Perturbation and Symmetry Techniques Applied to Finance

A Dissertation
submitted to obtain the title of Doctor of Philosophy (Dr. rer. pol.)
from the Frankfurt School of Finance & Management.
Specialty: Quantitative Finance

Submitted by: Stephen Taylor

Dissertation Advisor: Jan Vecer

March 2010

Declaration of Authorship: I hereby certify that unless otherwise indicated in the text of references, or acknowledged, this dissertation is entirely the product of my own work.

Stephen Taylor

Contents

1. Introduction

A significant portion of the mathematical finance literature is devoted to constructing exact pricing formulas for a variety of contingent claims whose underlying assets are assumed to evolve according to a specified stochastic process. In particular, the Black-Scholes-Merton [18], [119], [120], pricing formula for a European call option on an asset which undergoes lognormal dynamics has served as a cornerstone for much of the development of modern pricing theory. Although the lognormal model has long been a guide for gaining insight into how derivatives' prices depend on market parameters, it is often too simplistic for practical use and has many drawbacks. Specifically, asset prices generally do not follow lognormal dynamics, and one is typically unable to calibrate this model in a manner that is consistent with current market option prices.

These issues, amongst others, led Dupire [48] to consider the more general class of local volatility models. A local volatility model allows an asset process to evolve by a one-dimensional Itō process, where the diffusion coefficient function is chosen so that the model's present day European call prices agree with current market quotes. Unfortunately, one cannot determine the transition density for a general Itō process (which is equivalent to constructing an explicit formula (up to quadrature) for a call option). There are only a handful of local volatility models where it is possible to derive an expression for the transition density given by a composition of elementary and special functions. Constructing such a solution is equivalent to solving a single linear parabolic equation of one spatial and one temporal variable. Generally, as the functional form of the diffusion coefficient becomes more complicated, the prospect of finding an exact solution to the associated transition density equation diminishes.

In lieu of exact formulas, there are several widely used approximate pricing techniques. For example, tree pricing, Monte Carlo, and finite difference methods are currently the most popular approximation methods used in practice. Alternatively, one can attempt to construct explicit approximation formulas for the transition density of a local volatility process that generically will only be valid in certain model parameter regimes. There are a variety of approximation techniques that have been applied to one dimensional local volatility models, (c.f. Corielli et. al. [39], Cheng et. al. [31], Kristensen and Mele [97], and Ekstrom and Tysk [52] for example) The advantage of such formulas is that they are considerably more computationally efficient, relative to their previously mentioned counterparts, and, at times, retain qualitative model information in their functional form. The main drawback of these approximations is that they can be inaccurate for certain model parameters, and a thorough error analysis against an established approximation method is usually required prior to real world use.

The literature regarding explicit approximation formulas, which are generically constructed using different types of perturbation theory, has been growing over the past several years. In particular, heat kernel perturbation theory has proven to be effective in creating some of the most accurate approximation formulas to date. Roughly speaking, heat kernel perturbation theory utilizes a Taylor series expansion ansatz in time together with a geometrically motivated prefactor which is chosen to simplify subsequent calculations. The heat kernel ansatz is most naturally stated using the language of differential geometry. Hagan and Lesnewski [78], [79], [80], [106] were the first to introduce differential geometric methods in finance. Henry-Labordere [98], [99], [101], used heat kernel methods to construct implied volatility formulas for local volatility models, the SABR model, and the SABR Libor Market Model. More recently, Gatheral, et. al. [64] have also considered heat kernel expansions in the context of local volatility models. Improving upon the work of Henry-Labordere, Paulot

[134] constructed (to our knowledge) the most accurate explicit implied volatility formula for the SABR model. Forde [58, 59] has rigorized and extended this work to a larger class of stochastic volatility models. We refer the reader to Medvedev's work [114] for a list of several references to the finance perturbation theory literature.

There are two main advantages of heat kernel methods over other forms of perturbation theory. First, the only error in a formula constructed solely using a heat kernel ansatz is due to a Taylor series assumption in time. In particular, an expression for a transition density constructed using heat kernel perturbation theory will increase in accuracy as time decreases. Consequently, implied volatility smile approximation formulas constructed from a heat kernel expansion tend to be more accurate at out of the money strikes than others constructed with alternative formalisms that are perturbative in both time and strike. Secondly, the heat kernel ansatz solves its associated parabolic equation exactly to zeroth order in time. Thus in some sense, one does not carry along zeroth order complexities when trying to establish a first order correction for the transition density of a process.

Heat kernel perturbation theory is not without its limitations. Even in the case of two dimensional models, the heat kernel ansatz can not be expressed explicitly. This is due to the fact that one can not obtain an explicit distance function on a generic two dimensional geometry. There are other techniques in differential equations which one can use to attempt to reduce the complexity of a given pricing equation. One such technique is Lie symmetry analysis.

Lie symmetry analysis provides a means to reduce the dimension of a partial differential equation by using the symmetry of the equation to construct a natural coordinate system in which it takes a simpler form. This can be particularly useful in finance as one can reduce the dimension of a given pricing equation and then numerically simulate the reduced equation at a significantly faster speed than the original.

We study the symmetry groups of several equations relevant to finance including the heat equation, Black-Scholes-Merton equation, the classical Asian equation in the lognormal model, and several related Asian type equations. We examine both the heat and Black-Scholes-Merton equations in order to demonstrate our method for computing symmetry groups of partial difference equations. We then apply this method to the standard Asian equation and find many interesting symmetries which allow one to reduce the equation to a lower dimensional problem.

In particular, we first review the construction of the symmetry group of the Asian equation (c.f. Glasgow and Taylor [154]) and show that there are several ways one can reduce this equation by one spatial dimension. We next consider several additional equations which can be constructed the from Asian equation by a change of numeraire. Finally, we examine several relations between these different equations.

This work is organized in the following manner: In section 2, we provide basic conventions and background and review the construction of the heat kernel ansatz. Specifically, we first review topics in probability theory, stochastic differential equations, and Riemannian geometry which we will need throughout the remainder of the text. In addition, we review the heat kernel ansatz for a partial differential equation and then demonstrate how it can be used to construct explicit approximation formulas for the transition density of stochastic processes. Such formulas can be used in conjunction with integral approximation techniques to enable one to construct explicit formulas for path independent derivatives.

In section 3, we state the explicit form of the heat kernel ansatz in one dimension. We then apply this ansatz to several local volatility models. We first consider the Black-Scholes-Merton model, where we show that one can use the heat kernel method to construct the exact solution of the European call equation whose underlying evolves according to lognormal dynamics. Thus we provide a novel, although somewhat roundabout way, of solving the Black Scholes equation. Next, we examine the CEV local volatility model. We construct an n-th order approximation formula for this model and comment on convergence issues related to the resulting formula as well as their implications for the related formulas for the SABR model. Next, we look at the quadratic local volatility model. Here, we are able to derive the exact transition density for this model by using heat kernel perturbation theory to solve this relevant pricing PDE to arbitrarily high order and then inverting a related Taylor series to construct an exact solution for the transition density. We then turn to three new local volatility models called the cubic local volatility model, the affine-affine model, and the generalized CEV model. We are able to accurately approximate the transition densities of each of these models in certain model parameter regimes.

In section 4, we turn attention to studying two dimensional stochastic volatility models. We first consider a correspondence between stochastic volatility models and related constructions in differential geometry. We next focus on the SABR model. This stochastic volatility model is widely used in industry across asset classes. This model is popular because Hagan et. al, in [77], constructed an explicit implied volatility model for European call options in terms of SABR model parameters. However, this formula is known to degenerate for certain parameters values which include strike values for options that are highly out of the money. There are other methods to construct implied volatility approximation formulas for this model. Currently, the singly more accurate explicit implied volatility formula for this model was constructed by using heat kernel perturbation theory. We review, and partially extend, the work of Paulot [134] where we apply the heat kernel method to the SABR model. Next, we show that the resulting approximation method satisfies Lee's moment formula. After this, we look at a similar construction in the Heston model and examine a class of stochastic volatility models that one can also apply heat kernel methods to in order to construct explicit implied volatility formulas. We lastly, consider an alternative type of perturbation for stochastic volatility models and make connections with these and certain constructs in geometry.

Finally, in section 5, we show how Lie symmetry analysis can be applied to pricing problems in finance. We first review the subject as well as our method for computing the symmetry group of a given differential equation. We then demonstrate this method in the case of the heat and Black-Scholes-Merton equations. Next, we turn to the two dimensional Asian equation whose underlying evolves according to lognormal dynamics, and compute its full symmetry group. This leads to several dimensional reductions of the Asian equation which can be used to speed up the pricing of Asian options. We next analyze the symmetry groups of several other Asian type equations which are related to the original equation via a change of numeraire.

Acknowledgements: There are many people who have helped me throughout my stay in academia that I would like to acknowledge. In chronological order, I would like thank Ronald Selby for teaching me calculus at my secondary school. I thank Michael Dorff and Eric Hirschmann for their time in advising both my Masters theses and teaching me about complex analysis, geometry, and relativity. Next, I would like to thank Scott Glasgow for teaching me many things about differential equations, helping me understand the research process, and introducing me to the world of mathematical finance. I would also like to thank my past graduate student colleges Aaron Petersen and Marcelo Disconzi for many fruitful discussions about a variety of topics.

There are also many people at Bloomberg that have helped me shape portions of this work. I thank Fabio Mercurio for playing a significant role by suggesting research topics as well as teaching many things about interest rate models. I would like to thank all my colleagues in

Bruno Dupire's Quantitative Finance Research group for many useful comments related to this work, as well as continually teaching me many things on a variety of topics in real world finance.

Finally, I would like to thank my dissertation advisor Jan Vecer for educating me about how numeraire techniques can be applied in finance often in a manner which allows one to considerably simplify pricing problems. I would also like to thank him for suggesting many research questions regarding Asian options and taking the time to serve as my advisor. Also, I would like to thank the Frankfurt School for providing me with the means to complete my doctoral work.

I would lastly like to thank my parents for encouraging me to pursue higher education as well as instilling me with a strong desire to continually learn new things.

2. Background Material

We first summarize background material in probability theory that we will use throughout the remainder of the text. Next, we turn to recalling relevant topics related to stochastic differential equations, Riemannian geometry. We then combine use these results to construct the heat kernel expansion ansatz for differential equations.

2.1. **Probability Theory and Stochastic Processes.**

We first review basic aspects of probability theory and stochastic processes which will be utilized throughout the remainder of the text. In particular, we discuss relevant aspects of probability theory including, Brownian motion, stochastic processes constructed from Brownian motion, along with probability and transition density functions associated to these stochastic processes. Our aim is not to be rigorous or thorough and we defer to Borodin-Salminen [20], Evans [53], Jones [90], Joshi [91], Oksendal [130], Rudin [142], and Shreve [144] for deeper discussions of these topics.

We start with the notion of a probability space. Probability spaces are the fundamental background objects upon which stochastic processes are defined. A probability space is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a set called the event space, $\mathcal{F}$ is a sigma algebra over Ω, and $\mathbb{P}$ is a probability measure, i.e. a measure on Ω that satisfies the additional condition $\mathbb{P}(\Omega) = 1$. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a real-valued one-dimensional random variable is a function $X : \Omega \to \mathbb{R}$ that is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable; that is to say, $X^{-1}(\mathcal{B}) \in \mathcal{F}$ for every $B \in \mathcal{B}(\mathbb{R})$ where we let $\mathcal{B}(\mathbb{R})$ denote the collection of Borel subsets of $\mathbb{R}$. For any random variable X, there is an associated cumulative distribution function $F_X : \mathbb{R} \to [0, 1]$ defined on intervals $[a, b] \subset \mathbb{R}$ by $F_X([a, b]) = \mathbb{P}(a \leq X \leq b)$. If there is a function $p : \mathbb{R} \to \mathbb{R}^+$ such that

$$F_X([a, b]) = \int_a^b p(x)dx \tag{2.1}$$

for any $a, b \in \mathbb{R}$ with $a \leq b$, then $p(x)$ is called the density function (or probability density function) associated to X. If X is a standard normal random variable ($X \sim \mathcal{N}(0, 1)$), then we denote its density and distribution functions by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{and} \quad \Phi(x) = \int_{-\infty}^{x} \phi(u)du. \tag{2.2}$$

If a random variable X admits a density function p, then one is able to compute many important quantities associated to X by evaluating Riemann integrals which involve p. In particular, the expected value functional is defined by

$$\mu(X) = \mathbb{E}[X] \equiv \int_{\omega \in \Omega} X(\omega)d\mathbb{P}(\omega) = \int_{\mathbb{R}} xf(x)dx. \tag{2.3}$$

The moments of X are also given by similar expressions. In particular, one can compute the n-th moment of X^n by evaluating

$$\mu_n(X) = \mathbb{E}[(X - \mathbb{E}X)^n] = \int_{\mathbb{R}} (x - \mu)^n f(x)dx. \tag{2.4}$$

Here $\mu_0 = 1$, $\mu_1 = 0$, and $\mu_2 = \mathrm{Var}(X)$, where $\mathrm{Var}(X)$ is the variance of X; also, the third and forth moments define the skewness and kurtosis of X.

