricci curvature bounded from below combined with a gauge transformation. We then use the Tanaka-Meyer formula to estimate the price of a call option in terms of $\frac{1}{dx}\mathbb{E}(\sigma(X_t)^2 Y_t^2 1_{X_t \in dx})$ by integrating over time from zero to the maturity of the option, and using well known asymptotic results for the standard normal distribution function. This trick using the gauge transformation only works when $\rho = 0$, unless we impose a specific functional form for the drift of the volatility process and we assume that the local volatility function $\sigma(x)$ is constant (see subsection 4.4 for details on this). In particular we show that SABR model with $\beta = 1, \rho \le 0$ can still be handled using this trick. To the best of our knowledge, this is the first rigorous analysis of the small-time correction term for implied volatility under SABR model (the leading order term is computed using viscosity solutions in [BBF04]).

In Appendix B, we discuss how to explicitly compute the drift correction term that appears in the Bellaiche expansion, and when there is no local volatility component, this term takes an especially simple form. In principle, we can also formally derive a similar expansion for at-the-money call options, for which the small-time behaviour is qualitatively different, but this requires knowledge of the next term in Bellaiche heat kernel expansion (which is not given by Bellaiche), and we do not have a published reference for this next term, aside from slides by Laurence [Laur08], so we defer the details for future work.

In Proposition 4.8, we derive the corresponding expansions for call options using the Black-Scholes call option formula but with a time-dependent volatility function. This is needed in section 5, where we derive the correction terms for the implied volatility of non at-the-money options. The correction term for implied volatility is important because it takes account of the drift terms in the SDEs, which the result in [FJ11] fails to capture because Freidlin-Wentzell theory only works on a crude logarithmic scale. The correction term is also required to accurately approximate the price of a call option at small maturities. In section 7 we give closed-form formulae for all expressions of interest for the well known SABR model, and we verify our implied volatility expansion numerically by comparing against Monte Carlo simulations and the closed-form expression for the price of a call option under the SABR model with $\rho = 0, \beta = 1$ given in [AS12]. We find they are all in very close agreement. Finally, in section 8, we enrich the model by adding a single Poisson jump-to-default with hazard rate λ, and when $\rho = 0, \sigma(x) \equiv 1$, we show that the jump-to-default increases the implied volatility at log-moneyness x by the following amount

$$\lambda \frac{\hat{\sigma}^2(x)}{|x|\, y_1^*(x)} t + o(t) \tag{1}$$

for $x > 0$ as $t \to 0$, where $\hat{\sigma}(x)$ is the zeroth order implied volatility and $y_1^*(x_1)$ is the y-value of the intersection point for the shortest geodesic from the point (x_0, y_0) to the vertical line $\{x = x_1\}$ under the metric induced by the diffusion coefficient for the model, and the correction term in (1) blows up as the log-moneyness $x \searrow 0$.

3

Throughout the article, we write $a = o(b)$ if and only if the limit of a/b is 0 as some parameter of interest tends to zero or infinity (depending on the context), and write $a = O(b)$ if and only if $\limsup |a/b| < \infty$. If $a = O(b)$ and $b = O(a)$ hold simultaneously, we write $a \asymp b$.

2   The Heat kernel expansion

Consider a diffusion process on $\mathbb{R}^n$ with infinitesimal generator L. In local coordinates, L takes the form

$$L = \frac{1}{2} \sum_{1 \leq i,j \leq n} a^{ij}(\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{1 \leq i \leq n} b^i(\mathbf{x}) \frac{\partial}{\partial x_i} \quad \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Let $M = \mathbb{R}^n$ with metric $(g_{ij}) = (a^{ij})^{-1}$ so that M is a smooth Riemannian manifold with a single chart given by the identity map. We can write L as $\frac{1}{2}\Delta + \mathcal{A}$, where $\Delta = \sum_{i,j} \frac{1}{\sqrt{|g|}} \partial_i(\sqrt{|g|}\, g^{ij} \partial_j)$ is the Laplace-Beltrami operator and $\mathcal{A}^i = b^i - \frac{1}{2}\sum_j \frac{1}{\sqrt{|g|}} \partial_j(\sqrt{|g|}\, g^{ij})$ is a smooth first order differential operator and $|g| = |\det g_{ij}|$ (recall that $(g^{ij}) = (g_{ij})^{-1}$).

The heat kernel of L is a continuous function $p_t(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}, \mathbf{y}, t)$ defined on $M \times M \times (0, \infty)$ which is twice differentiable in $\mathbf{x}$ and once differentiable in t, which satisfies the backward Kolmogorov equation

$$(-\partial_t + L)p = 0, \tag{2}$$

[2] such that for any bounded continuous function f with compact support we have

$$\lim_{t \to 0} \int_M \hat{p}_t(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y} = f(\mathbf{x}), \quad \mathbf{y} = (y_1, \ldots, y_n) \in M, \tag{3}$$

where $\hat{p}_t(\mathbf{x}, \mathbf{y}) := p_t(\mathbf{x}, \mathbf{y})\sqrt{|g(\mathbf{y})|}$ [3](see e.g. page 135 in Chavel[Chav84] for details, or Definition 5.7.9 in [KS91] for a non-geometric reference). $\hat{p}_t(\mathbf{x}, \mathbf{y})$ is the probability density of X_t conditioned on $X_0 = \mathbf{x}$, with respect to Lebesgue measure $d\mathbf{y} = dy_1 \ldots dy_n$ (see e.g. equation (5.7.26) in [KS91]). Furthermore, if $\mathcal{A} = 0$, then $p_t(\mathbf{x}, \mathbf{y})$ is symmetric in $\mathbf{x}$ and $\mathbf{y}$ (see e.g. Theorem 1, chapter VI on page 138 of [Chav84]).

Remark 2.1 For a simple example to show the connection between the heat kernel $p_t(\mathbf{x}, \mathbf{y})$ and the transition density $\hat{p}_t(\mathbf{x}, \mathbf{y})$, consider a one-dimensional diffusion which satisfies $dX_t = \frac{1}{2}\sigma^2 X_t dt + \sigma X_t dW_t$ and $X_0 = x \in \mathbb{R}_+$, where $\sigma > 0$ is a constant and W is a standard Brownian motion. Then one can easily verify that the Laplace-Beltrami operator $\Delta = \sigma^2 x\, \partial_x + \sigma^2 x^2 \partial_x^2$ and $\mathcal{A} = 0$. Moreover, the Riemmanian metric g_{ij} on $\mathbb{R}$ associated with this model has a single element $g_{11}(x) = \frac{1}{\sigma^2 x^2}$, so the pre-factor $\sqrt{|g(\mathbf{y})|}$ is $\frac{1}{\sigma y}$. On the other hand, we know that the transition density for X is given by $\hat{p}_t(x, y) = \frac{1}{\sigma y \sqrt{2\pi t}} e^{-(\log \frac{y}{x})^2/(2\sigma^2 t)}$. Thus, the heat kernel is $p_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-(\log \frac{y}{x})^2/(2\sigma^2 t)}$, a function that is symmetric in x and y (see e.g. Chapters 5 and 6 of [HL08] for more examples of this nature).

Intuitively, one expects that the heat kernel on a Riemannian manifold should be a deformation of the heat kernel on Euclidean space. Molchanov [Mol75] made this idea precise by providing a small-time expansion for the heat kernel on a compact manifold. Subsequent authors have extended this result to more general manifolds. We state two such results.

Theorem 2.1 *(Theorem 5.1.1 in [Hsu02]). Let M be a complete n-dimensional Riemannian manifold. Let $C(M)$ be the subset of points $(\mathbf{x}, \mathbf{y})$ in $M \times M$ such that $\mathbf{x}$ lies in the cut locus of $\mathbf{y}$, which we denote by $\mathrm{Cut}(\mathbf{y})$. Let $d(\mathbf{x}, \mathbf{y})$ be the Riemannian distance between two points $(\mathbf{x}, \mathbf{y}) \in (M \times M) \setminus C(M)$. Let $p_t(\mathbf{x}, \mathbf{y})$ denote the heat kernel of $\frac{1}{2}\Delta$ on M (i.e. $\mathcal{A} = 0$). Then there exist smooth functions $u_n(\mathbf{x}, \mathbf{y})$ defined on $(M \times M) \setminus C(M)$ such that the following asymptotic expansion*

$$p_t(\mathbf{x}, \mathbf{y}) \sim (2\pi t)^{-n/2} e^{-d(\mathbf{x},\mathbf{y})^2/(2t)} \sum_{i=0}^{\infty} u_i(\mathbf{x}, \mathbf{y}) t^i \tag{4}$$

---

[2]Technically this is a forward equation in the t variable and a backward equation in $\mathbf{x}$, but it becomes a forward equation if we make the usual $t \mapsto T - t$ transformation.

[3]Note that we typically write (3) informally as $\hat{p}_0(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{y} - \mathbf{x})$, but (3) really says that $\hat{p}_t(\mathbf{x}, \mathbf{y}) d\mathbf{y}$ tends to the Dirac measure $\delta_{\mathbf{x}}(\mathbf{y})$ in the sense of weak convergence as $t \to 0$.

holds uniformly as $t \to 0$ over any compact subset of $(M \times M) \setminus C(M)$. Let $\exp_{\mathbf{x}} : T_{\mathbf{x}} M \longrightarrow M$ be the exponential map based at $\mathbf{x}$, then we have

$$u_0(\mathbf{x}, \mathbf{y}) = (J(\exp_{\mathbf{x}})(Y))^{-1/2}, \qquad Y = \exp_{\mathbf{x}}^{-1} \mathbf{y}. \tag{5}$$

Here $J(\exp_{\mathbf{x}})$ denotes the Jacobian of the exponential map where we use the flat metric induced by g on $T_{\mathbf{x}} M$ to define the Jacobian (see Remark 2.2 below).

Remark 2.2 Recall that if $f : M_1 \longrightarrow M_2$ is a differentiable map between oriented Riemannian n-manifolds (M_1, g_1) and (M_2, g_2), then $J(f)(p)$ is defined to be the ratio of the pullback of the volume form on M_2 to the volume form on M_1 at p. Let us explain this definition in detail. The differential of f at p defines a map $f_* : T_p M_1 \longrightarrow T_q M_2$ where $q = f(p)$. The pullback $f^* : \Lambda^k T_q^* M_2 \longrightarrow \Lambda^k T_p^* M_1$ is then defined by:

$$(f^* \mu)(v_1 \wedge \ldots \wedge v_k) = \mu(f_* v_1, \ldots, f_* v_k), \tag{6}$$

where $v_1, \ldots v_k \in T_p M_1$ and μ is a k-form. Now take local coordinates x for M_1 and y for M_2 centered on p and q. The space $\Lambda^n T_p^*$ is one dimensional and hence spanned by the volume form $\sqrt{|g_1|} \, dx^1 \wedge \ldots \wedge dx^n$. Thus for some $\lambda \in \mathbb{R}$ we have:

$$f^*(\sqrt{|g_2|} \, dy^1 \wedge \ldots \wedge dy^n) = \lambda \sqrt{|g_1|} \, dx^1 \wedge \ldots \wedge dx^n. \tag{7}$$

By definition $J(f)(p)$ is equal to λ. See subsections 2.1 and 2.2 to see how the expression for $J(\exp_{\mathbf{x}})(Y)$ simplifies when we work in geodesic normal coordinates.

We now recall the following extension of Theorem 2.1 by Bellaiche [Bel81]:

Theorem 2.2 *(Theorem 4.1 in [Bel81]). Let M be a C^4-Riemannian manifold and $\mathcal{A}$ a C^4-vector field. Then the heat kernel $p_t(\mathbf{x}, \mathbf{y})$ of the operator $\frac{1}{2}\Delta + \mathcal{A}$ satisfies:*

$$p_t(\mathbf{x}, \mathbf{y}) \sim (2\pi t)^{-n/2} u_0(\mathbf{x}, \mathbf{y}) \, e^{-\frac{1}{2} d(\mathbf{x}, \mathbf{y})^2 / t + A(\mathbf{x}, \mathbf{y})} \qquad (t \to 0) \tag{8}$$

for $(\mathbf{x}, \mathbf{y}) \in (M \times M) \setminus C(M)$. Here u_0 is defined as in (5) and

$$A(\mathbf{x}, \mathbf{y}) := \int_0^1 \langle \mathcal{A}, \dot{\gamma}(s) \rangle \, ds \tag{9}$$

and γ is the unique distance-minimizing geodesic [4] $\gamma : [0,1] \longrightarrow M$ joining $\mathbf{x}$ and $\mathbf{y}$. The estimate (8) is uniform on any compact subset of $(M \times M) \setminus C(M)$.

2.1 Geodesic normal coordinates and the geometric meaning of curvature

Now let $e_1, .., e_n$ be a basis of $T_p M$ which is orthonormal with respect to the scalar product on $T_p M$ induced by the metric g_{ij}. For each vector $v \in T_p M$, writing its components with respect to this basis, we obtain a map $\phi : T_p M \mapsto \mathbb{R}^n$, $v = v^i e_i \mapsto (v^1, ..., v^n)$. Then one has the associated geodesic normal coordinate system y given by $y = \phi \circ \exp_{\mathbf{x}}^{-1}$ (see page 21 in [Jost09]). In these coordinates, all first order partial derivatives of the metric vanish at zero, i.e. $g_{ij,k}(0) = 0$ for all i, j, k, and $\Gamma_{jk}^i = 0$, and the metric has the following Taylor expansion:

$$g_{ij} = \delta_{ij} - \sum_{a,b} \frac{1}{3} (R_{iajb} + R_{ibja}) y_a y_b + O(|y|^3)$$

(see e.g. [MA73]), where R_{jkl}^i is the Riemann curvature tensor and we are following standard conventions for raising and lowering indices using the metric. This formula provides the basic geometric interpretation of curvature as the deviation of the metric from the Euclidean metric in normal coordinates. Indeed, Riemann originally introduced curvature using such an expansion; the definition using the Levi-Civita connection was only introduced later.