Stochastic processes are collections of random variables on a fixed probability space. More specifically, a stochastic process X_t is a continuum of random variables $\{X_t | t \in [a, b] \subset \mathbb{R}\}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The only stochastic processes we consider will be constructed from Brownian motions (Wiener processes) and deterministic quantities. Thus we will not enter into the realm of Lévy processes or semimartingales. A Brownian motion is a stochastic process W_t on a time interval $\mathcal{T} = [0, T]$ which initially takes the value $W_0 = 0$, is continuous almost everywhere, and for $t, s \in \mathcal{T}$ with $t < s$, $W_t - W_s$ is a normal random variable with mean W_s and variance $t - s$, i.e. $W_t - W_s \sim \mathcal{N}(W_s, t - s)$. Given a Brownian motion W_t, a filtration $\mathcal{F}(t)$ for W_t is a continuum of sigma algebras such that $\mathcal{F}(t_1) \subset \mathcal{F}(t_2)$ for $t_1 < t_2$ where W_t is $\mathcal{F}(t)$ measurable, and has independent increments. A filtration can be viewed as a refinement of a sigma algebra over time and $\mathcal{F}(t)$ can be interpreted as the totality of all information known at time t which monotonically increases as time progresses. We will use the notation $\mathbb{E}[X|\mathcal{F}(s)]$ to represent the conditional expected value of a process X, given that the value of X is known at time $t = s$.

Brownian motions have many properties which make them somewhat amenable processes to work with. One of these is the martingale property $\mathbb{E}[W(t)|\mathcal{F}(s)] = W(s)$ which demonstrates that a Brownian motion has no tendency on average to rise of fall above its known level at time s and has trivial expected value if $s = 0$. If we let W_t evolve from $t = 0$ to $t = t_1$ and find it has value $W_{t_1} = \mu_1$, then the martingale property requires that $W_t \sim \mathcal{N}(\mu_1, t - t_1)$ for all $t > t_1$. Note that the variance of W_t grows like $\sqrt{t - t_1}$ in time; hence, a Brownian motion is diffusive in nature.

We can construct new processes from Brownian motions by rescaling W_t and summing it together with a deterministic function. For example, let $f_t = f(t)$ be a measurable function and consider the process $a_t = f_t + \sigma W_t$ for some constant σ. Then σ rescales the variance of the Brownian motion and f_t introduces a deterministic drift term. At time zero, the expected value of a_t is just the expected value of f, i.e. $\mathbb{E}(a_t) = \mathbb{E}(f)$. One can choose values of σ in a_t to suitably adjust the inherent randomness of the process. For $\sigma \approx 0$, the process will be mostly deterministic and accurately approximate $f(t)$ whereas when σ becomes large, graphs of the process will become increasingly randomly perturbed about the graph of f.

We now wish to construct another important family of stochastic process called Itō processes which are fundamental objects in continuous time finance and also the most general process that we consider. Let $\Pi = \{t_0, \dots, t_n\}$ be an n-point partition of the interval $[0, t]$ with $t_i < t_{i+1}$, $t_0 = 0$, and $t_{n-1} = t$, and for a given deterministic process σ_t, i.e. a function, consider the stochastic process

$$I_n(t) = \sum_{\Pi} \sigma_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}), \tag{2.5}$$

where we note here that σ is evaluated at the left endpoints of the subintervals of Π. The left endpoint evaluation of the integrand of the Itō process is important as it makes the process non-anticipative by construction. In fact, many of the useful properties of Itō processes which make them desirable tools for financial modeling stem from this aspect of their definition.

Taking a limit of any refining sequence of partitions that simultaneously satisfy $|\Pi| \to 0$ and $n \to \infty$, allows us to construct a new unique (modulo measure zero sets) stochastic process

$$I_t = \lim_{|\Pi| \to 0} \sum_{\Pi} \sigma_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}) \equiv \int_0^t \sigma(s) dW_s. \tag{2.6}$$

The process I_t is called an Itō integral, and it is an example of an Itō process I_t defined by

$$I_t = I_0 + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW(s) \tag{2.7}$$

where $\mu \in L^1[0, t]$, and $\sigma \in L^2[0, t]$ are deterministic functions where here L^1 and L^2 represent L^p function spaces. Itō processes are the basic dynamical objects used to model the evolution of continuous time financial quantities.

Now let S denote an Itō process which starts at an initial value S_t at time t and evolves to some future value S_T at time $T > t$. Then we will let $p(t, S_t, T, S_T)$ denote the probability transition density of S. Here p given the probability of S_t evolving to S_T. For example, if S is a Brownian motion, then p is a normalized Gaussian centered at S_t. Transition densities are important in finance since if one can obtain an exact form for the transition density for an Itō process, then he can price a call option on an asset which evolves by this process up to quadrature. We will see later that the transition density for an Itō process can always be described as the solution of a parabolic partial differential equation.

2.2. **Stochastic Differential Equations.** Working with Itō processes written in the form of equation (2.7) quickly becomes quite tedious. Stochastic differential equations are alternative representations of Itō processes in the analogous manner that any linear Volterra integral equation of the first kind can be converted to an associated first order differential equation (This should not be taken literally as we lose no regularity in the SDE case). For $i = 1, \dots, n$, let x_t^i be n Itō processes defined by

$$x_t^i = x_0^i + \int_0^t b^i(s, x_s)ds + \int_0^t \sum_{j=1}^m \sigma_j^i(s, x_s)dW_s^j \tag{2.8}$$

where x_0^i are the initial values of x_t^i, x_s represents the collection of all the x_t^i, $b^i : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$,$\sigma_j^i : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^+$ are measurable and have suitable regularity for the process to be well-defined, and W^j are m Brownian motions with correlation matrix ρ. We can formally write this system of Itō processes as a system of stochastic differential equations (SDEs)

$$dx_t^i = b^i(t, x_t)dt + \sum_{j=1}^m \sigma_j^i(t, x_t)dW_t^j. \tag{2.9}$$

There are two ways in which these equations are coupled together. The first is that the coefficient functions b^i and σ_j^i for the i-th equation can depend on the other x_t^j, which is exactly how one couples systems of deterministic first order differential equations. Secondly, the Brownian motions are coupled via $\mathbb{E}[W_t^j W_t^i] = \rho^{ij} dt$ where ρ^{ij} is a symmetric positive definite covariance matrix. Note that SDEs can only be first order by construction. However, we can effectively think of such systems as higher order linear equations in the same sense that second order linear ordinary differential equations can be decomposed into systems of first order equations.

There is a somewhat weak standard existence/uniqueness theorem for systems of SDEs which is given by the following.

Theorem 2.1. [53, p.86] *(Evans) Let $\mu : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times [0, T] \to \mathbb{R}^{m \times n}$ be continuous Lipschitz functions with Lipshitz constant L that satisfy the linear growth conditions*

$$|\mu(x, t)| \le L(1 + |x|), \quad |\sigma(x, t)| \le L(1 + |x|) \tag{2.10}$$

for all $t \in [0, T]$ and $x \in \mathbb{R}^n$. Let x_0 be an $\mathbb{R}^n$ valued random variables with $\mathbb{E}(|x_0|^2) < \infty$ and let W be an m-dimensional Brownian motion independent of x_0. Then there exists a unique solution $x \in L^2_{\mathbb{R}^n}(0, T)$ of the stochastic differential equation system

$$dx = \mu(x, t)dt + \sigma(x, t)dW, \quad x(0) = x_0. \tag{2.11}$$

modulo sets of measure zero.

Unfortunately, many SDE systems that commonly arise in financial modeling do not satisfy the hypotheses of this theorem. For such systems, we will operate under the typical, but possibly dangerous, assumption that there is indeed a well-defined unique solution of the SDE system's underlying stochastic processes. There are a variety of more specialized results for SDEs similar in spirit to this theorem.

If we know that a system of SDEs admits a unique solution, we often want to understand the associated unique dynamics of functions of the underlying stochastic processes. Itō's lemma provides us with this information:

Lemma 2.2. [53, p. 66] *(Evans) Consider an n-dimensional system of SDEs*

$$dx^i = \mu^i(t,x)dt + \sum_{j=1}^{m} \sigma^i_j(t,x)dW^j \tag{2.12}$$

where $\mu \in L^1(0,T)$ and $\sigma \in L^2(0,T)$. Let $u : [0,T] \times \mathbb{R} \to \mathbb{R}$ be $C^{1,2}([0,T] \times \mathbb{R})$. Define a process $Y(t) = u(x(t),t)$. Then Y has dynamics given by

$$dY = \left(\frac{\partial Y}{\partial t} + \sum_i \mu^i \frac{\partial Y}{\partial x^i} + \frac{1}{2} \sum_{i,j,k} \sigma^i_k \sigma^j_k \frac{\partial^2 Y}{\partial x^i \partial x^j} \right) dt + \sum_{i,j} \sigma^i_j(x,t) \frac{\partial Y}{\partial x^i} dW^j. \tag{2.13}$$

We note this can be written in the somewhat more appealing form

$$dY = \frac{\partial Y}{\partial t} dt + \sum_i^n \frac{\partial Y}{\partial x^i} dx^i + \frac{1}{2} \sum_{i,j}^n \frac{\partial^2 Y}{\partial x^i \partial x^j} dx^i dx^j, \tag{2.14}$$

where one can substitute equation (2.12) into this equation and use the formal rules $dt^2 = dtdW^i = 0$, $dW^i dW^j = \delta^{ij} dt$ (which follow from properties of the Itō integral Evans [53]) to derive equation (2.13). We will often employ this computational convenience when utilizing Itō's lemma in future calculations. We now comment on how SDEs can be used to model the evolution of financial quantities.

2.3. **Financial Applications of Stochastic Differential Equations.** There are a myriad number of ways SDEs are used for modeling purposes in finance. We are interested in studying their use in modeling the evolution of equities, exchange rates, interest rates, etc, which will allow us to price contingent claims (derivative contracts) whose underlying assets evolve according to a specified SDE model.

A market model (or model for short) is a collection of n Itō processes x^i which are specified by a system of SDEs

$$dx^i_t = \mu^i(x,t)dt + \sum_{j=1}^{m} \sigma^i_j(x,t)dW^j, \tag{2.15}$$

where the x^i_t have initial values x^i_0 and μ^i, σ^i_j are chosen appropriately so that the system admits a unique solution. Also, the Brownian motions satisfy $\mathbb{E}[dW^i dW^j] = \rho^{ij} dt$ where here $\mathbb{E}$ is defined with respect to the probability measure $\mathbb{P}$ under which W^j are Brownian. In addition to this system of equations, there is a separate SDE in a market model given by

$$dB_t = r_t B_t dt, \quad B_0 = 1. \tag{2.16}$$

Here B_t is interpreted as a process which models the growth of one unit of currency according to a risk free rate r_t. The risk free rate r_t can be taken to be constant, or modeled by another stochastic process (see for instance Brigo and Mercurio [23, Ch 3]) and is assumed not to depend on the x^i since we will assure that the x^i represent only a few asset prices whose values will not significantly impact the global risk-free rate. Equation (2.16) can be integrated exactly. In particular,

$$B(t) = \exp\left(\int_0^t r(s)ds\right). \tag{2.17}$$

An important quantity associated with $B(t)$ is the discount factor or zero coupon bond price at time $t = 0$ given by $D(0,t) = B(0,t)^{-1}$. The zero coupon bond curve encodes the market's current view on future expected values of the risk free rate. It allows us to compute the present day value of an asset whose value is known at time t by multiplication. In particular, if an asset has value $C(t)$ at some future time $t > 0$, then at the present time $t = 0$, it has present day value $D(0,t)C(t)$. We will typically assume that the yield curve is flat.

The x_i are typically interpreted as "market observables", and are commonly taken to model asset prices, short rates, forward rates, or foreign exchange rates. We will call μ^i the drift associated to x^i and σ^i_j the instantaneous volatility associated with the j-th Brownian motion of x^i.

In particular, we will usually be interested in the case of European derivatives. These are contracts that depend on a market model's assets that can only be exercised at some fixed future date $T > 0$. Given a market model, we can define a derivative which depends on underlying assets x_i by a payoff function $C_T = C(x_t)$ which may depend on the paths of the x_i (which is the case of Asian and other exotic options). Our task then is to find the fair value C_t of the contract at prior times $t \in [0, T)$.

The purpose of pricing theory is not to attempt to predict a single evolution of asset prices. Rather, one of the main questions it attempts to address is: given a market model, i.e. a collection of assumed dynamics for n assets and a risk free interest rate, what is the fair value of a specified derivative whose underlying asset(s) evolve according to the specified market-model dynamics? A market model is typically characterized by several market parameters including (but not limited to) volatility of the underlying, a mean reversion level, a mean reversion speed or an exponent with controls the relative grown of the underlying. Before pricing a given contract, one needs to choose values for these model parameters. This process is called calibration and in the context of equities is performed by forcing model call prices to agree with current market prices.

There are a variety of SDEs one can use to model the evolution of the underlying financial observables. One simple example which is often used in a variety of contexts in practice due to its robust analytical tractability is to assume that the price process $S(t)$ evolves according to lognormal dynamics

$$dS = \mu S dt + \sigma S dW \tag{2.18}$$

where μ and σ are constants and $S(0) = S_0$. However, this two parameter family of models often does not simulate realistic price evolution. For instance, one can see from a histogram of the log of the daily returns from the S&P Index that the associated density is not normal. We can extend this to the more general class of time dependent local volatility models with drift

$$dS = \mu(S,t)dt + \sigma(S,t)dW, \tag{2.19}$$

which now are parameterized by the choice of two functions $\mu(S,t)$, $\sigma(S,t)$. In practice, one can choose these functions to be consistent with prevailing market conditions, in order to attempt to construct a more accurate approximation of the distributional properties of the price process (this process is called calibration). There is no need to limit the stochastic part of this SDE to being a Brownian motion. In fact, one can consider jump processes among other constructions in order to better approximate the true price process.

In order to construct a formula for the price of a generic European derivative, we first give a heuristic definition of the fundamental concept of arbitrage. A market model is said to be arbitrage free if there is no trading strategy one can execute on a portfolio of assets $x_i(t)$ of trivial initial value which has positive probability of having a positive value at any future time $T > 0$ and zero probability of having a negative future value. Alternatively, a market model is arbitrage free if there is no way to make an instantaneous risk free profit by trading in market assets; more colloquially, there is no free lunch. In other words, a market has no arbitrage if there is no way to definitively profit at a future time by trading assets in a portfolio with trivial initial value. Although real markets are not completely arbitrage free, such an assumption approximately holds for efficient markets and we will always operate under this assumption. The theory of arbitrage has a long history and we refer the reader to the standard Delbaen and Schachermayer [42] for a more detailed discussion.

The fundamental theorem of asset pricing will allow us to construct a pricing formula for derivatives in arbitrage free markets. We state it here in a simplified form.

Theorem 2.3. *(Fundamental Theorem of Asset Pricing) Consider a market model where the x_t^i are defined for $t \in [0, T]$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the market model is arbitrage free iff there exists a probability measure $\bar{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ such that the discounted assets $\bar{x}_t^i \equiv D(0,t)x_t^i$ are martingales with respect to $\bar{\mathbb{P}}$.*

The discounted asset dynamics under $\bar{\mathbb{P}}$ have trivial drift and as a result the measure is called the risk-neutral measure. Given such a measure, we can represent the price of a European derivative by using the following lemma:

Lemma 2.4. *Given a risk neutral measure $\bar{\mathbb{P}}$ for a market model, then for each $t \in [0, T)$, there exists a unique price π_t for a European derivative C_t with underlying assets x_i given by*

$$\pi_t = \mathbb{E}^{\bar{\mathbb{P}}}(D(t,T)C_T|\mathcal{F}_t). \tag{2.20}$$

See Brigo and Mercurio [23, Chpt. 2] for an explanation of these theorems as well as original references.

Here $\mathbb{E}^{\bar{\mathbb{P}}}$ is the conditional expectation operator with respect to the risk neutral measure given all available information $\mathcal{F}_t$ at time t and $D(t,T)$ is a discount factor given by $D(t,T) = D(0,T)/D(0,t)$.

Computing exact prices of derivatives in generic market models is a challenging task. In fact, only a handful of analytic pricing formulas are known, and most of these exist only for simple derivatives with correspondingly simple, low dimensional market dynamics. Often one either has to resort to Monte Carlo methods Glasserman [71], Kloedan and Platen [96] or numerical PDE Andersen and Piterbarg [2], Duffie [47] techniques to accurately price derivatives.

To demonstrate this theory, we now give a well known, but very important, example of pricing a European call option in the case of log-normal underlying dynamics. Suppose a discounted asset S_t evolves by log-normal dynamics for $t \in [0, T]$, under the risk neutral measure, given by

$$dS_t = \sigma S_t W_t, \quad S(0) = S_0, \tag{2.21}$$

for some volatility constant σ. Then we can use Itō's lemma to compute

$$d(\ln S) = \frac{1}{S}dS - \frac{1}{S^2}dS^2 = \sigma dW_t - \frac{\sigma^2}{2}dt, \tag{2.22}$$

and integrate over $[0, t]$ for $t < T$ to find

$$S_t = S_0 \exp\left(-\frac{\sigma^2}{2}t + \sigma W_t\right). \tag{2.23}$$

We now wish to price a European call on the asset S with strike K. This is a contract that allows its holder to purchase the asset S priced in dollars at time T for K dollars. The payoff function for such a call is given by $C_T = \max(S_T - K, 0)$, and at time $t = 0$, our pricing equation in the case of zero interest rates gives

$$C(0, S_0) = \mathbb{E}[[S_T - K]^+]. \tag{2.24}$$

Now since W_t is a $\mathcal{N}(0, T)$ random variable under the risk neutral measure, we can evaluate the call price by computing the integral

$$C(0, S_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(S_0 \exp\left(-\frac{\sigma^2}{2}T + \sigma\sqrt{T}x\right) - K\right)^+ \exp\left(-\frac{x^2}{2}\right) dx \tag{2.25}$$

$$= S_0\Phi(d_1) - K\Phi(d_2), \tag{2.26}$$

where here

$$d_i = \frac{\ln(S_0/K) + (-1)^{i-1}\sigma^2 T/2}{\sigma\sqrt{T}}, \tag{2.27}$$

which is the well known Black-Scholes-Merton pricing formula for a European call option. We denote this function by

$$C_{BSM}(\sigma, S_0, K, T) = S_0\Phi(d_1) - K\Phi(d_2). \tag{2.28}$$

This formula, although not viable real-world pricing, contains a considerable amount of structure which one can use to build intuition for how call options should be priced. One such important property of the formula is that the Vega

$$\frac{\partial C_{BSM}}{\partial \sigma} = S_0\sqrt{T}\Phi'(d_1(\sigma)) > 0. \tag{2.29}$$

Thus the BSM formula is monotonically increasing in the volatility σ and therefore there is an injective correspondence between prices of European call options and BSM volatility constants σ.

If one were to consider more complicated dynamics for the asset(s) and needs to price a European call on such an asset(s) it may not be possible to integrate the corresponding system of SDEs exactly, nevertheless evaluate the call option expectation (not to mention price more complicated derivatives). Suppose we have such a model and label it model A. From risk neutral pricing theory, we know there exists a unique price for a call at time zero given by C_0^A given that the underlying market is complete. Now suppose that model A depends on n parameters $a_1, \ldots, a_n$. The Black implied volatility of this model is a function $\sigma = \sigma(a_i)$ such that $C_0^A = C_{BSM}(\sigma)$. Note this is unique by monotonicity of $C_{BSM}(\sigma)$. Thus if we can construct a Black implied volatility function for a model, we can compute prices of European calls in this model simply by evaluating $C_{BSM}(\sigma)$. One of our aims will be to construct accurate Black implied volatility functions for complicated models.

Now in the BSM model, the volatility σ is assumed to be constant. In reality, the volatility is known to depend on both strike and maturity. In particular, the function $\sigma(K)$ is known as the smile and is not constant in the majority of derivatives markets. A large literature regarding the smile has developed over the past decade. Smile modeling was first considered in the equity and foreign-exchange setting by Dupire in [48],[49] ,[50], in the context of local volatility models. In addition, one can assume the instantaneous volatility functions of local volatility models are stochastic (the resulting model is called a stochastic volatility model). Stochastic volatility models allow us to give an alternative, although more complicated, way to model the smile as well. We will consider both cases of local volatility and stochastic volatility models later on.

We finally turn to another very important theorem in mathematical finance that allows us to evaluate expectations by solving associated deterministic parabolic equations; this provides an alternative way to price derivatives. The Feynman-Kac theorem states:

Theorem 2.5. *(Feynman-Kac [94]) Let $C(t,x) \in C^{(1,2)}$ and let $r_t > 0$ be a continuous interest rate process. Then assuming x_i and r are given by a market model that admits a unique solution, the price of a generic derivative*

$$C(t,x) = \mathbb{E}\left[\exp\left(-\int_t^T r_s ds\right)\Big|\mathcal{F}_t\right] \tag{2.30}$$

is given as the solution of the backward parabolic deterministic equation

$$-\partial_t C(t,x) = LC(t,x) - r_t C(t,x) \tag{2.31}$$

with terminal data given by $C(T,x) = C(T)$ where L is an elliptic operator given by

$$L = \mu^i \frac{\partial}{\partial x^i} + \frac{1}{2}\sum_{i,j}^{n}\sum_{k}^{m} \sigma_k^i \sigma_k^j \frac{\partial^2}{\partial x^i \partial x^j}. \tag{2.32}$$

The Feynman-Kac formula provides a link between SDE and PDE techniques for derivatives pricing. In particular, one can evaluate (2.31) by performing a Monte Carlo simulation, or numerically solving the differential equation (2.32). However there are several drawbacks to these approaches. In addition to many numerical issues that arise in pricing by Monte-Carlo or numerical PDE simulation, one major drawback is that both of these techniques generally become increasingly computationally intensive as the complexity of the underlying SDEs grows. The particular type of approximation method we will use requires the language of differential geometry. We now review a few pertinent aspects of this subject prior to utilizing them to approximation pricing PDE.

We will focus on PDE related pricing methods. Specifically, we will attempt to approximate the solution of equations of the form (2.32) using perturbation theory methods. This will result in approximation formulas which will provide a means for fast valuation of derivatives in complex market models relative to numerical PDE and Monte Carlo techniques.

2.4. **Riemannian Geometry.** One of our main objectives is to study the interaction between SDEs and certain constructions in Riemannian geometry. Before we can consider these constructions, we need to review some aspects of Riemannian geometry. We refer the reader to Cheeger and Ebin [30], Do Carmo [44], Do Carmo [45], Frankel [55], Hassani [82], Jose [92], Milnor [122], Petersen [135], Nakahara [126], for more in depth discussions of the subject.

In the following we will let M be a C^∞ differentiable manifold with tangent bundle TM and let $g: TM \otimes TM \to \mathbb{R}$ be a positive definite symmetric $(0,2)$ tensor, i.e. a Riemannian metric on M (see Warner [167] for basic manifold theory definitions). The pair (M,g) is called a Riemannian manifold. Typically, M is taken to have non-trivial topology and in fact most of the fundamental results of Riemannian geometry explore the relationship between the topology of M and the space of admissible metrics g that can be defined upon M. However, all of our analysis will be local, and thus without loss of generality, we will always take M to be an embedded graph in a Euclidean space, although we seldom take g to be the standard Euclidean metric.

We also will let E be a smooth vector bundle over M and let $\Gamma(E)$ denote the space of smooth sections of E. We do not define a vector bundle here and refer the interested reader to the standard references Milnor [121], Steenrod [149] for more references. A connection $\nabla^A : \Gamma(E) \to \Gamma(T^*M \otimes E)$ on E is a linear map which for any $v \in E$ and $f \in C^\infty(M)$ satisfies a Leibnitz rule

$$\nabla^A(fv) = df \otimes v + \nabla^A dv. \tag{2.33}$$

where here d is exterior differentiation. We will only be interested in local properties of vector bundles and since every vector bundle is locally trivial, we lose no generality in assuming that $E = M \times \mathbb{R}^n$ (in particular we are only interested in line bundles, i.e. bundles with $n = 1$). If we let e_i be the standard basis elements of $\mathbb{R}^n$, then we can represent a generic section v of E by $v(x) = v^i(x)e_i$, and the Leibnitz rule requires that

$$\nabla^A v = \nabla^A v^i \otimes e_i + \sigma^i \nabla^A e_i. \tag{2.34}$$

Now $\nabla^A e_i \in \Gamma(T^*M \otimes E)$ so there is a matrix A_i^j of one forms $A_i^j = A_{ik}^j dx^k$ such that $\nabla^A e_i = A_i^j \otimes e_j$. In particular, we have

$$\nabla^A v = (dv^i + \sigma^i A_j^i) \otimes e_i, \tag{2.35}$$

and the connection acts on components of the section like

$$\nabla_k^A v^i = \partial_k \sigma^i + A_{kj}^i \sigma^j. \tag{2.36}$$

We can write the connection operator more conveniently in local coordinates as

$$\nabla_k^A = \partial_k + A_k, \quad \text{where} \quad A_k = A_{kj}^i, \tag{2.37}$$

where in the case of a line bundle the A_k are just 1-by-1 matrices of one forms, or simply just the local coordinate components of one forms.