---

[4] All geodesics have constant speed, and here we are just choosing the speed so the geodesic reaches point $\mathbf{y}$ in unit time

2.2 Calculating $J(\exp_{\mathbf{x}})(Y)$ and $u_0(\mathbf{x}, \mathbf{y})$ in normal coordinates

If we now consider Remark 2.2 for special case when $M_1 = T_{\mathbf{x}}M$, $M_2 = M$, $f = \exp_{\mathbf{x}}$, $p = \mathbf{x}$ and $q = \mathbf{y} = \exp_{\mathbf{x}}(Y)$, and take coordinates for the tangent space $\phi : M_1 = T_{\mathbf{x}}M \to \mathbb{R}^n$ as above and normal coordinates $y : M \to \mathbb{R}^n$ on M, then written in these coordinates f and f_* are just the identity function. The metric $g_1 = \delta_{ij}$. The metric $g_2 = g_{ij}$ is the metric associated with using normal coordinates on M (which is the identity at $\mathbf{x}$). Equation (7) becomes:

$$\exp^* \left(\sqrt{|g|} \, dy^1 \wedge ... \wedge dy^n \right) = \lambda \, d\phi^1 \wedge ... \wedge d\phi^n \,. \tag{10}$$

Using that $\exp^*$ is also the identity, we see that in these coordinates $\lambda = J(\exp_{\mathbf{x}})(Y) = \sqrt{|g|}$.

From this we also obtain

$$
\begin{aligned}
u_0(\mathbf{x}, \mathbf{y}) &= (\sqrt{|g|})^{-\frac{1}{2}} \\
&= \left[1 - \sum_{i,a,b} \frac{1}{3}(R_{iaib} + R_{ibia}) y_a y_b + o(|y|^2) \right]^{-\frac{1}{4}} \\
&= 1 + \frac{1}{12} \sum_{i,a,b} (R_{iaib} + R_{ibia}) y_a y_b + o(|y|^2) \,.
\end{aligned}
$$

We now specialize to the two-dimensional case. The symmetries of the curvature tensor tell us that it has only one independent component. At the origin in normal coordinates one has $R_{1212} = R_{2121} = -R_{1221} = -R_{2112} = \kappa$, where κ is the Gaussian curvature of the surface. All the other components of R vanish. Thus we have

$$u_0(\mathbf{x}, \mathbf{y}) = 1 + \frac{1}{12} \kappa |y|^2 + o(|y|^2) \,.$$

Alternatively, one can introduce geodesic polar coordinates $y = (r\cos(\theta), r\sin(\theta))$ in which case one obtains the same expansion but with $|y|$ replaced by r and now $r = d(\mathbf{x}, \mathbf{y})$ is the Riemannian distance.

2.3 Geodesic polar coordinates

Let $n = 2$ and let $f(r, \theta) = \exp_p(rv(\theta))$ for $p \in M$, with $0 < r < i(p)$ (where $i(p) = d(p, C_p)$ is the injectivity radius at p) and $-\pi < \theta \le \pi$, where $|v(\theta)| = 1$, $|v'(\theta)| = 1$. (r, θ) are called geodesic polar coordinates at p. In these coordinates, the metric $\mathbf{g}(r, \theta)$ has coefficients

$$g_{rr} = \left| \frac{\partial f}{\partial r} \right|^2 = |v(\theta)|^2 = 1, \; g_{r\theta} = 0 \,, \; g_{\theta\theta} = \left| \frac{\partial f}{\partial \theta} \right|^2$$

(see page 122 in do Carmo [doC92]), where $|\cdot|$ refers to the norm under the original metric g. Then $J(r) = \frac{\partial f}{\partial \theta}|(r, \theta)$ is a *Jacobi field*, and J has the asymptotic behaviour $|J|^2 = r^2 - \frac{1}{3} \kappa r^4 + o(r^4)$ as $r \to 0$. Thus the metric $\mathbf{g}(r, \theta)$ can be locally approximated as

$$ds^2 \approx dr^2 + r^2 \left(1 - \frac{1}{3} \kappa r^2 \right) d\theta^2 \,,$$

for $r \ll 1$, and we see that the curvature describes the departure of the metric from the usual metric $ds^2 = dr^2 + r^2 d\theta^2$ for polar coordinates in $\mathbb{R}^2$.

Remark 2.3 The heat kernel can be constructed geometrically by the method of parametrix starting from an approximate heat kernel in local coordinates. Page 148 in Chavel [Chav84] gives a nice sketch proof using geodesic polar coordinates. In these coordinates, we can write the $u_0(\mathbf{x}, \mathbf{y})$ term in the heat kernel expansion as

$$u_0(\mathbf{x}, \mathbf{y}) = \left(\frac{\sqrt{\mathbf{g}(r, \theta)}}{r} \right)^{-\frac{1}{2}} \,.$$

Remark 2.4 If $d(\mathbf{x}, \mathbf{y}) < d(\mathbf{x}, \text{Cut}(\mathbf{x}))$, then in local coordinates $u_0(\mathbf{x}, \mathbf{y}) = \sqrt{\Delta^{\text{VVM}}(\mathbf{x}, \mathbf{y})}$, where

$$\Delta^{\text{VVM}}(\mathbf{x}, \mathbf{y}) = g(\mathbf{x})^{-\frac{1}{2}} \det\left(-\frac{\partial^2 \phi(\mathbf{x}, \mathbf{y})}{\partial x_i \partial y_j}\right) g(\mathbf{y})^{-\frac{1}{2}} \tag{11}$$

is the so-called *van Vleck-Morette determinant* (see McAvity & Osborn [MO91] and equation (4.38) in Vassilevich [Vass03]), and $\phi(\mathbf{x}, \mathbf{y}) = \frac{1}{2}d(\mathbf{x}, \mathbf{y})^2$. (11) is useful when we can compute $d(\mathbf{x}, \mathbf{y})$ explicitly by solving the geodesic equations.

3 Local-stochastic volatility models

We work with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ throughout, with a filtration $\mathcal{F}_t$ supporting two independent Brownian motions which satisfies the usual conditions.

We now consider a general uncorrelated local-stochastic volatility model for a forward price process S_t defined by the following stochastic differential equations for $X_t = \log S_t$:

$$\begin{cases} dX_t = -\frac{1}{2}\sigma(X_t)^2 Y_t^2 dt + \sigma(X_t) Y_t \, dW_t^1, \\ dY_t = \mu(Y_t)dt + \alpha(Y_t)dW_t^2, \end{cases} \tag{12}$$

where $(X_0, Y_0) = (x_0, y_0) \in \mathbb{R} \times \mathbb{R}_+$, and W^1, W^2 are two independent standard Brownian motions. We need to impose that the correlation is zero for the gauge transformation trick in subsection 4.1 to work. However, the presence of the local volatility component $\sigma(x)$ can still produce an implied volatility skew.

We let M denote the upper half plane $\{(x, y) : y > 0\}$ with Riemmanian metric $(g_{ij}) = (a^{ij})^{-1}$, where a_{ij} is the diffusion coefficient for the model in (12), so the line element for g is given by

$$ds^2 = \sum_{i,j} g_{ij} dx_i dx_j = \frac{1}{\sigma(x)^2 y^2} dx^2 + \frac{1}{\alpha(y)^2} dy^2, \tag{13}$$

and the Laplace-Beltrami operator $\Delta = \sum_j \frac{1}{\sqrt{|g|}} \partial_j (\sqrt{|g|} \, g^{ij})$ satisfies

$$\frac{1}{2}\Delta = \frac{1}{2}y^2\sigma'(x)\sigma(x)\partial_x + \frac{1}{2}\left[\alpha'(y)\alpha(y) - \frac{\alpha(y)^2}{y}\right]\partial_y + \frac{1}{2}y^2\sigma(x)^2\partial_x^2 + \frac{1}{2}\alpha(y)^2\partial_y^2, \tag{14}$$

so we have

$$\mathcal{A} = \left[-\frac{1}{2}y^2\sigma(x)^2 - \frac{1}{2}y^2\sigma'(x)\sigma(x)\right]\partial_x + \left[\mu(y) - \frac{1}{2}(\alpha'(y)\alpha(y) - \frac{\alpha(y)^2}{y})\right]\partial_y.$$

Remark 3.1 See Appendix B for details on how to calculate $A(\mathbf{x}, \mathbf{y})$ as defined in (9).

We can easily compute the curvature tensor for the metric g directly from the standard formulae for the Christoffel symbols in local coordinates and standard formulae for the curvature tensor R.[5] We can then compute the Gaussian curvature as

$$\kappa = \frac{R_{1212}}{g_{11}g_{22} - g_{12}^2},$$

and from this we see that for any $(x, y) \in M$, we have[6]

$$\kappa(x, y) = \frac{\alpha(y)(-2\alpha(y) + y\alpha'(y))}{y^2}. \tag{15}$$

(note that $\kappa(x, y)$ does not depend on x or $\sigma(x)$) We now make the following additional assumptions:

---

[5]The Mathematica sheet to calculate the curvature is available on request.
[6]The formula for $\kappa(x, y)$ in (15) is also valid when W^1 and W^2 are correlated.

Assumption 3.1

- μ, α, σ are C^∞ and α, σ are strictly positive and α is strictly increasing and $\sigma \in C_b^2$, and μ, α are such that Y_t has a unique strong solution for any given $Y_0 > 0$ (e.g. Theorem 2.9 in [KS91]). And σ is Lipschitz continuous.

- $\alpha(y) \sim A_1 y$, $\alpha'(y) \sim A_1 = \alpha'(0+)$, $\alpha''(y) \to 0$ as $y \to 0$, and $\alpha(y) \sim B_1 y^p$, $\alpha'(y) \sim B_1 p y^{p-1}$, $\alpha''(y)/y^{p-2} \to B_1 p(p-1)$ as $y \to \infty$ for some constants $A_1, B_1 > 0$ and $p \in (0,1]$. This ensures that the associated Riemannian manifold on the upper half plane is complete - the distance to $y = 0$ and $y = \infty$ under the metric g_{ij} is infinite, and ensures that $y = 0$ and $y = \infty$ are unattainable boundaries. The condition at $y = \infty$ ensures that X_t has a fatter (and thus more realistic) right tail than if we chose a bounded volatility function $f(y)$ as in [FJ11] (see [AP07], [Jour04], [LM07] for more details).

- For all sufficiently small $y > 0$, $\mu(y) \geq 0$ and $\mu(1/y) \leq 0$. Moreover, μ is such that, as $y \to 0$, $\overline{V}(y)$ and $\overline{V}(1/y)$ are bounded from above, where

$$\overline{V}(y) := \mu(y)g(y) + \frac{1}{2}\alpha(y)^2 \left[g(y)^2 + g'(y)\right], \quad \text{with } g(y) := -\frac{\mu(y)}{\alpha(y)^2} + \frac{1}{2}\left(\frac{\alpha'(y)}{\alpha(y)} - \frac{1}{y}\right).$$

  These assumptions are needed to make the Gauge transformation trick with the Davies heat kernel estimate work.

- $0 < \underline{\sigma} \leq \sigma \leq \bar{\sigma} < \infty$ for some constants $\underline{\sigma}, \bar{\sigma}$.

- We assume that $\sigma(x)^2 + \sigma'(x)^2 - 2\sigma(x)\sigma''(x) > 0$ for all x, which is clearly true if σ is constant. This condition is required for the gauge transformation trick to work, and essentially just excludes excessive skew/convexity of the local volatility function $\sigma(x)$.

Proposition 3.2 *Under Assumption 3.1, the system of two-dimensional stochastic differential equations in (12) has a unique strong solution.*

Proof. Follows from standard arguments. ■

Assumption 3.3 *We assume that $-2\alpha(y) + y\alpha'(y) \leq 0$ which implies that $\kappa(x,y) \leq 0 \quad \forall(x,y) \in M$, which (by Hadamard's theorem, see page 149 in [doC92]) implies that the cut locus of M is empty.*

Remark 3.2 We list a few stochastic volatility (SV) models that satisfy both Assumption 3.1 and Assumption 3.3 in Table 1.

Table 1: Examples of SV models

Model	$\mu(y)$	$\alpha(y)$
SABR with $\beta = 1$	0	νy with $\nu > 0$
Mean-reverting SABR with $\beta = 1$	$\eta(\theta - y)$ with $\eta, \theta > 0$	νy with $\nu > 0$
Exponential OU	$\eta(\theta - \log y)y$ with $\eta > 0, \theta \in \mathbb{R}$	νy with $\nu > 0$

Remark 3.3 Using (15), we see that $\kappa(x,y)$ is smooth for all $(x,y) \in M$ and independent of x, and

$$\kappa(x,y) \sim - A_1^2 \qquad\qquad\qquad\qquad (y \to 0),$$

$$\kappa(x,y) \sim - B_1^2(2-p)y^{2(p-1)} \to \begin{cases} -B_1^2 & \text{if } p = 1 \\ 0 & \text{if } p \in (0,1) \end{cases} \qquad (y \to \infty),$$

so κ is bounded from below.

Remark 3.4 Our conditions include the SABR model for $\beta = 1$ (which corresponds to $p = 1$) but not the Heston model, because for the latter the associated manifold is not complete, and completeness is needed for the Davies heat kernel estimate below. Small-time asymptotics for the Heston model are obtained in [FJL12] using Fourier methods and saddle point estimates for contour integrals.

3.1 Tail behaviour of the model

For the SABR model with zero correlation (i.e. $\alpha(y) = \alpha y$, $\mu(y) = 0$), it is well known that for $m > 0$, $\mathbb{E}(S_t^m) < \infty$ if and only if $m \leq 1$. If $d\langle W^1, W^2 \rangle_t = \rho dt$ with $\rho < 0$, this condition changes to $\mathbb{E}(S_t^m) < \infty$ if and only $\rho \leq -\sqrt{(m-1)/m}$ (see Theorem 2.3 in [LM07]). Moreover, by Theorems 2.5 and 2.6 in [LM07], this result also applies to our model in (12) if $p = 1$ because $\alpha(y)$ does not have quadratic growth at $y = \infty$, i.e. the b_∞ term in equation (29) in [LM07] is zero. For $p < 1$, the conditions are more complicated, we refer the reader to Theorem 3.2 in [LM07] for details.

3.2 Example: the SABR model

For the hyperbolic metric $ds^2 = \frac{1}{y^2}(dx^2 + dy^2)$ on the upper half plane $\mathbb{H}^2$ (which is associated with the SABR model $dS_t = S_t^\beta Y_t dW_t^1, dY_t = Y_t dW_t^2, \ dW_t^1 dW_t^2 = \rho dt$ with $\beta = 1, \rho = 0$, see section 7), we have that $\kappa = -1$ (see e.g. Molchanov [Mol75] or chapter 5 in [doC92]). For $\mathbb{H}^2$ we have the simple explicit formula due to McKean [McK70]

$$p_t(\mathbf{x}, \mathbf{y}) \quad = \quad \frac{\sqrt{2}\,e^{-t/8}}{(2\pi t)^{3/2}} \int_{d(\mathbf{x}, \mathbf{y})}^\infty \frac{r e^{-r^2/2t}}{\sqrt{\cosh r - \cosh d(\mathbf{x}, \mathbf{y})}} dr$$

(see e.g. Theorem 3.1 in Matsumoto & Yor [MY05] for the corresponding formula for the n-dimensional hyperbolic space).