We now consider two examples of connections, in particular the Levi-Civita connection on the tangent bundle of (M, g) and an Abelian connection on a line bundle over (M, g).

The fundamental theorem of Riemannian geometry states:

Theorem 2.6. *(Do Carmo* [45] *, Petersen* [135]*) Given a Riemannian manifold (M, g), there is a unique connection ∇ defined on the tangent bundle TM that is torsion free, i.e.*

$$\nabla_X Y - \nabla_Y X = [X, Y] \tag{2.38}$$

and compatible with the metric ($\nabla g = 0$). This connection is called the Levi-Civita connection.

We denote the Levi-Civita connection by ∇ (without using a superscript) and its associated connection coefficients by Γ_{jk}^i which are called Christoffel symbols. The Levi-Civita connection acts locally on the components of a generic tensor $T : \otimes^r TM \otimes^q T^*M \to \mathbb{R}$ according to

$$\nabla_i T_{c\cdots d}^{a\cdots b} = \partial_i T_{c\cdots d}^{a\cdots b} + T_{a\cdots d}^{e\cdots b}\Gamma_{ie}^a + \cdots + T_{c\cdots b}^{a\cdots e}\Gamma_{ie}^b - T_{e\cdots d}^{a\cdots b}\Gamma_{ic}^e - \cdots - T_{c\cdots e}^{a\cdots b}\Gamma_{id}^e. \tag{2.39}$$

In particular, the metric compatibility condition demands that

$$0 = \nabla_k g_{ij} = \partial_k g_{ij} - \Gamma_{ki}^l g_{lj} - \Gamma_{kj}^l g_{il}, \tag{2.40}$$

which can be inverted to produce a formula for the Christoffel symbols (which are not tensors),

$$\Gamma_{bc}^a = \frac{1}{2} g^{ad} \left(\partial_b g_{dc} + \partial_c g_{db} - \partial_d g_{bc} \right). \tag{2.41}$$

These Christoffel symbols are symmetric in their lower two indices $\Gamma_{bc}^a = \Gamma_{cb}^a$.

We will mostly be concerned with connections on line bundles. In this case, the connection coefficients are given by $\Gamma^i_{jk} + A^i_{jk}$, so using our shorthand notation, ∇^A acts on the components of a vector v^i like

$$\nabla^A_j v^i = \partial_j v^i + \Gamma_j v^i + A_j v^i = \nabla_j v^i + A_j v^i \tag{2.42}$$

where here A_j are the components of a covector field. We are free to choose the A_j in specifying a connection on a line bundle. If one takes this action of ∇^A_j as axiomatic, then he need not concern himself with the previous vector bundle discussion.

Now given a connection ∇_A, we can also consider the curvature operator associated with A, denoted by $\nabla^2_A = \nabla_A \circ \nabla_A$. This operator acts on a section v like

$$\nabla^2_A v = \Omega v, \quad \text{where} \quad \Omega = dA + A \wedge A. \tag{2.43}$$

The matrix valued curvature two form $\mathcal{F}$ is given locally componentwise by

$$\mathcal{F}_{ij} = \partial_i A_j - \partial_j A_i + [A_i, A_j]. \tag{2.44}$$

In the case of the Levi-Civita connection, this produces the well known local coordinate formula for the Riemann curvature tensor of g (where we use the sign convention of Wald [165]),

$$R_{ijk}{}^l = \partial_j \Gamma^l_{ik} - \partial_i \Gamma^l_{jk} + \Gamma^a_{ik}\Gamma^l_{aj} - \Gamma^a_{jk}\Gamma^l_{ai} \tag{2.45}$$

Next we define the components of the Ricci curvature R_{ij} and the scalar curvature R by taking contractions of the Riemann tensor

$$R_{ij} = \sum_k R_{ikj}{}^k, \quad R = g^{ij} R_{ij}. \tag{2.46}$$

We need to use all of these quantities to compute approximations for PDE associated to different market models. We will mostly be interested in one and two dimensional models in which case the Riemann, Ricci, and scalar curvature tensors all have only one independent component which is trivial in one dimension or the Gauss curvature in two dimensions. In three dimensions, the Ricci tensor has six independent components, and the Riemann tensor is completely determined by the Ricci tensor. Only in four or higher dimensions does the Riemann tensor contain more information than Ricci, e.g. it has twenty-four independent components in dimension four whereas the Ricci tensor only has ten.

Another important geometric notion that we need to utilize is the concept of the distance function associated to a Riemannian metric. In particular, given two points $p, q \in M$, we define the distance function $d(p,q) : M \otimes M \to \mathbb{R}^+$. To do this, let $\mathcal{C}$ be the space of all C^1 curves joining p to q and for $\gamma \in \mathcal{C}$ denote the length functional $l : \mathcal{C} \to \mathbb{R}^+$ given by

$$l[\gamma] = \int_p^q \sqrt{g_{ij}\dot{\gamma}^i(t)\dot{\gamma}^j(t)}dt, \tag{2.47}$$

where here t is a coordinate that parameterizes γ by arclength. The distance function is then defined by

$$d(p,q) = \inf_{\gamma \in \mathcal{C}} \{l[\gamma]\}. \tag{2.48}$$

One can reduce the computation of the infimum to solving a system of ODEs by employing the standard variational calculus technique of demanding the first variation of l vanish, which is equivalent to solving the Euler-Lagrange equations associated with the length Lagrangian. The result is an associated system of differential equations for the components of the tangent vector to a geodesic joining p to q. The quasi-linear ODE system that one needs to solve to find the minimizing paths are the geodesic equations,

$$\ddot{\gamma}^i + \Gamma^i_{jk}\dot{\gamma}^j\dot{\gamma}^k = 0, \quad \gamma^i(0) = \gamma^i_0. \tag{2.49}$$

These equations generally form a coupled system of quasi-linear ODE and thus local existence and uniqueness of the system is guaranteed by standard ODE theorems. There is no global existence result for these equations. To understand why, note that on the sphere the are infinitely many geodesics joining any two antipodal points and hence uniqueness is violated.

There are cases where one can guarantee global existence/uniqueness of the geodesic equation. For instance, in the case (M, g) has negative sectional curvature and trivial fundamental group, if follows from the Hadamard-Cartan theorem Petersen [135, p.162] that geodesics are globally and uniquely defined on (M, g). Many cases of geometries we consider will satisfy this constraint, in particular the case of hyperbolic geometry.

We now turn our attention to discussing the relationship between SDEs and differential geometry.

2.5. **Finance/Geometry Relation.** We now review the construction of the pricing equation for a general path independent derivative whose payoff depends on n generic time homogeneous Itō processes. We then review how this equation can be related to a geometric heat equation, which will motivate the heat kernel expansion ansatz. Finally, we write the ansatz out explicitly in the one dimensional setting.

For $i = 1, \ldots, n$, let $x^i_t \equiv x^i(t)$ be n Itō processes which evolve according to a system of coupled stochastic differential equations (SDEs)

$$dx^i_t = \mu^i(x_t)dt + \sigma^i_j(x_t)dW^j, \quad x^i(0) = x^i_0, \tag{2.50}$$

where $j = 1, \ldots, n$, $x_t = \{x^1_t, \ldots, x^n_t\}$ denotes a function's dependence on potentially all the x^i_t, W^i are n Brownian motions with covariance matrix ρ^{ij}, i.e. $\mathbb{E}[dW^i dW^j] = \rho^{ij}dt$, and μ^i, σ^i_j are suitably regular functions. We assume these dynamics are risk neutral and that $\mathbb{E}$ is the expectation operator with respect to (w.r.t.) the risk neutral measure. We will also sometimes write $x^i = x^i_t$ for short when no contextual conflicts are present. The coefficient functions of this SDE system do not have explicit time dependence; we will always assume that this holds and note that the application of heat kernel expansions to models whose instantaneous volatility function depends explicitly on time remains an open area of research.

Let $F(x_t, t)$ be a pricing function for a contingent claim with a path independent payoff function $F(x_T, T)$ for some fixed $T > t > 0$. Then Itō's lemma requires that F evolves according to

$$dF = \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial x^i}dx^i + \frac{1}{2}\frac{\partial^2 F}{\partial x^i \partial x^j}dx^i dx^j, \tag{2.51}$$

where here we adopt the Einstein summation convention where repeated indices indicate implicit summation.

Moreover, since we take the above SDEs to be risk neutral, the discounted price process $e^{r\tau}F(x_t)$ (assuming a flat discounting curve), where $\tau = T - t$, is a martingale. Therefore

$$d(e^{r\tau}F(x_t, t)) = [e^{r\tau}F_t - re^{r\tau}F]dt + e^{r\tau}F_i dx^i + \frac{1}{2}e^{r\tau}F_{ij}dx^i dx^j \tag{2.52}$$

$$= e^{r\tau}\left[(F_t - rF)dt + F_i(\mu^i dt + \sigma^i_j dW^j) + \frac{1}{2}F_{ij}(\mu^i dt + \sigma^i_k dW^k)(\mu^j dt + \sigma^j_l dW^l)\right] \tag{2.53}$$

$$(2.54) \qquad = e^{r\tau}\left[F_t - rF + F_i\mu^i + \frac{1}{2}F_{ij}\sigma^i_k\sigma^j_l\rho^{kl}\right]dt + e^{r\tau}F_i\sigma^i_j dW^j,$$

where here we use subscripts on F to indicate partial differentiation with respect to the local coordinate functions, i.e. $F_i \equiv \partial_{x^i}F$. Now let $\Sigma^{ij} \equiv \sigma^i_k\sigma^j_l\rho^{kl}$ be a positive definite volatility matrix. Since the above process is a martingale, it must be driftless, and consequently F satisfies a backwards Fokker-Planck-Kolomogorov equation

$$(2.55) \qquad F_t - rF + \mu^i\frac{\partial F}{\partial x^i} + \frac{1}{2}\Sigma^{ij}\frac{\partial^2 F}{\partial x^i \partial x^j} = 0.$$

Alternatively, since we are restricting to the class of path independent payoff functions, given the values of the underlying processes x_t at time t and a filtration $\mathcal{F}_t$, we can represent F according to

$$(2.56) \qquad F(x_t) = \mathbb{E}\left[e^{r\tau}F(x_T)|\mathcal{F}_t\right] = e^{r\tau}\int_{\mathbb{R}} F(x_T)\phi(T, x_T|t, x_t)dx_T,$$

where here $\phi(T, x_T|t, x_t)$ is the joint transition density of the underlying processes x^i, i.e. it represents the probability that x will evolve to $x = x_T$ at time t given that the x^i had initial values x_t at time t.

Next, we want to construct a partial differential equation (PDE) for ϕ by computing F_t, F_i, and F_{ij} and substituting the results into equation (2.56). Differentiating with respect to t, we find that

$$(2.57) \qquad F_t(x_t) = \left[e^{r\tau}\int_{\mathbb{R}} F(x_T)\phi_t(x_T)dx_T - re^{r\tau}\int_{\mathbb{R}} F(x_T)\phi(x_T)dx_T\right]$$

$$(2.58) \qquad = e^{r\tau}\int_{\mathbb{R}} F(x_T)\phi_t(x_T)dx_T - rF(x_t).$$

We denote the spatial coordinates by $x^i = x^i_t$ so that the pricing PDE takes the form

$$(2.59) \qquad F_t - rF + \mu^i\frac{\partial F}{\partial x^i_t} + \frac{1}{2}\Sigma^{ij}\frac{\partial^2 F}{\partial x^i_t \partial x^j_t} = 0.$$

The gradient and Hessian of F just act on ϕ; after substituting them into equation (2.59), we find

$$(2.60) \qquad 0 = e^{r\tau}\int_{\mathbb{R}}\left[\phi_t(x_T) + \mu^i\frac{\partial\phi(x_T)}{\partial x^i_t} + \frac{1}{2}\Sigma^{ij}\frac{\partial^2\phi(x_T)}{\partial x^i_t\partial x^j_t}\right]F(x_T)dx_T.$$

This must hold for any payoff function F, so it is equivalent to a backwards parabolic PDE for the probability transition density

$$(2.61) \qquad 0 = \frac{\partial\phi}{\partial t} + \mu^i\frac{\partial\phi}{\partial x^i_t} + \frac{1}{2}\Sigma^{ij}\frac{\partial^2\phi}{\partial x^i_t\partial x^j_t}, \quad \phi(T-t, x, x_t) = \delta(x - x_t),$$

which is also subject to the spatial density boundary condition $\phi(x) \to 0$ as $|x_t - x^i| \to \infty$ (c.f Evans [54] for background on parabolic equations). Here $\delta(x - x_t)$ is the delta function on $\mathbb{R}^n$ centered at the point x_t and expressed in Euclidean coordinates.