4 Small-time asymptotics for call options

4.1 A gauge transformation to remove the $\mathcal{A}$ term

Here and throughout, we let $p_t^0(\mathbf{x}, \mathbf{y})$ denotes the heat kernel associated with the usual Laplace-Beltrami operator (i.e. with $\mathcal{A} = 0$).

The following lemma computes an upper bound for $p_t(\mathbf{x}, \mathbf{y})$ in terms of the heat kernel $p_t^0(\mathbf{x}, \mathbf{y})$ for the case when $\mathcal{A} = 0$. This is needed so we can appeal to the Davies heat kernel estimate that follows.

Lemma 4.1 *We have the following upper bound for $\hat{p}_t(\mathbf{x}, \mathbf{y}) = \sqrt{|g(\mathbf{y})|}\,p_t(\mathbf{x}, \mathbf{y})$:*

$$\hat{p}_t(\mathbf{x}, \mathbf{y}) \quad \leq \quad \frac{\chi(x_0, y_0)}{\chi(x_1, y_1)} e^{V_{\max}t}\,\hat{p}_t^0(\mathbf{x}, \mathbf{y})\,, \tag{16}$$

for some constant $V_{\max} < \infty$, where $\mathbf{x} = (x_0, y_0)$, $\mathbf{y} = (x_1, y_1)$ and

$$\chi(x, y) \quad = \quad \sqrt{\sigma(x)}\,e^{\frac{1}{2}x}\frac{\sqrt{\alpha(y)}}{\sqrt{y}}\,e^{-\int_1^y \frac{\mu(u)}{\alpha(u)^2}du}$$

(see also pages 108-9 in [HL08] for related discussion).

Proof. If we set $(x, y) = (x_0, y_0)$, then we know that $\hat{p}_t(\mathbf{x}, \mathbf{y})$ is a solution to the backward Kolmogorov equation

$$\partial_t \hat{p} \quad = \quad \left(\mathcal{A} + \frac{1}{2}\Delta \right) \hat{p}, \tag{17}$$

subject to $\hat{p}_0(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{y} - \mathbf{x})$, where the spatial partial derivatives in Δ and $\mathcal{A}$ are with respect to the backward variable $\mathbf{x} = (x, y)$. If we now let $\hat{p}_t(\mathbf{x}, \mathbf{y}) = \frac{\chi(x,y)}{\chi(x_1, y_1)} q_t(\mathbf{x}, \mathbf{y})$, then the PDE transforms to

$$\partial_t q \quad = \quad y^2 \sigma(x)^2 \left(\frac{\partial_x \chi}{\chi} - \frac{1}{2} \right) \partial_x q + \left(\mu(y) + \alpha(y)^2 \frac{\partial_y \chi}{\chi} \right) \partial_y q + \frac{1}{2}y^2\sigma(x)^2 \partial_x^2 q + \frac{1}{2}\alpha(y)^2 \partial_y^2 q + V(x, y)\,q$$

$$= \quad \frac{1}{2}\Delta q + V(x, y)\,q \tag{18}$$

with $q_0(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{y} - \mathbf{x})$, where $V(x, y) = \frac{(\mathcal{A} + \frac{1}{2}\Delta)\chi}{\chi}(x, y)$. In the second equality above, we used the fact that

$$\frac{\partial_x \chi}{\chi} = \frac{1}{2}\left(1 + \frac{\sigma'(x)}{\sigma(x)}\right) \equiv f(x), \quad \frac{\partial_y \chi}{\chi} = -\frac{\mu(y)}{\alpha(y)^2} + \frac{1}{2}\left(\frac{\alpha'(y)}{\alpha(y)} - \frac{1}{y}\right) = g(y).$$

In fact, one can use the above equations to obtain that

$$
\begin{aligned}
V(x, y) &= \frac{1}{2}y^2\sigma(x)^2\left[-f(x) + f(x)^2 + f'(x)\right] + \mu(y)g(y) + \frac{1}{2}\alpha(y)^2\left[g(y)^2 + g'(y)\right] \\
&= -\frac{1}{8}y^2\left[\sigma(x)^2 + \sigma'(x)^2 - 2\sigma(x)\sigma''(x)\right] + \mu(y)g(y) + \frac{1}{2}\alpha(y)^2\left[g(y)^2 + g'(y)\right].
\end{aligned}
$$

From the final bullet point in Assumption 3.1 we know that

$$V(x, y) \leq \overline{V}(y) = \mu(y)g(y) + \frac{1}{2}\alpha(y)^2\left[g(y)^2 + g'(y)\right], \quad \forall (x, y) \in \mathbb{R} \times \mathbb{R}_+.$$

Moreover, by Assumption 3.1 we know that $\overline{V}(y)$ is uniformly bounded from above as $y \to 0$ or $y \to \infty$. Thus, we know that $\overline{V}(y)$ is uniformly bounded from above for all $y \in (0, \infty)$ by some constant which we call $V_{\max}$ (which is independent of x), and the same constant bounds $V(x, y)$ from above for all (x, y). Now let $\hat{\mathbb{P}}$ denote the probability measure under which (X_t, Y_t) defined in (12) has infinitesimal generator $L = \frac{1}{2}\Delta$ (i.e. with $\mathcal{A} = 0$). Then for any bounded Borel function f over $\mathbb{R} \times \mathbb{R}_+$, and $\mathbf{x} = (x, y)$, $\mathbf{y} = (x', y')$, we have

$$\int f(\mathbf{y})q_t(\mathbf{x}, \mathbf{y})d\mathbf{y} = \mathbb{E}_{\mathbf{x}}^{\hat{\mathbb{P}}}\left(e^{\int_0^t V(X_s, Y_s)ds}f(X_t, Y_t)\right) = \int f(\mathbf{y})\mathbb{E}_{\mathbf{x}}^{\hat{\mathbb{P}}}\left(e^{\int_0^t V(X_s, Y_s)ds}1_{(X_t, Y_t) \in d\mathbf{y}}\right),$$

where we have used Feynman-Kac formula in the first equality. By the arbitrariness of f, we have that

$$q_t(\mathbf{x}, \mathbf{y})d\mathbf{y} = \mathbb{E}_{\mathbf{x}}^{\hat{\mathbb{P}}}\left(e^{\int_0^t V(X_s, Y_s)ds}1_{(X_t, Y_t) \in d\mathbf{y}}\right) \leq e^{V_{\max}t}\hat{\mathbb{P}}_{\mathbf{x}}\left((X_t, Y_t) \in d\mathbf{y}\right) = e^{V_{\max}t}\hat{p}_t^0(\mathbf{x}, \mathbf{y})d\mathbf{y}.$$

It follows that

$$\hat{p}_t(\mathbf{x}, \mathbf{y}) = \frac{\chi(x_0, y_0)}{\chi(x_1, y_1)}q_t(\mathbf{x}, \mathbf{y}) \leq \frac{\chi(x_0, y_0)}{\chi(x_1, y_1)}e^{V_{\max}t}\hat{p}_t^0(\mathbf{x}, \mathbf{y}).$$

∎

Remark 4.1 The gauge transformation trick here only works when the correlation $\rho = 0$, unless we impose a specific functional form for $\mu(y)$, see subsection 4.4 for details. We would expect a similar result to Theorem 4.6 to hold for $\rho \neq 0$ and general $\mu(y)$, but to prove this would require re-writing the whole of [Dav88] for the more general case when $\mathcal{A} \neq 0$ ([Dav88] only deals with the self-adjoint case when $\mathcal{A} = 0$). Otherwise, we could impose that the volatility is given by $f(y)$ for some bounded function of y, and use the Norris-Stroock [NS91] tail estimate for the fundamental solution to the heat equation with a uniformly elliptic coefficients instead of the Davies estimate, but this type of model would not have realistic fat tail behaviour.

Corollary 4.2 *Using that $\alpha(y) \sim A_1 y$ as $y \to 0$ and $\alpha(y) \sim B_1 y^p$ as $y \to \infty$, we have a positive constant $\underline{C} > 0$ such that:*

1. *For all sufficiently small $y > 0$,*

$$\chi(x, y) \geq \underline{C}\sqrt{A_1}\sqrt{\sigma(x)}\, e^{\frac{1}{2}x};$$

2. *For all sufficiently large $y > 0$,*

$$\chi(x, y) \geq \underline{C}\sqrt{B_1}\sqrt{\sigma(x)}\, e^{\frac{1}{2}x}y^{-\frac{1}{2}(1-p)}.$$

Proof. This is an immediate consequence of the formula given for χ and our assumptions on the asymptotics of μ and α. Note that we use our assumptions on the sign of μ near 0 and ∞ to bound $-\int_1^y \frac{\mu(u)}{\alpha(u)^2}du$ from below. ∎

4.2 The Davies upper bound for the heat kernel

By a simple modification of Theorem 16 in Davies [Dav88] (which deals with the heat equation $\partial_t u - \partial_{xx}^2 u = 0$ without the $\frac{1}{2}$ factor), we have the following:

Theorem 4.3 *If $\mathcal{M}$ is a complete Riemannian manifold of dimension N such that, for some constant $\beta \geq 0$,*

$$\mathrm{Ric}(\mathbf{x}) \geq -(N-1)\beta^2,$$

where Ric denotes the Ricci curvature[7], then there exists a constant c_δ depending on δ such that

$$0 \leq p_t^0(\mathbf{x}, \mathbf{y}) \leq c_\delta \, |B(\mathbf{x}, t^{\frac{1}{2}})|^{-\frac{1}{2}} |B(\mathbf{y}, t^{\frac{1}{2}})|^{-\frac{1}{2}} e^{-d(\mathbf{x},\mathbf{y})^2/(2+\delta)t}$$

for $0 < t < 1$, where $|B(\mathbf{x}, r)|$ denotes the Riemannian volume of the ball $B(\mathbf{x}, r) = \{\mathbf{y} \in M : d(\mathbf{x}, \mathbf{y}) < r\}$ (see also page 198 in [Chav84] for a similar result).

We now return to our manifold M. Let $\mathbf{x} = (x_0, y_0)$, $\mathbf{y} = (x_1, y_1)$ denote two points on M and let $d(x_0, y_0; x_1, y_1) = d(\mathbf{x}, \mathbf{y})$. From the assumption that $\kappa(x, y) \leq 0$ and the Günther volume comparison theorem on page 213 in [Jost09], we have that

$$|B(\mathbf{x}, r)| \geq |B^E(\mathbf{x}, r)| = \pi r^2, \tag{19}$$

where $|B^E(\mathbf{x}, r)|$ denotes the volume of a ball under the standard Euclidean metric. Thus setting $r = t^{\frac{1}{2}}$ we have the following corollary of Theorem 4.3:

Corollary 4.4 *Using Remark 3.3 and Theorem 4.3, we have the following upper bound*

$$p_t^0(\mathbf{x}, \mathbf{y}) \leq \frac{c_\delta}{\pi t} e^{-d(\mathbf{x},\mathbf{y})^2/(2+\delta)t},$$

which (combined with (16)) implies that

$$\hat{p}_t(\mathbf{x}, \mathbf{y}) \leq \frac{\chi(x_0, y_0)}{\chi(x_1, y_1)} e^{V_{\max} t} \sqrt{|g(x_1, y_1)|} \frac{c_\delta}{\pi t} e^{-d(\mathbf{x},\mathbf{y})^2/(2+\delta)t}.$$

Lemma 4.5 *From a simple asymptotic analysis of a vertical line we have*

$$d(\mathbf{x}, \mathbf{y}) \asymp d_0(y_1) \qquad (y_1 \to 0) \quad \text{where } d_0(y_1) := -\frac{1}{A_1} \log y_1,$$

$$d(\mathbf{x}, \mathbf{y}) \asymp d_\infty(y_1) \qquad (y_1 \to \infty) \quad \text{where } d_\infty(y_1) := \begin{cases} \frac{1}{B_1} \log y_1, & (p = 1), \\ \frac{y_1^{1-p}}{B_1(1-p)}, & (p \in (0, 1)) \end{cases}$$

for $\mathbf{x} = (x_0, y_0)$ and x_1 fixed (where $\mathbf{y} = (x_1, y_1)$).

4.3 Small-time expansion for non at-the-money call options

In Theorem 4.6, we state the main result in the paper: a small-time expansion for out-of-the-money call options under the general local-stochastic volatility model in (12). To prove this result, we proceed along similar lines to section 3 in [Pau10]. We introduce the following notation:

$$\phi(y_1) = \frac{1}{2} d(\mathbf{x}, \mathbf{y})^2,$$

$$\psi(y_1) = y_1^2 \, \mathcal{P}(\mathbf{x}, \mathbf{y}) \, u_0(\mathbf{x}, \mathbf{y}) \sqrt{|g(x_1, y_1)|},$$

where $\mathcal{P}(\mathbf{x}, \mathbf{y}) = e^{A(\mathbf{x}, \mathbf{y})}$, $\mathbf{x} = (x_0, y_0)$, $\mathbf{y} = (x_1, y_1)$. Then the following theorem characterizes the small-time behaviour of non-at-the-money call options.

---

[7]The Ricci curvature is just equal to the Gaussian curvature when the dimension $n = 2$, which is also equal to the sectional curvature.

Theorem 4.6 *Consider the stochastic volatility model defined in* (12) *and assume the initial stock price* $S_0 = 1$.[8] *Then we have the following small-time expansion for the price of a call option with strike* $K \neq S_0$:

$$\mathbb{E}(S_t - K)^+ - (S_0 - K)^+ \quad \sim \quad \frac{A^{\mathrm{SV}}(x_1)}{\sqrt{2\pi}} e^{-\phi(y_1^*)/t} t^{\frac{3}{2}} \qquad (t \to 0),$$

where $x_1 = \log K$,

$$A^{\mathrm{SV}}(x_1) \quad = \quad \frac{K\sigma(x_1)^2 \psi(y_1^*)}{\sqrt{\phi''(y_1^*)}} \frac{1}{2\phi(y_1^*)}$$

and $y_1^* = y_1^*(x_1)$ *is the y-value where the shortest geodesic from* (x_0, y_0) *hits the line* $\{x = x_1\}$ *under the metric* g_{ij} *in* (13).