One can convert this PDE into a forward equation by expressing it in terms of τ,

$$(2.62) \qquad \frac{\partial\phi}{\partial\tau} = \mu^i\frac{\partial\phi}{\partial x^i_t} + \frac{1}{2}\Sigma^{ij}\frac{\partial^2\phi}{\partial x^i_t\partial x^j_t}, \quad \phi(0, x, x_t) = \delta(x - x_t).$$

We seek to approximate solutions to equations of this form using heat kernel perturbation theory. In particular, we make an ansatz for ϕ which solves this equation exactly to zeroth order in hope of allowing us to simplify subsequent perturbative computations for higher order correction terms. There are many references for this construction (see e.g. Arnol'd [10], Avramidi [12], Labordere [98], Paulot [134]). The reader uninterested with the details of the construction may skip to equation (2.95) where we summarize the one dimensional form of the heat kernel ansatz that will be used later in our examples.

In order to motivate the heat kernel ansatz, we first review a correspondence between elliptic operators on $\mathbb{R}^n$ and connections on line bundles over a Riemannian manifold (M, g); here M is a C^∞ manifold (which can roughly be interpreted as a smooth subset of a Euclidean space) and g is a smooth set of symmetric positive definite matrices, indexed by points in M, which contains all necessary information related to computing distances on M. Consider an elliptic operator

$$L = \frac{1}{2}\Sigma^{ij}\partial_i\partial_j + \mu^i\partial_i, \tag{2.63}$$

on $\mathbb{R}^n$. We can represent L by an equivalent operator of the form

$$L = \Delta^A - Q = g^{ij}\nabla_i^A\nabla_j^A - Q, \tag{2.64}$$

on a line bundle $\mathcal{L}$ over M where here ∇_i^A is a connection which can be decomposed as $\nabla_i^A = \nabla_i + A_i$ where ∇_i is the Levi-Civita connection associated to g and A_i are the components of a real-valued section of the cotangent bundle of M, i.e. $A \in \Gamma(T^*M)$, and Q is a section of $\mathrm{End}\mathcal{L} \approx \mathcal{L} \otimes \mathcal{L}^*$, (c.f. Avramidi [12]). All of our analysis will be local, and we will only require the fact that ∇_i^A acts on a function $p : M \to \mathbb{R}$ according to

$$\nabla_i^A p = (\nabla_i + A_i)p = \nabla_i p + A_i p, \tag{2.65}$$

and on the components v_j of a covector $v \in T^*(M)$ like

$$\nabla_i^A v_j = \partial_i v_j - \Gamma_{ij}^k v_k, \quad \text{where} \quad \Gamma_{ij}^k = \frac{1}{2}g^{km}(\partial_i g_{kj} + \partial_k g_{ij} - \partial_k g_{ij}), \tag{2.66}$$

are the Christoffel symbols associated with g. In particular, in local coordinates, we can compute

$$Lp = g^{ij}(\nabla_i + A_i)(\nabla_j + A_j)p - Qp = g^{ij}(\nabla_i + A_i)(\partial_j p + A_j p) - Qp \tag{2.67}$$

$$= g^{ij}\left[\partial_i\partial_j p - \Gamma_{ij}^k\partial_k p + p(\partial_i A_j - \Gamma_{ij}^k A_k) + A_j p_i + A_i p_j + A_i A_j p\right] - Qp \tag{2.68}$$

$$= g^{ij}\left[\partial_i\partial_j p + 2A_j p_i - \Gamma_{ij}^k p_k + \left(\partial_i A_j - \Gamma_{ij}^k A_k + A_i A_j\right)p\right] - Qp. \tag{2.69}$$

We now can identity the diffusion and advection terms of this expression with those of equation (2.63) to find that

$$\frac{1}{2}\Sigma^{ij} = g^{ij}, \quad \mu^i = 2g^{ij}A_j - g^{jk}\Gamma_{jk}^i, \quad 0 = g^{ij}\left(\partial_i A_j - \Gamma_{ij}^k A_k + A_i A_j\right) - Q. \tag{2.70}$$

Through these equations, we can express an elliptic operator either by a choice of (Σ^{ij}, μ_i), or equivalently, by specifying a triple (g^{ij}, A_i, Q). Specifically, they can be inverted in order to write the geometric quantities in terms of the financial ones,

$$g^{ij} = \frac{1}{2}\Sigma^{ij}, \quad A_k = \frac{1}{2}\left[g_{ik}\mu^i + g_{ik}g^{jm}\Gamma_{jm}^i\right], \quad Q = g^{ij}(\partial_i A_j - \Gamma_{ij}^k A_k + A_i A_j). \tag{2.71}$$

We now give two operator identities that will prove useful later. The first concerns the A-Laplacian Δ^A defined by,

$$\Delta^A p \equiv g^{ij}(\nabla_i + A_i)(\nabla_j + A_j)p = g^{ij}(\nabla_i\nabla_j p + \nabla_i(A_j p) + A_i\nabla_j p + A_i A_j p) \tag{2.72}$$

$$= \Delta_g p + g^{ij}(A_j\partial_i p + p(\partial_i A_j - \Gamma_{ij}^k A_k) + A_i\partial_j p + A_i A_j p) \tag{2.73}$$

$$= \Delta_g p + 2g^{ij}A_i\partial_j p + g^{ij}[\partial_i A_j - \Gamma^k_{ij}A_k + A_iA_j]p = \Delta_g p + 2g^{ij}A_i\partial_j p + Qp, \tag{2.74}$$

which can be expressed more concisely as

$$(\Delta^A - Q)p = \Delta_g p + 2g^{ij}A_i\partial_j p. \tag{2.75}$$

The second identity involves the Levi-Civita Laplacian (Laplace-Beltrami operator) Δ_g and is expressed in local coordinates by

$$\Delta_g p = \frac{1}{\sqrt{g}}\partial_i\left(\sqrt{g}g^{ij}\partial_j p\right) = g^{ij}\left(\partial_i\partial_j p - \Gamma^k_{ik}\partial_j p\right). \tag{2.76}$$

where here $\sqrt{g}$ is the determinant of the metric.

2.6. **Heat Kernel Expansion Formula.** With the above tools at hand, we now can construct the heat kernel ansatz. First, consider a general second order elliptic differential operator on $\mathbb{R}^n$ expressed in terms of ∇^A_i given by

$$L\phi = g^{ij}\nabla^A_i\nabla^A_j\phi - Q\phi = \frac{1}{\sqrt{g}}(\partial_i + A_i)\left[\sqrt{g}g^{ij}(\partial_j + A_j)\phi\right] - Q\phi. \tag{2.77}$$

The heat equation $\partial_\tau\phi = L\phi$ can be solved exactly to zeroth order in τ. We assume that ϕ is given by this zeroth other solution multiplied by an arbitrary function Ω, where we must have $\Omega(0, x, x') = 1$ for consistency. The resulting expression for ϕ is called the heat kernel expansion and for $x, x' \in \mathbb{R}^n$, is given by

$$\phi(\tau, x, x') = \frac{\sqrt{g(x')}}{(4\pi\tau)^{n/2}}\mathcal{P}(x', x)\Delta^{1/2}(x, x')\exp\left(-\frac{d^2(x, x')}{4\tau}\right)\Omega(\tau, x, x'), \tag{2.78}$$

where here $d(x, x')$ is the distance function from x to x' associated with the metric g, and

$$\mathcal{P}(x', x) = \exp\left(-\int_C A\right) = \exp\left(-\int_{x'}^{x} A_i dx^i\right). \tag{2.79}$$

Here C is a minimizing oriented geodesic from $x' = x(0)$ to $x = x(t)$ parametrized by arclength. In the one dimensional setting, such a geodesic always exists, although this issue is more subtle in higher dimensions c.f. Forde [58]. Finally,

$$\Delta(x', x) = \frac{1}{\sqrt{g(x)g(x')}}\det\left(-\frac{\partial^2 d^2(x, x')}{\partial x_i \partial x'_j}\right), \tag{2.80}$$

is known as the van-Vleck-Morette determinant. If we substitute this ansatz into the previous equation, then after simplification, we find that Ω must satisfy

$$\left(\frac{\partial}{\partial\tau} + \frac{1}{\tau}(\nabla^i\sigma)\nabla_i - \mathcal{P}^{-1}\Delta^{-1/2}L\Delta^{1/2}\mathcal{P}\right)\Omega(\tau, x, x') = 0, \tag{2.81}$$

with initial condition $\Omega(0, x, x') = 1$. Now assume that Ω is given by a formal power series in τ,

$$\Omega(\tau, x, x') = \sum_{k=0}^{\infty} a_k(x', x)\tau^k. \tag{2.82}$$

Next, let $\sigma(x, x') = d(x, x')^2/2$, and substitute Ω into equation (2.81) to find

$$0 = \sum_{k=0}^{\infty} k a_k \tau^{k-1} + \sum_{k=0}^{\infty}[(\nabla^i\sigma)\nabla_i a_k]\tau^{k-1} - \sum_{k=0}^{\infty}\mathcal{P}^{-1}\Delta^{-1/2}L\Delta^{1/2}\mathcal{P}a_k\tau^k \tag{2.83}$$

$$= \sum_{k=1}^{\infty} k a_k \tau^{k-1} + \sum_{k=0}^{\infty} [(\nabla^i \sigma)\nabla_i a_k]\tau^{k-1} - \sum_{k=1}^{\infty} \mathcal{P}^{-1}\Delta^{-1/2} L \Delta^{1/2} \mathcal{P} a_{k-1}\tau^{k-1} \tag{2.84}$$

$$= (\nabla^i \sigma)(\nabla_i a_0)\tau^{-1} + \sum_{k=1}^{\infty} [k a_k + (\nabla^i \sigma)\nabla_i a_k - \mathcal{P}^{-1}\Delta^{-1/2} L \Delta^{1/2} \mathcal{P} a_{k-1}]\tau^{k-1}. \tag{2.85}$$

Now the coefficients of the different powers of the τ^i must vanish identically. The initial condition $\Omega(0, x, x') = 1$ together with $\nabla_i a_0 = 0$ require that $a_0 = 1$. We find that the rest of the a_k are given by a recursive hierarchy of differential equations

$$k a_k + d\partial_d a_k - \mathcal{P}^{-1}\Delta^{-1/2} L \Delta^{1/2} \mathcal{P} a_{k-1} = 0, \tag{2.86}$$

where here we use $\nabla^i \sigma = d\nabla^i d$ and $(\nabla^i \sigma)\nabla_i = d\nabla_{\nabla d} \equiv d\partial_d$. One can integrate this system to find an iterative formula for the a_k,

$$a_k(x', x) = \frac{1}{d^k} \int_C d^{k-1} \mathcal{P}^{-1}(x', x)\Delta^{-1/2} L \Delta^{1/2} \mathcal{P}(x', x) a_{k-1}, \quad k \geq 1. \tag{2.87}$$

The goal of heat kernel perturbation theory is to attempt to evaluate or approximate (usually by the tractable diagonal coefficients $a_k(x, x)$) the a_k integrals in a manner such that the resulting explicit form for ϕ approximates the true solution of equation (2.63) to a high degree of accuracy for a desired domain of model parameters. Computing a_k exactly is generally only possible in the simplest geometries (M, g) for dimensions $n \geq 2$. However, when $n = 1$, these reduce to integrals over $\mathbb{R}$ and are calculable in a wide variety of models.