Remark 4.2 Because the curvature $\kappa \leq 0$, from the argument on page 209 in [doC92], we know that there is a unique distance minimizing geodesic from (x_0, y_0) to the line $\{x = x_1\}$, and we have the *transversality* condition $\mathbf{g}(\frac{d\gamma^*}{dt}, (0,1))|_{(x_1, y_1^*)} = 0$, i.e. the shortest geodesic comes in perpendicular to the vertical line under the metric g_{ij} (see page 14 in [FJ11] for more details on this point). Moreover, because the correlation $\rho = 0$, the shortest geodesic is also perpendicular in the usual Euclidean sense.

Remark 4.3 There are semi-explicit formulae for computing $y_1^*(x)$ and $\phi(y_1^*)$; these are given by two integral equations in equations (26) and (27) in the prequel article [FJ11]. More specifically, we first solve for $y_1^*(x)$ numerically in equation (27) which is just a line search i.e. a one-dimensional root-finding exercise, and we then plug $y_1^*(x)$ into equation (26) to compute the distance of this shortest geodesic to the vertical line $\{x = x_1\}$. These calculations are quite low level and tedious, so we do not repeat these calculations in this article.

Proof. Applying the generalized change-of-variable formula for semimartingales (Theorem 3.7.1, part (v) in Karatzas & Shreve [KS91]) for $f(S) = (S - K)^+$ and using that S is a martingale, we have

$$\mathbb{E}\left(f(S_t) - f(S_0)\right) \quad = \quad \mathbb{E}\left(\int_{-\infty}^{\infty} \Lambda_t(S)\delta_K(dS)\right) \quad = \quad \mathbb{E}\left(\Lambda_t(K)\right),$$

where $\Lambda_t(a)$ is the semimartingale local time for S_t at level a and $\delta_K(dS)$ denotes the Dirac measure concentrated at $S = K$ (see also equation (3.6.47) in [KS91]). On the other hand, for any bounded, continuous function on $\mathbb{R}_+$, g, we have

$$\mathbb{E}\left(\int_0^t g(S_u)d\langle S_u \rangle\right) \quad = \quad \mathbb{E}\left(\int_0^t g(S_u)S_u^2\sigma(X_u)^2 Y_u^2 du\right)$$

$$= \quad \int_0^t \mathbb{E}(g(e^{X_u})e^{2X_u}\sigma(X_u)^2\mathbb{E}(Y_u^2|X_u))du$$

$$= \quad \int_0^t \int_{-\infty}^{\infty} g(e^x)e^{2x}\sigma(x)^2\mathbb{E}(Y_u^2|X_u = x)\mathbb{P}(X_u \in dx)\,du$$

$$= \quad \int_{-\infty}^{\infty} g(e^x)e^{2x}\sigma(x)^2 \int_0^t \mathbb{E}(Y_u^2|X_u = x)\,\hat{p}_u(x_0, y_0; x)\,du\,dx, \qquad (20)$$

where $\mathbb{P}(X_u \in dx | X_0 = x_0, Y_0 = y_0) = \hat{p}_u(x_0, y_0; x)dx$. But by Theorem 3.7.1, part (iii) in Karatzas & Shreve [KS91], we have

$$\mathbb{E}\left(\int_0^t g(S_u)d\langle S_u \rangle\right) \quad = \quad 2\,\mathbb{E}\left(\int_0^{\infty} g(S)\Lambda_t(S)dS\right) \quad = \quad 2\,\mathbb{E}\left(\int_{-\infty}^{\infty} g(e^x)\Lambda_t(e^x)e^x dx\right). \qquad (21)$$

By the arbitrariness of g, comparing (20) and (21) we see that

$$2e^x\mathbb{E}\left(\Lambda_t(e^x)\right) \quad = \quad e^{2x}\sigma(x)^2\int_0^t \mathbb{E}(Y_u^2|X_u = x)\hat{p}_u(x_0, y_0; x)du, \qquad \forall x \in \mathbb{R}.$$

---

[8] the result is easily adapted in the general case $S_0 \neq 1$ by considering $X_t = \log\frac{S_t}{S_0}$.

In particular

$$\mathbb{E}(f(S_t) - f(S_0)) \;=\; \mathbb{E}(\Lambda_t(K)) \;=\; \mathbb{E}(\Lambda_t(e^{x_1})) \;=\; \frac{1}{2}K\sigma(x_1)^2 \int_0^t \mathbb{E}(Y_u^2|X_u = x_1)\hat{p}_u(x_0,y_0;x_1)du \,. \qquad (22)$$

Let $1 < a < \infty$ with $a > y_1^*$. Applying the Bellaiche heat kernel expansion on the compact interval $[\frac{1}{a}, a]$, we know that for any fixed $\varepsilon \in (0,1)$, there exists a $t^* = t^*(\varepsilon)$ such that for all $t < t^*$ we have

$$\begin{aligned}
\mathbb{E}(Y_t^2|X_t = x_1)\hat{p}_t(x_0,y_0;x_1) \;&=\; \int_{-\infty}^{\infty} y_1^2 \hat{p}_t(x_0,y_0;x_1,y_1)dy_1 \leq (1+\varepsilon)\int_{\frac{1}{a}}^{a} \frac{\psi(y_1)}{2\pi t}\, e^{-\phi(y_1)/t}dy_1 \;+\; I_0 \;+\; I_\infty, \\
&=\; (1+\varepsilon)\frac{\psi(y_1^*)}{\sqrt{2\pi t\,\phi''(y_1^*)}}\, e^{-\phi(y_1^*)/t}\left[1 + O(t^{\frac{1}{2}})\right] \;+\; I_0 \;+\; I_\infty \,, \qquad (23)
\end{aligned}$$

where $I_0 = \int_0^{\frac{1}{a}} y_1^2 \hat{p}_t(x_0,y_0;x_1,y_1)dy_1$, $I_\infty = \int_a^{\infty} y_1^2 \hat{p}_t(x_0,y_0;x_1,y_1)dy_1$, y_1^* is defined in the statement of the theorem and we have used Laplace's method around the minimizer at $y_1 = y_1^*$ (see Proposition 2.1, page 323 in Stein & Sharkarchi [SS03]), which we are allowed to do because the distance function d, the metric (g_{ij}) and $u_0(\mathbf{x},\mathbf{y})$ are all smooth away from the cut locus of x (and the cut locus is empty because $\kappa \leq 0$), so ψ and ϕ are smooth. Similarly we obtain the lower bound

$$\mathbb{E}(Y_t^2|X_t = x_1)\hat{p}_t(x_0,y_0;x_1) \;\geq\; (1-\varepsilon)\frac{\psi(y_1^*)}{\sqrt{2\pi t\,\phi''(y_1^*)}}\, e^{-\phi(y_1^*)/t}\left[1 + O(t^{\frac{1}{2}})\right] \,.$$

Fix $\delta > 0$. Then by Corollary 4.2 , Corollary 4.4 and Lemma 4.5, we know that for $a = a(\varepsilon)$ sufficiently large we have

$$\begin{aligned}
I_0 \;&\leq\; \int_0^{\frac{1}{a}} y_1^2 \cdot \frac{\chi(x_0,y_0)}{\chi(x_1,y_1)} e^{V_{\max}t} \sqrt{|g(\mathbf{y})|}\, \frac{c_\delta}{\pi t} e^{-d(\mathbf{x},\mathbf{y})^2/(2+\delta)t}dy_1 \\
&\leq\; \frac{c_\delta}{\pi t} e^{V_{\max}t} \int_0^{\frac{1}{a}} y_1^2 \sqrt{|g(\mathbf{y})|}\, C_0\, e^{-d_0(y_1)^2/C_0'(2+\delta)t}dy_1 \,,
\end{aligned}$$

for some positive constants $C_0 = C_0(x_0,y_0,x_1)$ and $C_0' = C_0'(x_0,y_0,x_1)$, where $\sqrt{|g(\mathbf{y})|} \equiv \sqrt{|g(x_1,y_1)|}$. But from the assumptions on $\alpha(y)$ we also know that $\sqrt{|g(\mathbf{y})|} \sim \frac{1}{A_1\sigma(x_1)y_1^2}$ as $y_1 \to 0$. Setting $\varphi_{\varepsilon,\delta} = C_0'(2+\delta)(1+\varepsilon)$ and $c_{\varepsilon,\delta,t} = \frac{C_0}{A_1\sigma(x_1)}c_\delta e^{V_{\max}t}(1+\varepsilon)/\pi$, then we obtain

$$\begin{aligned}
I_0 \;&\leq\; \frac{c_{\varepsilon,\delta,t}}{t} \int_0^{\frac{1}{a}} e^{-d_0(y_1)^2/(\varphi_{\varepsilon,\delta}t)}dy_1 \\
&\leq\; \frac{c_{\varepsilon,\delta,t}}{t} \int_0^{\frac{1}{a}} \frac{1}{y_1} e^{-[\frac{1}{A_1}\log y_1]^2/(\varphi_{\varepsilon,\delta}t)}dy_1 \\
&=\; \frac{c_{\varepsilon,\delta,t}}{t} \int_{-\infty}^{\log\frac{1}{a}} e^{-\frac{w^2}{A_1^2\varphi_{\varepsilon,\delta}t}}dw \\
&=\; \frac{c_{\varepsilon,\delta,t}}{t} \nu\sqrt{2\pi} \int_{-\infty}^{\log\frac{1}{a}} \frac{1}{\nu\sqrt{2\pi}} e^{-\frac{w^2}{2\nu^2}}dw \qquad \text{(where } \nu = A_1\varphi_{\varepsilon,\delta}^{\frac{1}{2}}\sqrt{t}/\sqrt{2}) \\
&=\; \frac{c_{\varepsilon,\delta,t}}{t}\nu\sqrt{2\pi}\,\Phi(z) \;\leq\; \frac{c_{\varepsilon,\delta,t}}{t}\nu\frac{e^{-z^2/2}}{|z|} \,,
\end{aligned}$$

where $\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx$, $z = \frac{\log\frac{1}{a}}{\nu}$, and we have used that $\frac{1}{y_1} \geq 1$ for $0 < y_1 < \frac{1}{a} < 1$ and the inequality $\Phi(z) \leq \frac{1}{|z|\sqrt{2\pi}}e^{-z^2/2}$ for $z < 0$ in the last line.

For I_∞, we only consider the case that $p \in (0,1)$; the case of $p = 1$ can be treated with an argument similar as the one for I_0. Again from Corollaries 4.2 and 4.4 and Lemma 4.5, we know that for $a = a(\varepsilon)$ sufficiently large we have

$$\begin{aligned}
I_\infty \;&\leq\; \int_a^{\infty} y_1^2 \cdot \frac{\chi(x_0,y_0)}{\chi(x_1,y_1)} e^{V_{\max}t} \sqrt{|g(\mathbf{y})|}\, \frac{c_\delta}{\pi t} e^{-d(\mathbf{x},\mathbf{y})^2/((2+\delta)t)}dy_1 \\
&\leq\; \frac{c_\delta}{\pi t} e^{V_{\max}t} \int_a^{\infty} y_1^2\, C_\infty y_1^{\frac{1}{2}(1-p)} \sqrt{|g(\mathbf{y})|}\, e^{-d_\infty(y_1)^2/(C_\infty'(2+\delta)t)}dy_1,
\end{aligned}$$

for some constants $C_\infty = C_\infty(x_0, y_0, x_1)$ and $C'_\infty = C'_\infty(x_0, y_0, x_1)$. But for $p \in (0,1)$, $d_\infty(y_1) \sim \frac{y_1^{1-p}}{B_1(1-p)}$ and $\sqrt{|g(\mathbf{y})|} \sim \frac{1}{\sigma(x_1)B_1 y_1^{1+p}}$ as $y_1 \to \infty$. Thus, we have

$$I_\infty \leq \frac{\bar{c}_{\varepsilon,\delta,t}}{t} \int_a^\infty y_1^{\frac{3}{2}(1-p)} e^{-\frac{y_1^{2(1-p)}}{B_1^2(1-p)^2}/(\bar{\varphi}_{\varepsilon,\delta}t)} dy_1$$

where $\bar{c}_{\varepsilon,\delta,t} = \frac{C_\infty}{B_1\sigma(x_1)} c_\delta e^{V_{\max}t}(1+\varepsilon)/\pi$ and $\bar{\varphi}_{\varepsilon,\delta} = C'_\infty(2+\delta)(1+\varepsilon)$. Making the change of variable $u = y_1^{1-p}$ in I_∞ so $dy_1 = \frac{1}{1-p} u^{\frac{p}{1-p}} du$, we see that for a sufficiently large, we have

$$
\begin{aligned}
I_\infty &\leq \frac{\bar{c}_{\varepsilon,\delta,t}}{t} \int_{a^{1-p}}^\infty u^{\frac{3}{2}} e^{-\frac{u^2}{B_1^2(1-p)^2 \bar{\varphi}_{\varepsilon,\delta}t}} \frac{u^{\frac{p}{1-p}}}{1-p} du \\
&= \frac{\bar{c}_{\varepsilon,\delta,t}}{(1-p)t} \int_{a^{1-p}}^\infty u^{\frac{3-p}{2(1-p)}} e^{-\frac{u^2}{B_1^2(1-p)^2 \bar{\varphi}_{\varepsilon,\delta}t}} du \\
&\leq \frac{\bar{c}_{\varepsilon,\delta,t}}{(1-p)t} \nu\sqrt{2\pi} \int_{a^{1-p}}^\infty \frac{1}{\nu\sqrt{2\pi}} e^{-\frac{u^2}{2\nu^2}} du \\
&\qquad \text{(for } a = a(\varepsilon) \text{ sufficiently large, where } \nu = B_1(1-p)\bar{\varphi}_{\varepsilon,\delta}^{\frac{1}{2}}\sqrt{t}) \\
&= \frac{\bar{c}_{\varepsilon,\delta,t}}{(1-p)t} \nu\sqrt{2\pi}\, \Phi^c\left(\frac{a^{1-p}}{\nu}\right) \\
&\leq \frac{\bar{c}_{\varepsilon,\delta,t}}{(1-p)t} \nu \frac{e^{-\frac{1}{2}z^2}}{z},
\end{aligned}
\tag{24}
$$

where $\Phi^c = 1 - \Phi$ and $z = \frac{a^{1-p}}{\nu} = \frac{a^{1-p}}{B_1(1-p)\bar{\varphi}_{\varepsilon,\delta}^{\frac{1}{2}}\sqrt{t}}$. Thus for a sufficiently large, z^2 will exceed $\phi(y_1^*)/t$ so I_0 and I_∞ are both higher order terms than the leading term $e^{-\phi(y_1^*)/t}/\sqrt{t}$ in (23), thus we can ignore them at leading order. A similar argument holds for the right tail integral I_∞ when $p = 1$. Thus we conclude that

$$\mathbb{E}(Y_t^2|X_t = x_1)\hat{p}_t(x_0, y_0; x_1) \sim \frac{\psi(y_1^*)}{\sqrt{2\pi t\, \phi''(y_1^*)}} e^{-\phi(y_1^*)/t}, \qquad (t \to 0). \tag{25}$$

We now have to estimate the integral in (22), using (25). Using the well known asymptotic relation

$$\frac{1}{2} \int_0^t \frac{1}{\sqrt{2\pi s}} e^{-k^2/2s} ds = \frac{1}{k^2\sqrt{2\pi}} t^{\frac{3}{2}} e^{-\frac{k^2}{2t}} [1 + O(t)] \qquad (t \to 0),$$

and comparing with (22) we see that for all $\varepsilon > 0$ there exists a $t^* = t^*(\varepsilon)$ such that for all $t \leq t^*$ we have

$$
\begin{aligned}
\mathbb{E}(S_t - K)^+ - (S_0 - K)^+ &\leq \frac{1}{2} K\sigma(x_1)^2 \int_0^t \frac{\psi(y_1^*)}{\sqrt{2\pi s\phi''(y_1^*)}} e^{-\phi(y_1^*)/s}(1+\varepsilon) ds \\
&= K\sigma(x_1)^2 \frac{\psi(y_1^*)}{\sqrt{\phi''(y_1^*)}} \frac{1}{2} \int_0^t \frac{1}{\sqrt{2\pi s}} e^{-\phi(y_1^*)/s}(1+\varepsilon) ds \\
&= K\sigma(x_1)^2 \frac{\psi(y_1^*)}{\sqrt{\phi''(y_1^*)}} \frac{1}{2\phi(y_1^*)\sqrt{2\pi}} t^{\frac{3}{2}} e^{-\frac{\phi(y_1^*)}{t}} [1 + O(t)](1+\varepsilon) \\
&\leq \frac{A^{\mathrm{SV}}(x_1)}{\sqrt{2\pi}} e^{-\phi(y_1^*)/t} t^{\frac{3}{2}} (1 + 2\varepsilon)
\end{aligned}
$$

(recall that $A^{\mathrm{SV}}(x_1) = \frac{K\sigma(x_1)^2\psi(y_1^*)}{\sqrt{\phi''(y_1^*)}} \frac{1}{2\phi(y_1^*)}$). We proceed similarly for the lower bound. ∎

4.4 Non-zero correlation

If $d\langle W^1, W^2\rangle_t = \rho dt$ for $\rho \neq 0$, we can still make the gauge transformation trick work, if $\mu(y)$ takes a certain functional form in terms of $\alpha(y)$ and $\sigma(x)$ is constant, as the following proposition demonstrates. However, we assume that $\rho \leq 0$ to ensure that the stock price process $S_t = e^{X_t}$ is a martingale (see e.g. [Jour04], [LM07] to see examples of where this fails for $\rho > 0$).

Proposition 4.7 *For $\rho \neq 0, \pm 1$, we can find a gauge transformation to remove the $\mathcal{A}$ term if $\sigma(x)$ is constant, and*

$$\mu(y) = \frac{\alpha(y)}{2y}[y\alpha'(y) - \alpha(y)].$$

Under this condition, and assuming μ and α also satisfy the other conditions in Assumption 3.1, the potential $V(x,y)$ induced by the gauge transformation is bounded from above, so (if $\rho \leq 0$) Theorem 4.6 still holds subject to minor modifications of the proof, using the following distance estimates for $\rho \neq 0$

$$d(\mathbf{x}, \mathbf{y}) \asymp d_0(y_1), \qquad (y_1 \to 0) \quad \text{where } d_0(y_1) := -\frac{1}{\bar\rho A_1} \log y_1,$$

$$d(\mathbf{x}, \mathbf{y}) \asymp d_\infty(y_1), \qquad (y_1 \to \infty) \quad \text{where } d_\infty(y_1) := \begin{cases} \frac{1}{\bar\rho B_1} \log y_1, & (p = 1), \\ \frac{y_1^{1-p}}{\bar\rho B_1(1-p)}, & (p \in (0,1)) \end{cases}$$

when $\sigma(x) \equiv 1$ and $\bar\rho = \sqrt{1-\rho^2}$.

Proof. See Appendix C. ∎

Remark 4.4 *Setting $\alpha(y) = \nu y$ for $\nu > 0$, we find that $\mu(y) = 0$ which is consistent with the SABR model (for $\beta = 1$), so the gauge transformation works for this case.*

4.5 Small-time behaviour for the Black-Scholes model

Let $C^{\mathrm{BS}}(S_0, K, t, \sigma)$ denote the price of a European call option at time zero under the Black-Scholes model with stock price S_0, strike $K = S_0 e^x$, time-to-maturity t, and volatility σ (with zero interest rates and dividends).

Proposition 4.8 *Let*

$$\hat\sigma_t = \sqrt{\sigma^2 + at}$$

for $t > 0$, and assume $t \in (0, \frac{\sigma^2}{|a|})$ if $a < 0$. Then $C^{\mathrm{BS}}(S, K, t, \hat\sigma_t)$ has the following asymptotic behaviour as $t \to 0$

$$C^{\mathrm{BS}}(S, K, t, \hat\sigma_t) = (S_0 - K)^+ + \frac{e^{-\frac{x^2}{2\sigma^2 t}}}{\sqrt{2\pi}} t^{\frac{3}{2}} e^{\frac{1}{2}\frac{ax^2}{\sigma^4}} A_{\mathrm{BS}}(x, \sigma)[1 + O(t)] \qquad (x \neq 0), \tag{26}$$

where $K = S_0 e^x$ and $A_{\mathrm{BS}}(x,\sigma) = S_0 e^{\frac{1}{2}x}\frac{\sigma^3}{x^2}$.

Proof. This is just the same as Proposition 3.4 in [FJL12]. ∎

Remark 4.5 *Note that we are not considering a time-dependent Black-Scholes model here per se, but rather a standard Black-Scholes model but with a volatility parameter which depends on t.*

5 Small-time behaviour of implied volatility

In this section we derive the corresponding asymptotic expansions for implied volatility.

Theorem 5.1 *For the model defined above, let $\hat\sigma_t(x_1)$ denote the implied volatility at time zero and maturity t for strike $K = e^{x_1}$, $K \neq S_0$. Then we have the following asymptotic behaviour for $\hat\sigma_t(x_1)$:*

$$\hat\sigma_t^2(x_1) = \hat\sigma^2(x_1) + a(x_1)t + o(t), \tag{27}$$

where

$$\hat\sigma(x_1) = \frac{|x_1 - x_0|}{d(x_0, y_0; x_1, y_1^*(x_1))}, \qquad a(x_1) = \frac{2\hat\sigma^4(x_1)}{x^2}\log\frac{A^{\mathrm{SV}}(x_1)}{A_{\mathrm{BS}}(x, \hat\sigma(x_1))}, \tag{28}$$

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

Proof. See Appendix A. ∎

Remark 5.1 Note that we have not said anything about uniform convergence in x in (27). Typically for these type of problems, convergence is uniform on compact sets away from zero, but the convergence is not uniform on intervals of the form $(0, a)$ or $(-a, 0)$ for $a > 0$ (but we do not need such a result in this article, so we defer the details for future work).

Remark 5.2 From (27) we see that

$$\hat{\sigma}_t(x_1) = \hat{\sigma}(x_1) + \frac{1}{2}\frac{a(x_1)}{\hat{\sigma}(x_1)}t + o(t). \tag{29}$$

6 Comparison against the asymptotic expansion in Lorig, Pagliarani & Pascucci

[LPP15] consider a general local-stochastic volatility model for a log stock price process X_t of the form

$$\left.\begin{aligned}
dX_t &= -\frac{1}{2}\sigma^2(t, X_t, Y_t)dt + \sigma(t, X_t, Y_t)dW_t, & X_0 &= x \in \mathbb{R}, \\
dY_t &= f(t, X_t, Y_t)dt + \beta(t, X_t, Y_t)dB_t, & Y_0 &= y \in \mathbb{R}, \\
d\langle W, B\rangle_t &= \rho(t, X_t, Y_t)\, dt, & |\rho| &< 1.
\end{aligned}\right\} \tag{30}$$

By expanding the coefficients of the infinitesimal generator of (X, Y) in a Taylor series about an arbitrary point $(\bar{x}, \bar{y})$, the authors obtain an explicit expansion for the price of a European call option and its associated implied volatility. Under suitable conditions, the price of a European option $u(t, x, y) := \mathbb{E}_{t,x,y}(\varphi(X_T))$ [9] satisfies the backward Kolmogorov equation

$$(\partial_t + \mathcal{A}(t))u = 0, \quad u(T, x, y) = \varphi(x), \tag{31}$$

where $\mathcal{A}(t)$ is the infinitesimal generator associated with the two-dimensional process (X, Y). We now briefly explain how the [LPP15] methodology works, in the one-dimensional case and for a general local-stochastic volatility model (see also [LPP14] and [PP14]).

6.1 The one-dimensional case

We first consider the one-dimensional heat equation

$$\left[\partial_t + a(x)\partial_x^2\right]u = 0.$$

If we now formally expand $a(x)$ around zero: $a(x) = a(0) + a'(0)x + \frac{1}{2}a''(0)x^2 + \ldots$ and set $u = u_0 + u_1 + u_2 + \ldots$, we obtain the following nested system of Cauchy problems:

$$\begin{aligned}
\left(\partial_t + a(0)\partial_x^2\right)u_0 &= 0, & u_0(T, x) &= \varphi(x), \\
\left(\partial_t + a(0)\partial_x^2\right)u_1 &= -a'(0)x\partial_x^2 u_0, & u_1(T, x) &= 0, \\
\left(\partial_t + a(0)\partial_x^2\right)u_2 &= -a'(0)x\partial_x^2 u_1 - \frac{1}{2}a''(0)x^2\partial_x^2 u_0, & u_2(T, x) &= 0,
\end{aligned}$$

and so on. In general each of these equations can be solved recursively using Duhamel's principle, and [LPP15] give an explicit formula for u_n. It is also helpful to consider an artificial parameter $\varepsilon \in (0, 1]$. We then set $a^\varepsilon(x) = a(0) + \varepsilon a'(0)x + \frac{1}{2}\varepsilon^2 a''(0)x^2$, $u^\varepsilon = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \ldots$, and to obtain the family of equations above, we collect terms of like order in ε, and then finally set $\varepsilon = 1$.

---

[9]In this section, $\mathbb{E}_{t,x,y}(\cdot)$ is shorthand for $\mathbb{E}(\cdot\,|X_t = x, Y_t = y)$, for all $(t, x, y) \in [0, T] \times \mathbb{R} \times \mathbb{R}$.

6.2 The general case

Now consider the generator $\mathcal{A}(t)$ associated with a general local-stochastic volatility model of the form in (30):

$$\mathcal{A}(t) = a(t,x,y)(\partial_x^2 - \partial_x) + f(t,x,y)\partial_y + b(t,x,y)\partial_y^2 + c(t,x,y)\partial_x\partial_y,$$

where the functions a, b, c are defined as

$$a(t,x,y) := \frac{1}{2}\sigma^2(t,x,y), \qquad b(t,x,y) := \frac{1}{2}\beta^2(t,x,y), \qquad c(t,x,y) := \rho(t,x,y)\sigma(t,x,y)\beta(t,x,y).$$

Expanding each function $\{a,b,c,f\}$ as a Taylor series about a fixed point $(\bar{x},\bar{y}) \in \mathbb{R}^2$:

$$\chi(t,x,y) := \sum_{n=0}^{\infty}\chi_n(t,x,y), \qquad\qquad\qquad \chi = \{a,b,c,f\},$$

$$\chi_n(t,x,y) := \sum_{k=0}^{n}\chi_{n-k,k}(t)\cdot(x-\bar{x})^{n-k}(y-\bar{y})^k, \qquad \chi_{n-k,k}(t) := \frac{1}{(n-k)!k!}\partial_x^{n-k}\partial_y^k\chi(t,\bar{x},\bar{y}),$$

the generator $\mathcal{A}(t)$ can now be written formally as

$$\mathcal{A}(t) = \sum_{n=0}^{\infty}\mathcal{A}_n(t), \qquad \mathcal{A}_n(t) = a_n(t,x,y)(\partial_x^2 - \partial_x) + f_n(t,x,y)\partial_y + b_n(t,x,y)\partial_y^2 + c_n(t,x,y)\partial_x\partial_y.$$

We now expand the unknown function u as a series $u = \sum_{n=0}^{\infty} u_n$. Inserting this expansion, as well as the expansion for $\mathcal{A}(t)$ into the Kolmogorov backward equation we again obtain a nested sequence of Cauchy problems

$$(\partial_t + \mathcal{A}_0(t))u_0 = 0, \qquad\qquad u_0(T,x,y) = \varphi(x), \qquad\qquad (32)$$

$$(\partial_t + \mathcal{A}_0(t))u_n = -\sum_{k=1}^{n}\mathcal{A}_k(t)u_{n-k}, \qquad u_n(T,x,y) = 0, \qquad n \geq 1. \qquad (33)$$

The sequence (u_n) can be solved explicitly, and a general expression for the nth term is given in Theorem 2.6 in [LPP15]. For European call/put options, the expansion lends itself to an explicit implied volatility expansion (see Section 3 of [LPP15]), and the number of terms in the price and implied volatility expansion grows like $n!$. As such, for practical purposes, one can only compute terms up to order $n = 4$. The advantage of the above method is that the nth order approximation (for both price and implied volatility) can be applied to any diffusion whose coefficients are C^n in the spatial variables. However, to prove the accuracy of the pricing approximation more stringent conditions (discussed below) must be enforced.