We now restrict our attention to the case of one dimensional models. Specifically, we consider a driftless local volatility model of the form

$$dS_t = C(S_t)dW, \quad S(0) = S_0, \tag{2.88}$$

where we take $C : \mathbb{R}^+ \to \mathbb{R}$ to be at least a C^2 function and uniformily positive. The transition density equation for this model is given by

$$\phi_\tau = \frac{1}{2} C(\alpha)^2 \phi_{\alpha\alpha}. \tag{2.89}$$

Applying the PDE/geometry correspondence, note that the single component of the inverse metric is given by $g^{\alpha\alpha} = C(\alpha)^2/2$. Thus the metric and the square root of its determinant are just $g_{\alpha\alpha} = 2/C(\alpha)^2$ and $\sqrt{g(\alpha)} = \sqrt{2}/C(\alpha)$. Using this, we can compute the single Christoffel symbol

$$\Gamma^\alpha_{\alpha\alpha} = \frac{1}{2} g^{\alpha\alpha}\partial_\alpha g_{\alpha\alpha} = -\frac{1}{C(\alpha)}\frac{\partial C(\alpha)}{\partial \alpha}, \tag{2.90}$$

which is just the component of the one form $-d\ln C(\alpha)$. Next, we find $A_\alpha = \frac{1}{2} g_{\alpha\alpha} g^{\alpha\alpha}\Gamma^\alpha_{\alpha\alpha} = \frac{1}{2}\Gamma^\alpha_{\alpha\alpha}$ and

$$\mathcal{P}(\alpha, S) = \exp\left(-\int_\alpha^S A_\alpha d\alpha\right) = \exp\left(\int_\alpha^S d\ln\sqrt{C(\alpha)}\right). \tag{2.91}$$

Evaluating the integral, we find that $\mathcal{P}(\alpha, S) = \sqrt{C(S)/C(\alpha)}$, which in turn implies $\mathcal{P}(S, \alpha) = \sqrt{C(\alpha)/C(S)}$. We can further compute

$$\mathcal{P}^{-1}\Delta_g \mathcal{P} + 2\mathcal{P}^{-1} g^{\alpha\alpha} A_\alpha \partial_\alpha \mathcal{P} = \frac{1}{8}\left(2CC'' - (C')^2\right). \tag{2.92}$$

Next, define a coordinate $s = \sqrt{2}\int_{\alpha}^{S} \frac{du}{C(u)}$ which parameterizes geodesics on $(\mathbb{R}, g)$ by arclength. Changing coordinates allows us to see that

$$g_{\alpha\alpha} = \left(\frac{\partial s}{\partial \alpha}\right)^2 g_{ss} = \frac{2}{C(\alpha)^2} g_{ss} \tag{2.93}$$

from which we note that the line element is given by

$$ds^2 = g_{\alpha\alpha} d\alpha^2, \tag{2.94}$$

i.e. in the s coordinate g is just the standard Euclidean metric on $\mathbb{R}$. In particular, this implies that $\Delta(S, \alpha) = 1$ which considerably simplifies the computation of the a_k.

We now summarize the the heat kernel ansatz in one dimension in a form that is expressed solely in terms of operators and functions on $\mathbb{R}$, namely,

$$\phi(\tau, \alpha, S) = \sqrt{\frac{C(\alpha)}{2\pi\tau C(S)^3}} \exp\left(-\frac{d^2(S, \alpha)}{4\tau}\right) \sum_{k=1}^{\infty} a_k(S, \alpha)\tau^k, \tag{2.95}$$

where the distance function d is

$$d(S, \alpha) = \sqrt{2}\int_{\alpha}^{S} \frac{du}{C(u)}. \tag{2.96}$$

Actually, the true distance function associated to g is given by taking the absolute value of the above; however we omit this off for simplicity as the sign of d will not effect the ansatz. The a_k are given by the integrals

$$a_k(S, \alpha) = -\frac{1}{d^k}\int_{S}^{\alpha} d^{k-1}\Big[\mathcal{P}(S, \alpha)^{-1}\partial_\alpha\left(\sqrt{g(\alpha)}g^{\alpha\alpha}\partial_\alpha\left(\mathcal{P}(S, \alpha)a_{k-1}(S, \alpha)\right)\right) \tag{2.97}$$

$$+2\mathcal{P}(S, \alpha)^{-1}\sqrt{g(\alpha)}g^{\alpha\alpha}A_\alpha\partial_\alpha\left(\mathcal{P}(S, \alpha)a_2(S, \alpha)\right)\Big]d\alpha. \tag{2.98}$$

We can represent the first heat kernel coefficient in the following convenient way,

$$a_1(S, \alpha) = -\frac{\sqrt{2}}{4d}\int_{S}^{\alpha}\left[C''(u) - \frac{(C'(u))^2}{2C(u)}\right]du. \tag{2.99}$$

When it is possible evaluate this integral, which can be done for a wide range of functions C, then one can insert the result into the a_k formula and attempt to compute a_2. Although one can typically compute the a_1 integral exactly, computing higher a_k is a potentially a difficult task. If we are able to compute a_k for a given local volatility model, we will say that we have constructed a k-th order approximation formula. We now turn to several examples starting with the CEV model.

Before proceeding to applications of this approximation, now briefly comment on how we will use the heat kernel expansion formula for purposes of transition density estimation and derivative pricing. If x^i are n stochastic processes with initial values x_0^i, then we have seen that the transition probability density $p(x^i, x_0^i)$ is determined by solving a second order linear parabolic equation with delta function initial data. We will use the heat kernel expansion to estimate the solution of this PDE initial value problem. We can then use an expression for the transition density to price a generic derivative by integrating the derivative's payoff function against the transition density. In some instances, evaluation of this integral is not possible, and we can turn to a Laplace/steepest decent approximation.

3. Local Volatility Model Applications

We now consider heat kernel expansions in the context of one-dimensional local volatility models. The underlying geometry of these models will always be Euclidean when we express the model dynamics in an appropriate local coordinate system. Thus the distance function is easy to determine and the Van-Vleck determinant will always be unity. We can even calculate the a_i heat kernel series coefficient functions exactly in the case of several one-dimensional models. We then comment on the errors associated with this approximation procedure.

We start by considering the Black-Scholes Merton model in order to illustrate these perturbative methods in a simple context. We then turn to the CEV and quadratic local volatility models. In each case, we are able to compute all the heat kernel coefficients. We then construct first and second order expansion formulas for the transition density for three new local volatility models which we call the cubic, affine-affine, and generalized CEV models.

3.1. **Black-Scholes-Merton Example.** We now consider the simple case of the one dimensional lognormal Black-Scholes-Merton (BSM) model [18], [119]. We can compute all quantities involved in the heat equation exactly in this model and in fact can even invert the series in the heat kernel expansion to produce the known analytic expression for the BSM transition density function thus providing an alternative way to solve the BSM equation (see Andereasen et. al [6] for a variety of ways to solve the BSM equation).

The risk neutral dynamics are given by

$$dS(t) = rS(t)dt + \sigma S(t)dW(t), \tag{3.1}$$

where r is the risk-free interest rate and $\sigma > 0$ is a volatility constant. Thus $\mu^S = rS$ and $\Sigma^{SS} = \sigma^2 S^2$. so that the transition probability density function $p(S, T, S_0, t) = p(\tau, S, S_0)$ must satisfy

$$\frac{\partial p}{\partial \tau} = rS_0 \frac{\partial p}{\partial S_0} + \frac{\sigma^2}{2} S_0^2 \frac{\partial^2 p}{\partial S_0^2}, \quad p(0, S_0, S) = \delta(S - S_0). \tag{3.2}$$

Now this equation can be solved exactly by making appropriate coordinate changes to reduce it to the heat equation or by using Fourier transform techniques and its solution is given by

$$p(S, T, S_0, t) = \frac{1}{S\sigma\sqrt{2\pi\tau}} \exp\left(-\frac{\left(\log(S/S_0) - (r - \sigma^2/2)\tau\right)^2}{2\sigma^2\tau} \right). \tag{3.3}$$

We now compute the quantities that go into the heat kernel expansion formula which approximates the solution of this PDE.

To ease notation, let $\alpha = S_0$. First, we know that the inverse metric is given by $g^{\alpha\alpha} = \frac{1}{2}\Sigma^{\alpha\alpha} = \frac{\sigma^2\alpha^2}{2}$, and hence $g_{\alpha\alpha} = 2/\sigma^2\alpha^2$. Now we define a new coordinate $s(\alpha) = \frac{\sqrt{2}}{\sigma}\int_\alpha^S d\ln u$ such that $\partial s/\partial\alpha = -\sqrt{2}/\alpha\sigma$. In this coordinate, the metric takes the form $g_{ss} = (\partial\alpha/\partial s)^2 g_{\alpha\alpha} = 1$, i.e. it is just the standard Euclidean metric and hence s is an arc-length coordinate; that is to say, s parameterizes geodesics on $(\mathbb{R}, g)$ by arclength.

We can thus compute the distance function associated to this metric immediately to be

$$d(x_1, x_2) = s(x_2) - s(x_1) = \frac{\sqrt{2}}{\sigma}(\ln x_2 - \ln x_1) = \frac{\sqrt{2}}{\sigma} \ln\left(\frac{x_2}{x_1}\right), \tag{3.4}$$

where we take $x_1 < x_2$. From the form of $d(s_0, s) = s - s_0$, it is immediate that $\Delta(s_0, s) = 1$. Since this does not depend on local coordinates, we have that $\Delta(\alpha, S) = 1$ as well. We finally need to compute $\mathcal{P}$ and a_α, for which we need to know A_α; we also compute Q:

$$A_\alpha = \frac{1}{2} g_{\alpha\alpha}\mu^\alpha + \frac{1}{2}\Gamma^\alpha_{\alpha\alpha} = \left(\frac{r}{\sigma^2} - \frac{1}{2}\right)\frac{1}{\alpha}, \tag{3.5}$$

$$Q = g^{\alpha\alpha}\left(\partial_\alpha A_\alpha - \Gamma^\alpha_{\alpha\alpha}A_\alpha + A_\alpha^2\right) = \frac{\sigma^2}{2}\left(\frac{r}{\sigma^2} - \frac{1}{2}\right)^2, \tag{3.6}$$

where we use the fact that $\Gamma^\alpha_{\alpha\alpha} = \frac{1}{2}g^{\alpha\alpha}\partial_\alpha g_{\alpha\alpha} = -1/\alpha$. Using this we can compute

$$\mathcal{P}(\alpha, S) = \exp\left(-\int_\alpha^S A_\alpha d\alpha\right) = \exp\left(-\left(\frac{r}{\sigma^2} - \frac{1}{2}\right)\int_\alpha^S \frac{1}{\alpha}d\alpha\right) = \left(\frac{S}{\alpha}\right)^{\frac{1}{2} - \frac{r}{\sigma^2}}. \tag{3.7}$$

where we used the formula

$$\mathcal{P}(x', x) = \exp\left(-\int_{C(x',x)} A_i dx^i\right), \tag{3.8}$$

where $C(x', x)$ is the arc length parametrized oriented geodesic starting at x' and ending at x. Note that the ordering on the arguments of $\mathcal{P}$ is important. It turns out that we get the same result if we compute $a_1(S, \alpha)$ instead of $a_1(\alpha, S)$ in the BSM case, however in more complicated models this does not hold. We will stick to the convention that the first argument of the a_k should always be the spatial variables of the relevant probability density PDE and the second should be an initial data constant.

Now we are interested in solving the BSM PDE

$$p_\tau = r\alpha p_\alpha + \frac{\sigma^2}{2}\alpha^2 p_{\alpha\alpha} = L_\alpha p(\tau, \alpha, x), \tag{3.9}$$

where we denote the elliptic part of the BSM density operator by L_α to illustrate that α is the spatial variable of the operator.

We have to "reverse" the arguments of our previous heat kernel formula to compensate for the fact that α is the initial asset price and find that the relevant heat kernel expansion formula is given by

$$p(\tau, \alpha, S) = \frac{\sqrt{g(S)}}{(4\pi\tau)^{n/2}}\mathcal{P}(S, \alpha)\Delta^{1/2}\exp\left(-\frac{d^2(S, \alpha)}{4}\right)\sum_{k=0}^\infty a_k(S, \alpha)\tau^k. \tag{3.10}$$

Now note that

$$\mathcal{P}(S, \alpha) = \exp\left(-\int_S^\alpha A_\alpha d\alpha\right) = \exp\left(\int_\alpha^S A_\alpha d\alpha\right) = \mathcal{P}(\alpha, S)^{-1}. \tag{3.11}$$

Now we want to compute

$$a_1(S, \alpha) = \frac{1}{d}\int_C \mathcal{P}^{-1}(S, \alpha)(\Delta^A - Q)\mathcal{P}(S, \alpha)ds = \frac{1}{d}\int_C \left[\mathcal{P}^{-1}\Delta_g\mathcal{P} + 2\mathcal{P}^{-1}g^{\alpha\alpha}A_\alpha\partial_\alpha\mathcal{P}\right]ds, \tag{3.12}$$

where here C is the geodesic that joins α to S and $\mathcal{P} = \mathcal{P}(S, \alpha)$ in the above. We now compute

$$\mathcal{P}^{-1}(S, \alpha)\Delta_g\mathcal{P}(S, \alpha) = \frac{\mathcal{P}^{-1}}{\sqrt{g}}\partial_\alpha\left(\sqrt{g}g^{\alpha\alpha}\partial_\alpha\mathcal{P}\right) = \frac{\sigma^2}{2}\left(\frac{1}{2} - \frac{r}{\sigma^2}\right)^2 = Q. \tag{3.13}$$

Next we compute

$$2\mathcal{P}^{-1}(S, \alpha)g^{\alpha\alpha}A_\alpha\partial_\alpha\mathcal{P}(S, \alpha) = -\sigma^2\left(\frac{1}{2} - \frac{r}{\sigma^2}\right)^2 = -2Q. \tag{3.14}$$

which when combined with the previous yields

$$a_1(S, \alpha) = \frac{1}{d}\int_0^d [Q - 2Q]ds = \frac{1}{d}\int_S^\alpha (-Q)\left(-\frac{\sqrt{2}}{\alpha\sigma}\right)d\alpha = -Q. \tag{3.15}$$

Similarly, we can use the formula for a_k to find that

$$a_k(S,\alpha) = \frac{1}{d^k}\int_0^d s^{k-1}\frac{(-Q)^k}{(k-1)!}ds = \frac{(-Q)^k}{k!} = \frac{(-1)^k(2r-\sigma^2)^{2k}}{k!2^{3k}\sigma^{2k}}. \tag{3.16}$$

Thus we can sum the a_k exactly by recognizing them as the Taylor series coefficients of an exponential function

$$\sum_{k=0}^{\infty} a_k\tau^k = \sum_{k=0}^{\infty}\frac{(-Q\tau)^k}{k!} = e^{-Q\tau}. \tag{3.17}$$

We can now put these pieces together in our heat kernel formula to find that the density function is given by

$$p(\tau,\alpha,S) = \frac{\sqrt{g(S)}}{\sqrt{4\pi\tau}}\mathcal{P}(S,\alpha)\exp\left(-\frac{d(\alpha,S)^2}{4\tau}\right)\sum_k a_k\tau^k \tag{3.18}$$

$$= \frac{1}{\sigma S\sqrt{2\pi\tau}}\exp\left(\left(\frac{r}{\sigma^2}-\frac{1}{2}\right)\ln\frac{S}{\alpha}\right)\exp\left(-\frac{\ln(S/\alpha)^2}{2\sigma^2\tau}\right)e^{-Q\tau} \tag{3.19}$$

$$= \frac{1}{S\sigma\sqrt{2\pi\tau}}\exp\left(-\frac{\left[\ln(S/\alpha)-(r-\sigma^2/2)\tau\right]^2}{2\sigma^2\tau}\right) \tag{3.20}$$

which is exactly the same transition density function that we found before. We now consider examples to determine how accurate different order approximations are compared to the exact solution of the model in order to build intuition for more complex models to come. In order to achieve this we need to define

$$p_n(\tau,\alpha,S) = \frac{1}{S\sigma\sqrt{2\pi\tau}}\exp\left(\left(\frac{r}{\sigma^2}-\frac{1}{2}\right)\ln\left(\frac{S}{\alpha}\right)-\frac{\ln^2(S/\alpha)}{2\sigma^2\tau}\right)\sum_{k=0}^{n}\frac{(-Q)^k}{k!}\tau^k. \tag{3.21}$$

We first note that when one considers graphs of the p_n for typical market values of the parameters σ and r, i.e. $\sigma \in [0,0.5]$ and $r \in [0,0.1]$, it is hard to distinguish the graph of p_0 from the exact solution for standard maturity times unless one considers minuscule codomain scales due to the relatively small size of the correction terms in the perturbation series (or the correspondingly highly accurate nature of the leading terms of the approximation). We thus first consider a plot using a reasonable interest rate value of $r = 0.05$ along with a high volatility $\sigma = 0.8$. In addition, we take a unit initial asset price $S_0 = 1$ and $\tau = 1$ and plot p_0 through p_4 along with the exact solution in Figure 1. There only appear to be two graphs (even though we are plotting six curves) in the figure. The zeroth order correction p_0 is represented by the greatest graph and $p_1,\ldots,p_4$ as well as the exact solution are depicted in the second graph; if one zooms in very closely, he can see a where a graph which corresponds to the first order approximation differs from the higher order graphs. Thus the p_i, for $i > 1$, approximate the exact solution very accurately to the point they are indistinguishable in our plot.

There are a few notable statistical features of these density functions. First we note that if we let $I_n = \int_{\mathbb{R}^+} p_n dS$, then we can numerically integrate to find that $I_0 = 1.05861$, $I_1 = 0.998315$, $I_2 = 1.00003$, and $I_3 = 1.00000$. The I_n give an estimate of how fast the p_n are converging to a valid density function, since such a function must integrate to one. Now let X be the random variable that has a density function given by p_∞, i.e. the exact solution, and let X_n be random variables which have p_n as psuedo-density functions (We use the qualifier "pseudo" here since none of the p_n can in fact be density functions since they do not integrate to unity. We do not normalize the p_n to fix this). We first note that

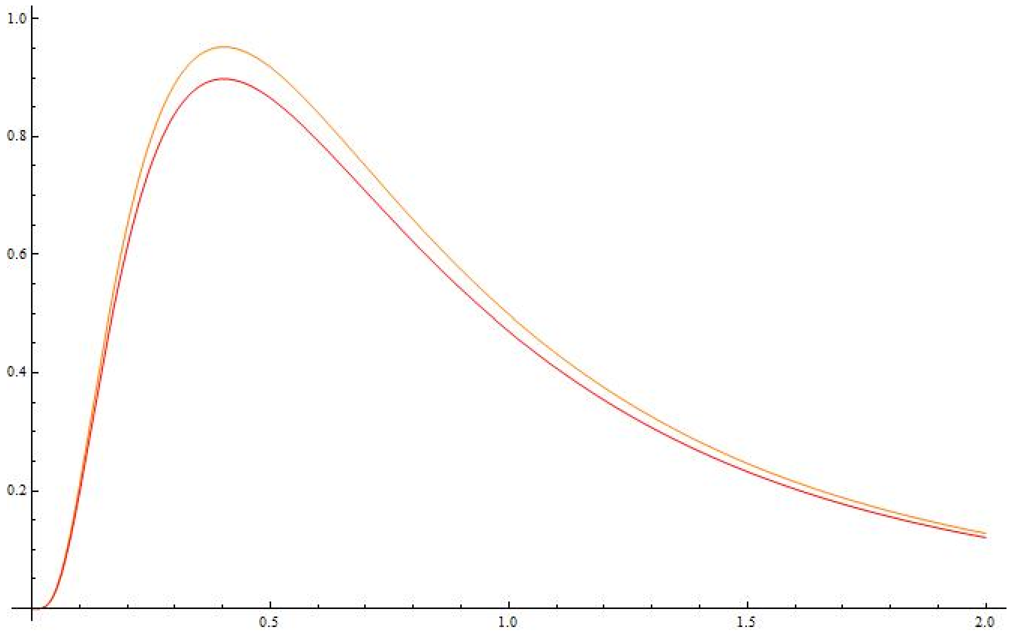

FIGURE 1. This is a plot of $p_0, \ldots, p_4$ from equation (3.21) and the exact density p from equation (3.3) for $\alpha = 1$, $\tau = 1$, $\sigma = 0.8$, and $r = 0.05$.

$\mathbb{E}X_0 = 1.11288$ which overestimates the mean of X which again is expected by inspection of the graph plots. Next, we find that $\mathbb{E}X_1 = 1.0495$, which is now an underestimate. This is due to the fact that the terms in the heat kernel expansion formula alternate in sign and thus we expect an alternating over/under-estimating of the mean. This is confirmed by further considering $\mathbb{E}X_2 = 1.05130$ and $\mathbb{E}X_3 = 1.05127$. Then we find to five decimal places, all the higher p_n have the same value as $\mathbb{E}X_3$, so $\mathbb{E}X = 1.05127$ which is a reasonable expected value given that we have specified the risk neutral dynamics at a five percent annual risk free rate.

The over/under estimation pattern is brought out more clearly in our next graph. In Figure 2, we again plot $p_0, \ldots, p_4$ together with the exact solution, but this time take an unusually high interest rate $r = 0.4$ together with a standard volatility $\sigma = 0.3$ while still keeping $f_0 = \tau = 1$.

Here p_0 is the greatest curve, p_1 the least, p_2 the second greatest, p_3 the second least, and p_4 as well as the exact solution plotted in the middle and cannot be distinguished. Here we find $\mathbb{E}X = 1.49182$.

We note that our sequence of psuedo-densities p_n converges uniformly to the exact density at every point in their domains. We will see in our next example that this is not the case for more general models.

Finally, we note that we can tell from the form of (3.16) how well the approximation will work for different values of the model parameters (r, σ). In particular, note that

$$|a_k| = \frac{(2r - \sigma^2)^{2k}}{k! 2^{3k} \sigma^{2k}}. \tag{3.22}$$

Thus when $2r \approx \sigma^2$ we see the higher order correction terms become small very quickly. Note that the a_k are generically small for reasonable choices of interest rates and volatilities. Also if one would apply this model in a highly volatile market where interest rates are correspondingly high, i.e. $\sigma = 0.5$ and $r = 0.12$, we would still have a very accuracte approximation even to low order in the a_i. Lastly, note that one can apply the ratio test to the a_i to see that the series $\sum_i a_i$ is indeed convergent. We will see in the next section that

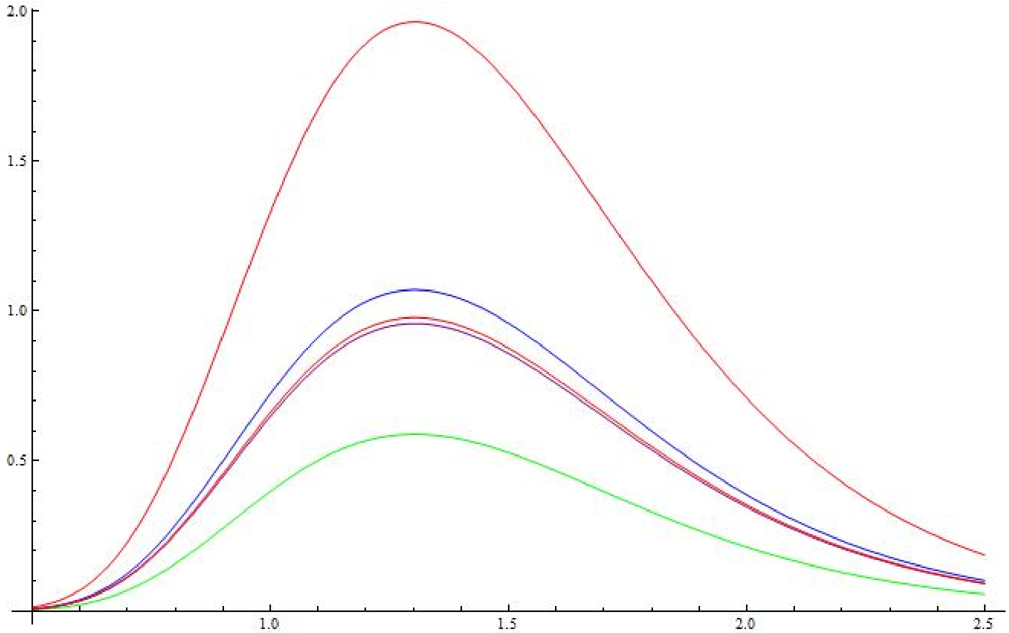

FIGURE 2. This is a plot of $p_0, \ldots, p_4$ from equation (3.21) and the exact density p from equation (3.3) for $\alpha = 1$, $\tau = 1$, $\sigma = 0.3$, and $r = 0.4$.

this is a special property of BSM dynamics that does not hold in the CEV setting where the associated a_i series is in fact always divergent except when it reduces to the BSM case.

We finally consider a slightly more global representation of the error associated with this approximation in Figure 3. Here we fix $\alpha = 1$, $T = 1$, and $r = 0.05$ and plot the difference between our first order approximation and the exact solution, namely $p - p_1$, for small asset values $S \in [0, 2]$ and the full range of realistic volatility values $\sigma \in [0, 1]$.

Note the error is generically small except in the case where σ is large and S is small. There also is an spike in the error in a neighborhood around $(S, \sigma) = (1, 0.15)$. This spike appears somewhat generic to all terms, and the error in this region is significantly reduced by taking higher order approximations to the density.

3.2. **Constant Elasticity of Variance Model.** We now consider a generalization of the previous BSM model called the Constant Elasticity of Variance model (CEV) (see Brecher and Lindsay [22] for a survey). The CEV model is a local volatility model and is widely used in many financial contexts for smile modeling. The model dynamics are given by

$$dS = \sigma S^{\beta} dW, \quad S(0) = S_0, \tag{3.23}$$

where we fix $\beta \in (0, 1)$ and $\sigma > 0$. This model can be thought of as the natural interpolation between the normal Bachelier ($\beta \to 0$) and lognormal Black-Scholes-Merton ($\beta \to 1$) models. We note that when $\beta = 1$, one can compute the heat kernel coefficients a_k exactly and in fact is able to invert the heat kernel expansion series and recover the exact well known transition density for a lognormal process. Note that this SDE has no drift. The presence of a drift term significantly complicates the perturbation theory involved in computing an expansion formula. We will consider adding a mean reversion/drift term when we later investigate a generalized CEV model.