6.3 Asymptotic error estimates

In [LPP15], the authors prove the following bound for the error introduced by replacing the exact European option price $u(t,x,y)$ with the N-th order approximation $\bar{u}_N(t,x,y) = \sum_{i=0}^{N} u_i(t,x,y)$.

Corollary 6.1 *Consider a European option with payoff function φ. Fix $\overline{T} > 0$ and $(\bar{x},\bar{y}) = (x,y)$. Suppose the coefficients $a(t,\cdot,\cdot)$, $b(t,\cdot,\cdot)$, $c(t,\cdot,\cdot)$ and $f(t,\cdot,\cdot)$ and all of their partial derivatives up to order $N+1$ are bounded by a positive constant $M < \infty$, uniformly with respect to $t \in [0,\overline{T}]$ and that*

$$\frac{1}{M}(\xi^2 + \eta^2) \leq a(t,x,y)\xi^2 + c(t,x,y)\xi\eta + b(t,x,y)\eta^2 \leq M(\xi^2 + \eta^2), \quad \forall t \in [0,\overline{T}],\ x,y,\xi,\eta \in \mathbb{R}.$$

Then for any $0 < t < T \leq \overline{T}, \varepsilon > 0$ there exists a constant C such that

$$|u(t,x,y) - \bar{u}_N(t,x,y)| \leq C(T-t)^{\frac{N+1}{2}}\int_{\mathbb{R}^2}\Gamma^{M+\varepsilon}(t,x,y;T,x',y')\varphi(x')dx'dy', \qquad 0 \leq t < T,\ (x,y) \in \mathbb{R}^2, \qquad (34)$$

where $\Gamma^{M+\varepsilon}(t,x;T,y)$ is the fundamental solution of the two-dimensional heat operator

$$H^{M+\varepsilon} = (M+\varepsilon)(\partial_x^2 + \partial_y^2) + \partial_t, \qquad\qquad (35)$$

and the constant C depends only on M, N, $\overline{T}$ and ε.

17

Proof. See Corollary 4.6 of [LPP15]. ∎

For the case of a European call option with log-moneyness $x_1 > 0$, we can re-write the error bound as

$$
\begin{aligned}
|u(t,x,y) - \bar{u}_N(t,x,y)| &\leq Ct^{\frac{N+1}{2}}\mathbb{E}(e^{\sigma W_t} - e^{x_1})^+ \\
&\sim Ct^{\frac{N+1}{2}}e^{-\frac{x_1^2}{2\sigma^2 t}}\frac{\sigma^3 e^{x_1}}{\sqrt{2\pi}\,x_1^2}t^{\frac{3}{2}} \qquad (t \to 0),
\end{aligned}
$$

where $\sigma^2 = M + \varepsilon$ and W_t is a standard one-dimensional Brownian motion starting at zero. But the error in our call option expansion is of the order

$$
e^{-\phi(y_1^*)/t}\,o(t^{\frac{3}{2}}) \qquad (t \to 0)\,.
$$

Thus we see that the [LPP15] result only gives a tighter error bound than our asymptotic call option estimate in Theorem 4.6 if

$$
N \geq O\left(\frac{1}{t\log\frac{1}{t}}\right)\left[\phi(y_1^*) - \frac{x_1^2}{2\sigma^2}\right] \qquad (t \to 0),
$$

which will require computing $O\left(\left(\frac{1}{t\log\frac{1}{t}}\right)!\right)$ terms in the [LPP15] expansion (recall that $\frac{1}{t\log\frac{1}{t}} \to \infty$ as $t \to 0$, but grows slower than $\frac{1}{t}$) (also note that $\phi(y_1^*) - \frac{x_1^2}{2\sigma^2}$ is positive because σ is an upper bound of the diffusion matrix via the ellipticity condition in Assumption 4.1i in [LPP15]).

7 Numerical Example: the SABR model

Consider the well known SABR model for $\beta = 1$ with unit vol-of-vol:

$$
\begin{cases} dX_t = -\frac{1}{2}Y_t^2 dt + Y_t dW_t^1\,, \\ dY_t = Y_t dW_t^2\,, \end{cases} \tag{36}
$$

with initial value $(X_0, Y_0) = (x_0, y_0) \in \mathbb{R} \times \mathbb{R}_+$ and independent standard Brownian motions W^1 and W^2. The metric associated with this model is the hyperbolic metric $ds^2 = \frac{1}{y^2}(dx^2 + dy^2)$ on the upper half plane $\mathbb{H}^2$, and $g_{11}(x,y) = g_{22}(x,y) = \frac{1}{y^2}$ and $g_{12}(x,y) = g_{21}(x,y) = 0$, so $\sqrt{|g(x,y)|} = \frac{1}{y^2}$. For the hyperbolic metric, it is known (see e.g. page 170 in [HL08] and Paulot [Pau10]) that

$$
\begin{aligned}
d(x_0, y_0; x_1, y_1) &= \cosh^{-1}[1 + \frac{(x_1-x_0)^2 + (y_1-y_0)^2}{2y_0 y_1}]\,, \\
y_1^* &= \sqrt{(x_1-x_0)^2 + y_0^2}\,, \\
d := d(x_0, y_0; x_1, y_1^*) &= \cosh^{-1}[\frac{\sqrt{(x_1-x_0)^2 + y_0^2}}{y_0}]\,, \\
E_{yy}(x_0, y_0; x_1, y_1^*) &= \phi''(y_1^*) = \frac{d(x_0, y_0; x_1, y_1^*)}{y_0 y_1^* \sinh d(x_0, y_0; x_1, y_1^*)}\,.
\end{aligned}
$$

We also have

$$
\Delta = \frac{1}{y^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\,, \quad \mathcal{A} = -\frac{1}{2}y^2\partial_x\,, \quad u_0(x_0, y_0; x_1, y_1) = \left(\frac{\sinh d(x_0, y_0; x_1, y_1)}{d(x_0, y_0; x_1, y_1)}\right)^{-\frac{1}{2}}\,,
$$

$$
A(x_0, y_0; x_1, y_1) = \int_0^1 \langle \mathcal{A}, \dot{\gamma} \rangle\, dt = \int_{x_0}^{x_1} \frac{1}{y^2} \cdot -\frac{1}{2}y^2\, dx = -\frac{1}{2}(x_1 - x_0)\,.
$$

Without loss of generality, we can set $x_0 = 0$, and we obtain

$$
A^{\text{SV}}(x_1) = \frac{K\psi(y_1^*)}{\sqrt{\phi''(y_1^*)}}\frac{1}{2\phi(y_1^*)} = \frac{Ke^{-\frac{1}{2}(x_1-x_0)}\left(\frac{\sinh d}{d}\right)^{-\frac{1}{2}}}{\sqrt{d/(y_0 y_1^* \sinh d)}}\frac{1}{d^2}\,.
$$
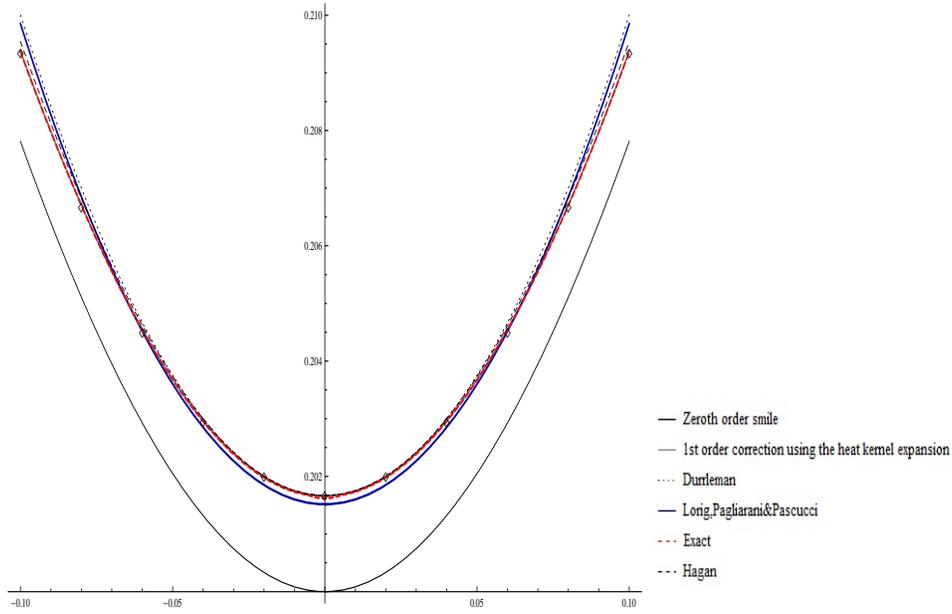
Figure 1: Here we consider the uncorrelated SABR model $dS_t = S_tY_t dW_t^1, dY_t = \alpha Y_t dW_t^2$ with $S_0 = 1$, $y_0 = .2$, $\alpha = 1$ and maturity $t = .1$ - we have plotted the leading order smile (the lower curve, solid grey thin line) and the corrected implied volatility smile (light grey) using (27). We have also plotted the exact smile (computed via numerical integration using the formula in [AS12]) in red-dashed (which is barely distinguishable from our corrected smile in light grey), the Lorig, Pagliarani & Pascucci truncated expansion with $N = 3$ (dark blue), the smile using a fast Monte Carlo scheme (points with diamond-shaped markers, using 6 million simulations and 1000 time steps), the smile using the well known Hagan et al. [HKLW02] formula (black dashed, which lies above the red-dashed curve and below the LPP smile) and the Durrleman [Dur04] approximation (in black dots) (cf. Theorem 3.1.1 in [Dur04], which also agrees with Theorem 4.1 in [FJ11])(Mathematica code available from MF on request)
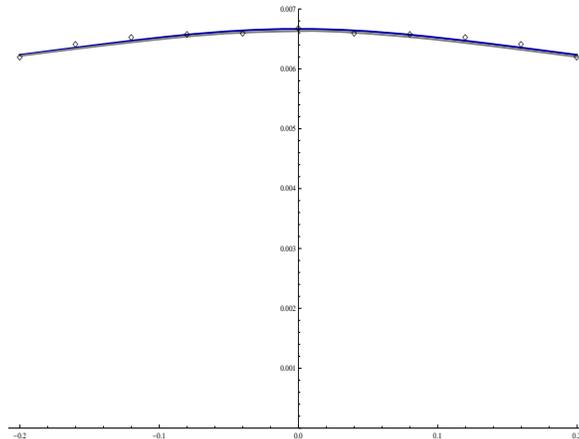
Figure 2: Here we have plotted the our correction term $a(x)$ (blue) against the $a(x)$ implied by [AS12] and the $a(x)$ implied by the Monte Carlo call prices, i.e. $(\sigma_{MC}^2(x) - \sigma_0(x)^2)/t$ (diamond-shaped markers). Our formula for $a(x)$ is also in exact (i.e. to machine precision) numerical agreement with the expansion given in section 4.3 in Paulot [Pau10], and in equation (6.10) in Busca et al. [BBF04].

In the first table below, we show the leading order smile, the first order corrected smile and the implied volatility from Monte Carlo σ_{MC}, as a function of the log-moneyness. In the second table, we show $a(x)$ against the $a(x)$ implied by the Monte Carlo implied volatility given by $(\sigma_{MC}^2(x) - \sigma_0(x)^2)/t$ and also the $a(x)$ implied by the formula in [AS12].

x	Zeroth order smile	First order correction	Monte Carlo	Relative Error
0.0001	0.2	0.201668	0.201664	-0.00212726%
0.04	0.201319	0.202961	0.20295	-0.00542944%
0.08	0.20511	0.206705	0.206709	0.00220052%
0.12	0.210961	0.212489	0.212503	0.00685656%
0.16	0.21838	0.21983	0.219846	0.00760195%
0.2	0.226919	0.228288	0.22828	-0.00324944%

x	$a(x)$	$[\sigma_{MC}^2(x) - \sigma_0(x)^2]/t$	$a(x)$ Antonov & Spectre	Relative error of $a(x)$ to MC
0.0001	0.00670034	0.00668307	0.00662428	-0.257743 %
0.04	0.00664065	0.00659592	0.00659901	-0.673579%
0.08	0.00656944	0.00658825	0.00653502	0.286241%
0.12	0.00646856	0.00653048	0.00643458	0.957229%
0.16	0.00635304	0.00642651	0.00632072	1.15654%
0.20	0.00623259	0.00619872	0.00620046	-0.54341%

8 Small-time asymptotics for stochastic volatility model with a single jump-to-default

We now extend the model in (12) by incorporating a single (independent) Poisson jump-to-default with constant intensity $\lambda > 0$, and set $\rho = 0$, $\sigma(x) \equiv 1$ for simplicity, so $S_t = e^{X_t} 1_{\tau > t}$ where X satisfies

$$\begin{cases} dX_t = (\lambda - \frac{1}{2} Y_t^2) dt + Y_t dW_t^1 , \\ dY_t = \mu(Y_t) dt + \alpha(Y_t) dW_t^2 , \end{cases} \tag{37}$$

and $\tau \sim \text{Exp}(\lambda)$ is the default time for S (which is independent of W^1, W^2), with $X_0 = 0$, $Y_0 = y_0 \in \mathbb{R}_+$. Then the price of a out-of-the-money call option at time zero with strike $K > S_0$ is $e^{-\lambda t} \mathbb{E}(e^{X_t} - K)^+$. But $e^{-\lambda t} = 1 + O(t)$, so the effect of the actual default on the call option will not be seen at the order that we are interested in. However the effect of the compensator drift term λdt *will* be felt at leading order, and will increase $\mathcal{P}(\mathbf{x}, \mathbf{y})$ as defined in Section 4.3 (and thus the call option price at leading order) by the following factor:

$$e^{\lambda \int_0^1 \frac{1}{y(x(t))^2} \frac{dx}{dt} dt} \tag{38}$$

where $x(t)$ is the x coordinate of the shortest geodesic joining $(0, y_0)$ to the vertical line $\{x = x_1\}$. By a standard property of geodesics (see e.g. [doC92] and also equation (16) in [FJ11]), we know that the speed of the geodesic is conserved, i.e.

$$L = \frac{1}{y^2} \left(\frac{dx}{dt} \right)^2 + \frac{1}{\alpha(y)^2} \left(\frac{dy}{dt} \right)^2 = E \tag{39}$$

for some Energy constant $E > 0$, where L is the Lagrangian. But from the second Euler-Lagrange equation, we also know that

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) = \frac{d}{dt} \left(\frac{1}{y^2} \frac{dx}{dt} \right) = 0 . \tag{40}$$

Thus $K_1 = \frac{1}{y^2} \frac{dx}{dt}$ is also a conserved quantity, so we can re-write (39) as

$$y^2 K_1^2 + \frac{1}{\alpha(y)^2} \left(\frac{dy}{dt} \right)^2 = E . \tag{41}$$

We know from standard properties of geodesics that the shortest distance from $(0, y_0)$ to the line $\{x = x_1\}$ is $d(x_1) = \sqrt{E}$ and the shortest geodesic from $(0, y_0)$ to $\{x = x_1\}$ is perpendicular to the y-axis at $(x_1, y_1^*(x_1))$ so $\frac{dy}{dt} = 0$ at this point, hence from (41) we have that $(y_1^*(x_1))^2 = E/K_1^2$. Thus

$$K_1 = \frac{\sqrt{E}}{y_1^*(x_1)} = \frac{d(x_1)}{y_1^*(x_1)} , \tag{42}$$

so (38) simplifies to $e^{\lambda d(x_1)/y_1^*(x_1)}$, and from (28) we find that the modified (i.e. jump-adjusted) correction term is given by

$$a^J(x) = a(x) + \frac{2\lambda \hat{\sigma}^4(x)}{x^2} \frac{d(x)}{y_1^*(x)} .$$

Then, using (29), we see that the presence of the jump-to-default increases the implied volatility by the amount:

$$\Delta \hat{\sigma}^J(x) := \lambda \frac{\hat{\sigma}^3(x)}{x^2} \frac{d(x)}{y_1^*(x)} t + o(t) = \lambda \frac{\hat{\sigma}^2(x)}{|x| \, y_1^*(x)} t + o(t) .$$

For an out-of-the-money put option with strike $K < S_0$, the small-time behaviour is qualitatively completely different, and we see that

$$\mathbb{E}(K - S_t)^+ = K(1 - e^{-\lambda t}) + e^{-\lambda t} const. \times e^{-\phi(y_1^*)/t} t^{\frac{3}{2}} [1 + o(1)] \qquad (t \to 0)$$

$$= K \sum_{n=1}^{\infty} \frac{(-1)^{n+1} (\lambda t)^n}{n!} + o(t^q) \qquad (t \to 0) \tag{43}$$

for any $q > 0$. From these observations, we note the following:

- For $x > 0$, the new correction term $a^J(x)$ tends to infinity as $x \to 0$, which means that $\lim_{x \searrow 0} \frac{\partial}{\partial t} \hat{\sigma}_t^2(x)|_{t=0} = \infty$; this may not equal $\frac{\partial}{\partial t} \hat{\sigma}_t(0)$ (which is not computed in this article), but for a general exponential Lévy model with a non-zero Brownian component, equation (1.14) in Figueroa-López et al. [FGH14] show that

$$\hat{\sigma}_t(0) \quad = \quad \sigma_0 \; + \; const. \times t^{1 - \frac{Y}{2}} + o(t^{1 - \frac{Y}{2}}), \tag{44}$$

where $Y \in (0, 2)$ measures the degree of jump activity, and (44) implies that $\frac{\partial}{\partial t} \hat{\sigma}_t(0)|_{t=0} = \infty$.

- The $K < S_0$ case is not so mathematically interesting for us; we see that the small-time put price expansion in (43) is essentially a Taylor series in powers t (none of whose coefficients are affected by the stochastic volatility). In this case, it can be shown that the implied volatility tends to ∞ as $t \to 0$ (see e.g. [FFL12] et al. for details), similar to a pure exponential Lévy model; we refer the reader to [FGH12] et al. for more results in this direction.

References

[AP07] Andersen, L.B.G., V.V. Piterbarg, "Moment explosions in stochastic volatility Models", *Finance and Stochastics*, 11(1), 29-50, 2007.

[AS12] Antonov, A. and M. Spector, "Advanced analytics for the SABR model", `http://ssrn.com/abstract=2026350`, 2012.

[Bel81] Bellaiche, C., "Comportement asymptotique de $p_t(\mathbf{x}, \mathbf{y})$", *Astérique*, 84-85, 151-187, 1981.

[BA88] Ben Arous, G., "Methods de Laplace et de la phase stationnaire sur le space de Wiener", *Stochastics*, 25: 125-153, 1988.

[BBF02] Berestycki, H., J. Busca and I. Florent, "Asymptotics and calibration of local volatility models", *Quantitative Finance*, 2, 61-69, 2002.

[BS13] Brunick, G. and S. Shreve, "Matching an Itô process by a solution of a stochastic differential equation", *Annals of Applied Probability*, 23, 1584-1628, 2013.

[BBF04] Berestycki, H., J. Busca and I. Florent, "Computing the implied volatility in stochastic volatility models" *Communications on Pure and Applied Mathematics*, 57(10), 1352-1373, 2004.

[Chav84] Chavel, I., "Eigenvalues in Riemannian Geometry", Pure and Applied Mathematics, vol. 115, Academic Press Inc., Orlando, FL, 1984.

[Dav88] Davies, E.B., "Gaussian upper bounds for the heat kernels of some second-order operators on Riemannian manifolds", *Journal of Functional Analysis*, 80, 16-32, 1988.

[DFJV11] Deuschel, J.D., P.K. Friz, A. Jacquier, S. Violante, "Marginal density expansions for diffusions and stochastic volatility, Part II: Theoretical foundations ", *Communications on Pure and Applied Mathematics*, 67(2): 321-350, 2014.

[DFJV11b] Deuschel, J.D., P.K. Friz, A. Jacquier, S. Violante, "Marginal density expansions for diffusions and stochastic volatility, Part I: Applications", *Communications on Pure and Applied Mathematics*, 67(1): 40-82, 2014.

[dW65] DeWitt, B.S. "Dynamical Theory of Groups and Fields", Gordon and Breach, 1965.

[doC92] do Carmo, M., "Riemannian Geometry", Birkhäuser, 1992.

[Don11] Donaldson, S., "Riemann Surfaces", Oxford University Press, 2011.

[Dur04] Durrleman, V., "From Implied to Spot Volatilities", PhD thesis, Princeton University, 2004.

[FFL12] Figueroa-López, J.E. and M. Forde, "The small-maturity smile for exponential Lévy models", *SIAM Journal on Financial Mathematics*, 3, 33-65, 2012.

[FGH14] Figueroa-López, J.E., R. Gong and C. Houdré, "High-order short-time expansions for ATM option prices of exponential Lévy models", *Mathematical Finance*, 26(3), 516-557, 2016.

[FGH12] Figueroa-López, J.E., R. Gong and C. Houdré, "Small-time expansions of the distributions, densities, and option prices of stochastic volatility models with Lévy jumps", *Stochastic Processes and their Applications*, 122, 1808-1839, 2012.

[FJ11] Forde, M. and A. Jacquier "Small-time asymptotics for implied volatility under a general Local-Stochastic volatility model", *Applied Mathematical Finance*, 18, 517-535, 2011.

[FJL12] Forde, M., A. Jacquier and R. Lee, "The small-time smile and term structure of implied volatility under the Heston model", *SIAM Journal on Financial Mathematics*, 3, 690-708, 2012.

[FdM13] Friz, P. and S. DeMarco, "Varadhan's formula, conditioned diffusions, and local volatilities", `http://arxiv.org/pdf/1311.1545.pdf`, 2013.

[GHLOW12] Gatheral, G., E. Hsu, E.P. Laurence, C. Ouyang and T.-H. Wang, "Asymptotics of implied volatility in local volatility models", (2012), *Mathematical Finance*, 22(4), 591-620, 2012.

[Gyö86] Gyöngy, I., "Mimicking the one-dimensional marginal distributions of processes having an Itô differential", *Probability Theory and Related Fields*, 71(4), 501-516, 1986.

[HKLW02] Hagan, P., D. Kumar, A. S. Lesniewski and D.E. Woodward, "Managing smile risk", *Wilmott Magazine*, 2002.

[HL08] Henry-Labordère, P., "Analysis, Geometry, and Modeling in Finance: Advanced Methods in Option Pricing", Chapman & Hall, 2008.

[Hsu] Hsu, E.P., "A brief introduction to Brownian motion on a Riemannian manifold", unpublished lecture notes.

[Hsu02] Hsu, E.P., "Stochastic Analysis on Manifolds", Graduate Studies in Mathematics, vol. 38, American Mathematical Society, Providence, RI, 2002.

[Jost09] Jost, J., "Riemannian Geometry and Geometric Analysis", Fifth edition, Springer, 2008.

[Jour04] Jourdain, B, "Loss of martingality in asset price models with lognormal stochastic volatility", *preprint Cermics*, 267, 2004.

[KS91] Karatzas, I. and S. Shreve, "Brownian Motion and Stochastic Calculus", Springer-Verlag, 1991.

[Laur08] Laurence, P., "Implied volatility, fundamental solutions, asymptotic analysis and symmetry methods", Caltech, April 2008.

[Laur10] Laurence, P., "Asymptotics for local volatility and SABR models", Global Derivatives, Paris, France, 2010.

[Lew07] Lewis, A., "Geometries and smile asymptotics for a class of Stochastic Volatility models", `www.optioncity.net`, 2007.

[LM07] Lions, P.-L. and M. Musiela, "Correlations and bounds for stochastic volatility models", *Annales de l'Institut Henri Poincare (C) Non Linear Analysis"*, 24(1), 1-16, 2007.

[LPP15] Lorig, M., S. Pagliarani and A. Pascucci, "Explicit implied volatilities for multifactor local-stochastic volatility models", to appear in *Mathematical Finance*, 2015.

[LPP14] Lorig, M., S. Pagliarani and A. Pascucci, "Analytic expansions for parabolic equations", *SIAM Journal on Applied Mathematics*, 75(2), 468-491, 2014.

[MA73] Misner, C.W. and J. Archibald, "Gravitation", W. H. Freeman and Company, 1973.

[McK70] McKean, H.P. , "An upper bound to the spectrum of Δ on a manifold of negative curvature", *Journal of Differential Geometry*, 4, 359-366, 1970 .

[MO91] McAvity, D.M. and H. Osborn, "A DeWitt expansion of the heat kernel for manifolds with a boundary", *Classical Quantum Gravity*, 8, 603-638, 1991.

[MP49] Minakshisundaram, S. and A. Pleijel, "Some properties of the eigenfunctions of the Laplace operator on Riemannian manifolds", *Canadian Journal of Mathematics*, 1, 242-256, 1949.

[Mol75] Molchanov, S., "Diffusion processes and Riemannian geometry", *Russian Mathematics Surveys* 30:1, 1-63, 1975.

[MY05] Matsumoto, H. and Yor, M., "Exponential functionals of Brownian motion II: Some related diffusion processes", *Probability Surveys*, 2, 348-384, 2005.

[Neel07] Neel, R., "The small-time asymptotics of the heat kernel at the cut locus", *Communications in Analysis and Geometry* 15(4), 845-890, 2007.

[NS91] Norris, J. and D.W. Stroock , "Estimates on the fundamental solution to heat flows with uniformly elliptic coefficients", *Proceedings London Mathematical Society*, 62, 375-402, 1991

[Olv74] Olver, F.W., "Asymptotics and Special Functions", Academic Press, 1974.

[Pau10] Paulot, L., "Asymptotic implied volatility at the second order with application to the SABR model", working paper, 2010.

[PP14] Pagliarani, S. and A. Pascucci, "Asymptotic expansions for degenerate parabolic Equations", *Comptes Rendus Mathematique*, 352(12), 1011-1016, 2014.

[SS03] Stein, E.M. and R. Sharkarchi, "Complex Analysis", Princeton University Press, 2003.

[Vass03] Vassilevich, D.V., "Heat kernel expansion: user's manual", *Physics Reports* 388, 279-360, hep-th/0306138, 2003.

A Proof of Theorem 5.1

The formulae given for $\hat{\sigma}$ and a are chosen such that $\phi^{\mathrm{BS}}(x,\hat{\sigma}) = \phi(y^*)$ and $\hat{A}^{\mathrm{BS}}(x,\hat{\sigma},a) = A^{SV}(x_1)$, where $\phi^{\mathrm{BS}}(x,\sigma) := \frac{x^2}{2\sigma^2}$ and $\hat{A}^{\mathrm{BS}}(x,\sigma,a) := A_{\mathrm{BS}}(x,\sigma)e^{\frac{1}{2}\frac{ax^2}{\sigma^4}}$. Given $\delta > 0$, choose $\varepsilon < \frac{1}{4}\frac{x^2\delta}{\sigma^4}$. From Proposition 4.8 and Theorem 4.6, there exists a $t^* > 0$ such that for all $t < t^*$

$$
C^{\mathrm{BS}}(S_0, K, t, \sqrt{\hat{\sigma}^2 + t(a+\delta)}) - (S_0 - K)^+ \;\geq\; \frac{1}{\sqrt{2\pi}}\hat{A}^{\mathrm{BS}}(x,\hat{\sigma}, a+\delta)e^{-\phi^{\mathrm{BS}}(x,\hat{\sigma})/t}t^{\frac{3}{2}}e^{-\varepsilon},
$$

$$
\mathbb{E}(S_t - K)^+ - (S_0 - K)^+ \;\leq\; \frac{1}{\sqrt{2\pi}}A^{\mathrm{SV}}e^{-\phi(y^*)/t}t^{\frac{3}{2}}e^{\varepsilon} = \frac{1}{\sqrt{2\pi}}\hat{A}^{\mathrm{BS}}(x,\hat{\sigma},a)e^{-\phi^{\mathrm{BS}}(x,\hat{\sigma})/t}t^{\frac{3}{2}}e^{\varepsilon}.
$$

We now observe that ε has been chosen so that

$$
\frac{\hat{A}^{\mathrm{BS}}(x,\hat{\sigma}, a+\delta)}{\hat{A}^{\mathrm{BS}}(x,\hat{\sigma}, a)} \;=\; e^{\frac{1}{2}\frac{x^2\delta}{\hat{\sigma}^4}} \;>\; e^{2\varepsilon}. \tag{A-1}
$$

So

$$
\mathbb{E}(S_t - K)^+ - (S_0 - K)^+ < \frac{1}{\sqrt{2\pi}}\hat{A}^{\mathrm{BS}}(x,\hat{\sigma}, a+\delta)e^{-\phi^{\mathrm{BS}}(x,\hat{\sigma})/t}t^{\frac{3}{2}}e^{-\varepsilon} \leq C^{\mathrm{BS}}(S, K, t, \sqrt{\hat{\sigma}^2 + t(a+\delta)}) - (S_0 - K)^+.
$$

By the monotonicity of C^{BS} as a function of volatility, we deduce

$$
\hat{\sigma}_t^2(x_1) \;\leq\; \hat{\sigma}(x_1)^2 + a(x_1)t + \delta t. \tag{A-2}
$$

We proceed similarly to prove a lower bound.

B Calculating $A(\mathbf{x}, \mathbf{y})$ explicitly

Recall that

$$
\mathcal{A} \;=\; \left[-\frac{1}{2}y^2 - \frac{1}{2}y^2\sigma'(x)\sigma(x)\right]\partial_x + \left[\mu(y) - \frac{1}{2}(\alpha'(y)\alpha(y) - \frac{\alpha(y)^2}{y})\right]\partial_y.
$$

24

We calculate that

$$
\begin{aligned}
A(x_0, y_0; x_1, y_1^*) &= \int_0^1 \langle \mathcal{A}, \dot{\gamma} \rangle \, dt \\
&= \int_0^1 \left[\frac{1}{\sigma(x)^2 y^2} \mathcal{A}^1 \frac{d\gamma^1}{dt} + \frac{1}{\alpha(y)^2} \mathcal{A}^2 \frac{d\gamma^2}{dt} \right] dt \\
&= \int_0^1 \left[\frac{1}{\sigma(x)^2 y^2} \left(-\frac{1}{2} y^2 - \frac{1}{2} y^2 \sigma'(x)\sigma(x) \right) \frac{d\gamma^1}{dt} + \frac{1}{\alpha(y)^2} \mathcal{A}^2 \frac{d\gamma^2}{dt} \right] dt \\
&= \oint_\gamma \left(-\frac{1}{2\sigma(x)^2} \left[1 + \sigma'(x)\sigma(x) \right] dx + \frac{1}{\alpha(y)^2} \left[\mu(y) - \frac{1}{2}(\alpha'(y)\alpha(y) - \frac{\alpha(y)^2}{y}) \right] dy \right).
\end{aligned}
$$

This is the integral of an exact 1-form so its value only depends upon the end points of γ. We deduce that

$$
A(x_0, y_0; x_1, y_1^*) = -\frac{1}{2} \int_{x_0}^{x_1} \frac{1}{\sigma(x)^2} \left[1 + \sigma'(x)\sigma(x) \right] dx + \int_{y_0}^{y_1^*} \frac{1}{\alpha(y)^2} \left[\mu(y) - \frac{1}{2}(\alpha'(y)\alpha(y) - \frac{\alpha(y)^2}{y}) \right] dy.
$$

C   Proof of Proposition 4.7

Denoting $\hat{p}(x_1, y_1, t) = \hat{p}_t(x_0, y_0; x_1, y_1)$, $q_t(x_1, y_1, t) = q_t(x_0, y_0, x_1, y_1)$, and substituting $\hat{p}(x, y, t) = h(x, y) \, q(x, y, t)$ into the original PDE (17), we need to find $h(x, y)$ such that the coefficients of $\partial_x q$ and $\partial_y q$ are $\frac{1}{2} y^2 \sigma'(x)\sigma(x)$ and $\frac{\alpha(y)}{2y}[y\alpha'(y) - \alpha(y)]$ respectively, i.e. agree with $\frac{1}{2}\Delta$ in the equivalent expression to (14) for the case when $\rho \neq 0$. Performing this substitution, we obtain

$$
\begin{aligned}
\partial_t q &= -\frac{1}{2} y^2 \sigma(x)^2 \left(\frac{\partial_x h}{h} q + \partial_x q \right) + \frac{1}{2} y^2 \sigma(x)^2 \left(\frac{\partial_x^2 h}{h} q + 2\frac{\partial_x h}{h} \partial_x q + \partial_x^2 q \right) + \mu(y) \left(\frac{\partial_y h}{h} q + \partial_y q \right) \\
&\quad + \frac{1}{2}\alpha(y)^2 \left(\frac{\partial_y^2 h}{h} q + 2\frac{\partial_y h}{h} \partial_y q + \partial_y^2 q \right) + \rho y \sigma(x)\alpha(y) \left(\frac{\partial_x \partial_y h}{h} q + \frac{\partial_x h}{h} \partial_y q + \frac{\partial_y h}{h} \partial_x q + \partial_x \partial_y q \right).
\end{aligned}
$$

Collecting coefficients of q and its derivatives, we have

$$
\begin{aligned}
\partial_t q &= y^2 \sigma(x)^2 \left(\frac{\partial_x h}{h} - \frac{1}{2} + \frac{\rho\alpha(y)}{\sigma(x)y} \frac{\partial_y h}{h} \right) \partial_x q + \left(\mu(y) + \alpha(y)^2 \frac{\partial_y h}{h} + \rho y \sigma(x)\alpha(y) \frac{\partial_x h}{h} \right) \partial_y q \\
&\quad + \frac{1}{2} y^2 \sigma(x)^2 \partial_x^2 q + \frac{1}{2}\alpha(y)^2 \partial_y^2 q + \rho y \sigma(x)\alpha(y)\partial_x \partial_y q + V(x, y)q,
\end{aligned}
$$

where

$$
V(x, y) = \frac{(\mathcal{A} + \frac{1}{2}\Delta)h}{h}. \tag{C-1}
$$

Thus, to determine the function h, we impose that

$$
y^2 \sigma(x)^2 \left(\frac{\partial_x h}{h} - \frac{1}{2} + \frac{\rho\alpha(y)}{\sigma(x)y} \frac{\partial_y h}{h} \right) = \frac{1}{2} y^2 \sigma'(x)\sigma(x), \tag{C-2}
$$

$$
\mu(y) + \alpha(y)^2 \frac{\partial_y h}{h} + \rho\sigma(x)y\alpha(y)\frac{\partial_x h}{h} = \frac{\alpha(y)}{2y}[y\alpha'(y) - \alpha(y)]. \tag{C-3}
$$

From (C-2) we obtain that $\partial_x h / h = \frac{1}{2}(1 + \frac{\sigma'(x)}{\sigma(x)}) - \frac{\rho\alpha(y)}{\sigma(x)y}\partial_y h/h$. Plugging this into (C-3), we have

$$
\frac{\partial_y h}{h} = \frac{1}{1 - \rho^2} \left[\frac{\alpha'(y)}{2\alpha(y)} - \frac{1}{2y} - \frac{\mu(y)}{\alpha(y)^2} - \frac{\rho y}{2\alpha(y)}(\sigma(x) + \sigma'(x)) \right] = g(x, y),
$$

so

$$
h(x, y) = h(x, 1) \exp \left[\int_1^y g(x, u)du \right].
$$

<div align="center">25</div>

Now suppose $h(\cdot,\cdot)$ does exist and $h(x_0, y_0) \neq 0$ for some $(x_0, y_0) \in \mathbb{R} \times \mathbb{R}_+$, then the function $B(x,y) := \log|h(x,y)|$ is a twice continuously differentiable function locally defined in a small neighborhood of (x_0, y_0). It is easily seen that in this neighborhood, we have

$$\partial_y B(x,y) = g(x,y), \quad \partial_x B(x,y) = \frac{1}{2}\left(1 + \frac{\sigma'(x)}{\sigma(x)}\right) - \frac{\rho\alpha(y)}{\sigma(x)y}g(x,y). \tag{C-4}$$

However, for $B(x,y)$ to be locally twice continuously differentiable, necessarily, $\partial_x\partial_y B(x,y) = \partial_y\partial_x B(x,y)$, which leads to

$$
\begin{aligned}
-\frac{\rho}{1-\rho^2}\frac{y}{2\alpha(y)}(\sigma'(x)+\sigma''(x)) &= \partial_y\left[\frac{1}{2}\left(1+\frac{\sigma'(x)}{\sigma(x)}\right) - \frac{\rho\alpha(y)}{\sigma(x)y}g(x,y)\right] \tag{C-5}\\
&= -\frac{\rho}{1-\rho^2}\frac{1}{\sigma(x)}\partial_y\left[\frac{\alpha(y)}{y}\left(\frac{\alpha'(y)}{2\alpha(y)} - \frac{1}{2y} - \frac{\mu(y)}{\alpha(y)^2} - \frac{\rho y}{2\alpha(y)}(\sigma(x)+\sigma'(x))\right)\right]\\
&= -\frac{\rho}{1-\rho^2}\frac{1}{\sigma(x)}\partial_y\left[\frac{\alpha(y)}{y}\left(\frac{\alpha'(y)}{2\alpha(y)} - \frac{1}{2y} - \frac{\mu(y)}{\alpha(y)^2}\right)\right].
\end{aligned}
$$

When $\rho = 0$, the above equality holds automatically. If $\rho \neq 0, \pm 1$, then the above equality is equivalent to

$$(\sigma'(x) + \sigma''(x))\sigma(x) = \frac{2\alpha(y)}{y}\partial_y\left[\frac{\alpha(y)}{y}\left(\frac{\alpha'(y)}{2\alpha(y)} - \frac{1}{2y} - \frac{\mu(y)}{\alpha(y)^2}\right)\right], \tag{C-6}$$

which cannot hold unless both sides are equal to a constant b. Suppose this is the case and $b < 0$, then $\frac{b}{\underline{\sigma}} < \sigma'(x) + \sigma''(x) < \frac{b}{\bar{\sigma}} < 0$. Multiplying both sides by e^x we obtain

$$\frac{b}{\underline{\sigma}}e^x \quad < \quad e^x(\sigma'(x) + \sigma''(x)) = (e^x\sigma'(x))' \le \frac{b}{\bar{\sigma}}e^x \quad < \quad 0. \tag{C-7}$$

If we now integrate both inequalities from 0 to $x > 0$, we obtain that

$$\frac{b}{\underline{\sigma}}(e^x - 1) \quad < \quad e^x\sigma'(x) - \sigma'(0) \quad < \quad \frac{b}{\bar{\sigma}}(e^x - 1) \quad < \quad 0. \tag{C-8}$$

Solving for $\sigma'(x)$ from the above inequalities, we have

$$e^{-x}\sigma'(0) + \frac{b}{\underline{\sigma}}(1 - e^{-x}) \quad < \quad \sigma'(x) \quad < \quad e^{-x}\sigma'(0) + \frac{b}{\bar{\sigma}}(1 - e^{-x}). \tag{C-9}$$

Letting $x \to \infty$, we see that $\sigma'(x) < \frac{b}{\bar{\sigma}} < 0$, which contradicts the assumption that σ is smooth and uniformly bounded. Similarly, we can show that b cannot be strictly positive. Hence, the only possibility is when $\sigma'(x) + \sigma''(x) \equiv 0$. In this case, the only positive bounded solution is $\sigma(x) \equiv \sigma_0$ for some positive constant σ_0.

In conclusion, a necessary condition for $h(x,y)$ to exist is that $\sigma(x)$ is a constant and

$$\frac{\alpha(y)}{y}\left(\frac{\alpha'(y)}{2\alpha(y)} - \frac{1}{2y} - \frac{\mu(y)}{\alpha(y)^2}\right) \equiv c \tag{C-10}$$

for some constant c. Conversely, if this holds, then, by the criterion for an exact differential (see e.g. page 61 of Donaldson[Don11]), we can find a B satisfying (C-4). Hence if we take $h := e^B$ it will satisfy (C-2) and (C-3).

Re-arranging (C-10), we obtain

$$\mu(y) = \frac{\alpha(y)}{2y}[y\alpha'(y) - \alpha(y)] - cy\alpha(y).$$

But substituting the asymptotic behaviour for $\alpha(y)$ as $y \to \infty$ given in Assumption 3.1, we find that $\mu(y) \to \infty$ unless $c \ge 0$ ($y\alpha(y)$ is the dominating term). On the other hand, by applying the mean value theorem to $\alpha'(y)$, we have $\alpha'(y) = A_1 + \alpha''(\zeta_y)y$, where $\zeta_y \in (0, y)$ is some point depending on y. Hence, as $y \to 0$, we have

$$\mu(y) = \frac{A_1}{2}(1 + o(1))\alpha''(\zeta_y)y^2 - cA_1y^2(1 + o(1)).$$

Because $\alpha''(y) \to 0$ as $y \to 0$, we know that $-cA_1y^2$ is the leading term of $\mu(y)$ as $y \to 0$. In order for $\mu(y) \ge 0$ for all sufficiently small $y > 0$, we must have $c \le 0$. Overall, the only suitable choice of c is zero. As a consequence, we have

that

$$\frac{\partial_y h}{h} = g(x, y) \equiv g(y) := -\frac{\rho\sigma_0}{2(1-\rho^2)}\frac{y}{\alpha(y)}, \quad \frac{\partial_x h}{h} = \frac{1}{2} - \frac{\rho\alpha(y)}{\sigma_0 y}\frac{\partial_y h}{h} = \frac{1}{2(1-\rho^2)} := C,$$

and hence $\partial_x^2 h = C^2 h$ and $\partial_y^2 h = hg' + g\partial_y h$. Using these relations and (C-1) we find that

$$
\begin{aligned}
V(x, y) &= \frac{(\frac{1}{2}\sigma_0^2 y^2 \partial_x^2 + \rho\sigma_0 y\alpha(y)\partial_x\partial_y + \frac{1}{2}\alpha(y)^2\partial_y^2 - \frac{1}{2}\sigma_0^2 y^2\partial_x + \mu(y)\partial_y)h}{h} \\
&= \frac{1}{2}\sigma_0^2 y^2 C(C-1) + \rho\sigma_0 y\alpha(y)Cg(y) + \frac{1}{2}\alpha(y)^2(g(y)^2 + g'(y)) + \mu(y)g(y) \\
&=: V(y).
\end{aligned}
\tag{C-11}
$$

Using that $\alpha(y) \sim A_1 y$ as $y \to 0$ and $\alpha(y) \sim B_1 y^p$ as $y \to \infty$ we find that

$$V(y) \quad \sim \quad -\frac{\sigma_0^2 y^2}{8(1-\rho^2)} \qquad \text{(as } y \to \infty \text{ and as } y \to 0),$$

and $V(x, y) < \infty$ for all x, y, so V is bounded from above, as required.