The solutions of the CEV SDE fall into two classes depending on whether $\beta \in (0, 1/2)$ or $\beta \in [1/2, 1)$. In the former case, it was shown by Feller [56] that the level $S = 0$ is attainable and one needs to specify whether this boundary is absorbing (meaning that the process remains trivial after hitting the zero level) or reflecting. If $\beta \in [0.5, 1)$, then the

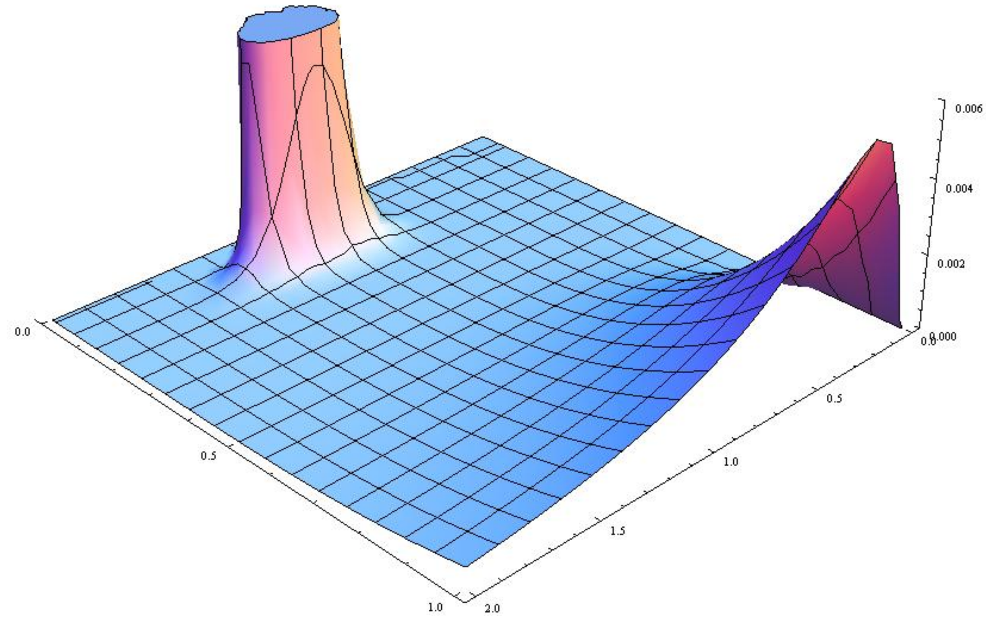

FIGURE 3. This is a first order absolute error plot of $p - p_1$ where p_1 is given by (3.21) and the exact transition density p by equation (3.3). Here $S \in [0, 2]$ and $\sigma \in [0, 1]$.

boundary is always absorbing. Also, if $\beta > 1$, the zero level is not attainable so there is no need to consider spatial boundary condition.

We now compute the geometric quantities that are relevant to this model. First we note $\Sigma^{\alpha\alpha} = \sigma^2\alpha^{2\beta}$ where $\alpha = S_0$ and $\mu^\alpha = 0$, so we wish to approximate the solution of the transition density PDE

$$p_\tau = \frac{1}{2}\sigma^2\alpha^{2\beta}p_{\alpha\alpha}, \quad p(0, \alpha, S) = \delta(S - \alpha). \tag{3.24}$$

Now in the case of an absorbing spatial boundary condition at the $S = 0$ level, this PDE has an exact solution given by

$$p(\tau, S, \alpha) = \frac{S^{\frac{1}{2}-2\beta}}{(1-\beta)\sigma^2\tau}\sqrt{\alpha}\exp\left(-\frac{S^{2(1-\beta)} + \alpha^{2(1-\beta)}}{2(1-\beta)^2\sigma^2\tau}\right) I_{\frac{1}{2(1-\beta)}}\left(\frac{(S\alpha)^{1-\beta}}{(1-\beta)^2\sigma^2\tau}\right), \tag{3.25}$$

where here $I_\nu(x)$ is the modified Bessel function of the first kind defined by

$$I_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{(x^2/4)^k}{k!\Gamma(\nu + k + 1)}, \tag{3.26}$$

and $\Gamma(x)$ is the standard interpolation of the factorial function, commonly called the Gamma function, and is defined for positive real values according to

$$\Gamma(x) = \int_{\mathbb{R}^+} u^{x-1}e^{-u}du. \tag{3.27}$$

One can verify this solution directly by substitution and use of the modified Bessel function identity

$$\frac{\partial I_\nu(x)}{\partial x} = \frac{1}{2}\left(I_{\nu-1}(x) + I_{\nu+1}(x)\right), \tag{3.28}$$

which can be iterated to compute the required second derivative formula.

Note that p is not analytic at the point $\tau = 0$. Also for certain parameter values, $\int pdS < 1$ due to the absorbing boundary condition, i.e. we lose probability mass. Stated another way, when p is integrated over $[0, \infty)$, the result is exactly one, whereas if it is integrated over $(0, \infty)$, the presence of a Dirac function in the density at the origin due to the boundary conditions will require the associated integral to be less than 1. In these cases, the true transition probability density can be constructed if one supplements this expression by adding an appropriately normalized delta function about the origin where the normalization constant is chosen so that the integral of the true density is one. However, when the volatility is small, the probability that any price path hits the zero level is also small and as a result $\int_{\mathbb{R}^+} p \approx 1$. For instance when $\alpha = 1$, $t = 10$, $\sigma = 0.3$, and $\beta = 0.6$, we can numerically integrate to find $\int_{\mathbb{R}^+} p = 0.950384$, and when we keep all the same parameters but lower the volatility to $\sigma = 0.2$, we find $\int_{\mathbb{R}^+} p = 0.999232$.

Now the inverse metric is given by $g^{\alpha\alpha} = \sigma^2\alpha^{2\beta}/2$, and thus $g_{\alpha\alpha} = 2/\sigma^2\alpha^{2\beta}$ and $\sqrt{g} = \sqrt{2}/\sigma\alpha^{\beta}$. Since there is no drift term in the SDE, we have

$$0 = 2g^{\alpha\alpha}A_\alpha - g^{\alpha\alpha}\Gamma^{\alpha}_{\alpha\alpha} \to A_\alpha = \frac{1}{2}\Gamma^{\alpha}_{\alpha\alpha}. \tag{3.29}$$

We next compute $\Gamma^{\alpha}_{\alpha\alpha} = -\beta/\alpha$, from which we find that $A_\alpha = -\beta/(2\alpha)$. Next, we note that

$$Q = g^{\alpha\alpha}\left(\partial_\alpha A_\alpha - \Gamma^{\alpha}_{\alpha\alpha}A_\alpha + A_\alpha^2\right) = -\frac{\sigma^2}{8}\beta(\beta-2)\alpha^{2\beta-2}. \tag{3.30}$$

Using this we find that

$$\mathcal{P}(\alpha, S) = \exp\left(-\int_\alpha^S A_\alpha d\alpha\right) = \exp\left(\frac{\beta}{2}\int_\alpha^S \frac{1}{\alpha}d\alpha\right) = \left(\frac{S}{\alpha}\right)^{\frac{\beta}{2}}, \tag{3.31}$$

which tells us that $\mathcal{P}(S, \alpha) = (\alpha/S)^{\beta/2}$. From this we can calculate

$$\mathcal{P}^{-1}(S,\alpha)\Delta_g\mathcal{P}(S,\alpha) = \frac{\sigma^2\beta(3\beta-2)}{8}\alpha^{2(\beta-1)}, \tag{3.32}$$

and also

$$2\mathcal{P}^{-1}g^{\alpha\alpha}A_\alpha\partial_\alpha\mathcal{P} = -\frac{\sigma^2\beta^2}{4}\alpha^{2(\beta-1)}, \tag{3.33}$$

so that the sum is just given by

$$\mathcal{P}^{-1}\Delta_g\mathcal{P} + 2\mathcal{P}^{-1}g^{\alpha\alpha}A_\alpha\partial_\alpha\mathcal{P} = \frac{\sigma^2\beta(\beta-2)}{8}\alpha^{2(\beta-1)}. \tag{3.34}$$

In order to use this to construct $a_1(S, \alpha)$, we need to find an arclength parameter for the geodesic joining S to α. We define

$$s(\alpha) = \frac{\sqrt{2}}{\sigma}\int_\alpha^S \frac{du}{u^\beta} = \frac{\sqrt{2}}{\sigma(1-\beta)}\left[S^{1-\beta} - \alpha^{1-\beta}\right]. \tag{3.35}$$

Note that $\partial s/\partial\alpha = -\sqrt{2}/\sigma\alpha^\beta$, so that $g_{ss} = (\partial\alpha/\partial s)^2 g_{\alpha\alpha} = 1$. Thus this coordinate induces a distance function

$$d(x_2, x_1) = \frac{\sqrt{2}}{\sigma(1-\beta)}[x_2^{1-\beta} - x_1^{1-\beta}]. \tag{3.36}$$

Also note that

$$ds = \frac{\partial s}{\partial\alpha}d\alpha = -\frac{\sqrt{2}}{\sigma\alpha^\beta}d\alpha, \tag{3.37}$$

and that $s(S) = 0$ and $s(\alpha) = d(S, \alpha)$. Now

(3.38)
$$a_1(S,\alpha) = \frac{\beta(\beta-2)\sigma^2}{8d}\int_0^d \alpha^{2(\beta-1)}ds = -\frac{\sqrt{2}\beta\sigma(\beta-2)}{8d}\int_S^\alpha \alpha^{\beta-2}d\alpha = \frac{\beta(\beta-2)\sigma^2}{8}(S\alpha)^{\beta-1}.$$

Next we can compute

$$a_2 = \frac{(\beta-2)(3\beta-4)(3\beta-2)}{128}(S\alpha)^{2\beta-2}. \tag{3.39}$$

Moreover, after a long calculation we find that

$$\frac{a_{n+1}}{a_n} = \frac{\sigma^2}{8(n+1)}[(2n+1)\beta-(2n+2)][(2n+1)\beta-2n](S\alpha)^{\beta-1}. \tag{3.40}$$

Putting the pieces back together, we find that

$$p(\tau,\alpha,S) = \frac{\sqrt{g(S)}}{\sqrt{4\pi\tau}}\mathcal{P}(S,\alpha)\exp\left(-\frac{d(\alpha,S)^2}{4\tau}\right)\sum_k a_k\tau^k \tag{3.41}$$

$$= \frac{1}{\sigma S^\beta\sqrt{2\pi\tau}}\left(\frac{\alpha}{S}\right)^{\beta/2}\exp\left(-\frac{(S^{1-\beta}-\alpha^{1-\beta})^2}{2\sigma^2(1-\beta)^2\tau}\right)\sum_{k=0}^\infty a_k\tau^k. \tag{3.42}$$

Just as in the previous BSM example, we will label the k-th partial sum of the above formula by p_k. We now comment on the non-convergence of this series. Note that if we apply the ratio series convergence test, we find

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \infty, \tag{3.43}$$

except in the BSM case where $\beta = 1$. Hence the series will always diverge for $\beta \neq 1$. One might expect this to invalidate the use of these perturbation methods for the CEV model. However, for $\beta \approx 1$, they approximate the true density quite well when the series is truncated after the first few a_k, as we will demonstrate in a few examples. However, if one takes k large, he finds this approximation becomes increasingly inaccurate in parameter regimes of increasing size.

We demonstrate this with plots of two examples, but first provide an alternative approximation formula from Labordere [98]. In Labordere [98, p.132], this approximation formula for the transition density is given by

$$p_{HL}(\tau,S,\alpha) = \frac{S^{-\beta}}{\sqrt{2\pi\sigma^2\tau}}\exp\left(-\frac{1}{2\sigma^2\tau}\left(\int_S^{S_0}\frac{1}{u^\beta}du\right)^2\right)\left(\frac{\alpha}{S}\right)^{\beta/2} \tag{3.44}$$

$$\times\left(1+\frac{1}{8}S_{av}^{2\beta-2}\beta(\beta-2)\tau\sigma^2 + \frac{\beta(\beta-2)(3\beta-2)(3\beta-4)}{128}S_{av}^{4\beta-4}(\tau\sigma^2)^2\right). \tag{3.45}$$

This is a second order formula where the heat kernel coefficients a_1 and a_2 have not been computed exactly but rather are approximated by their simpler diagonal values. In particular, the a_i are approximated by $a_i(\alpha, S) = a_i(S_{av}, S_{av})$ where $S_{av} = (S+\alpha)/2$ which takes advantage of the simple form of the diagonal heat kernel coefficients which in the one dimensional CEV case are given by

$$a_1(S,S) = Q(S), \quad a_2(S,S) = \frac{1}{2}\left(Q(S)^2 + \frac{\Delta_g Q(S)}{3}\right). \tag{3.46}$$

We now consider two examples of plots of the p_n. First we will set the model parameters to $\alpha = 1$, $T = 10$, $\sigma = 0.3$, and $\beta = 0.6$. In Figure 4, the greatest graph is the approximation from Labordere [98], next plot the exact transition density, our zeroth order correction, and our first order correction in the other plots.

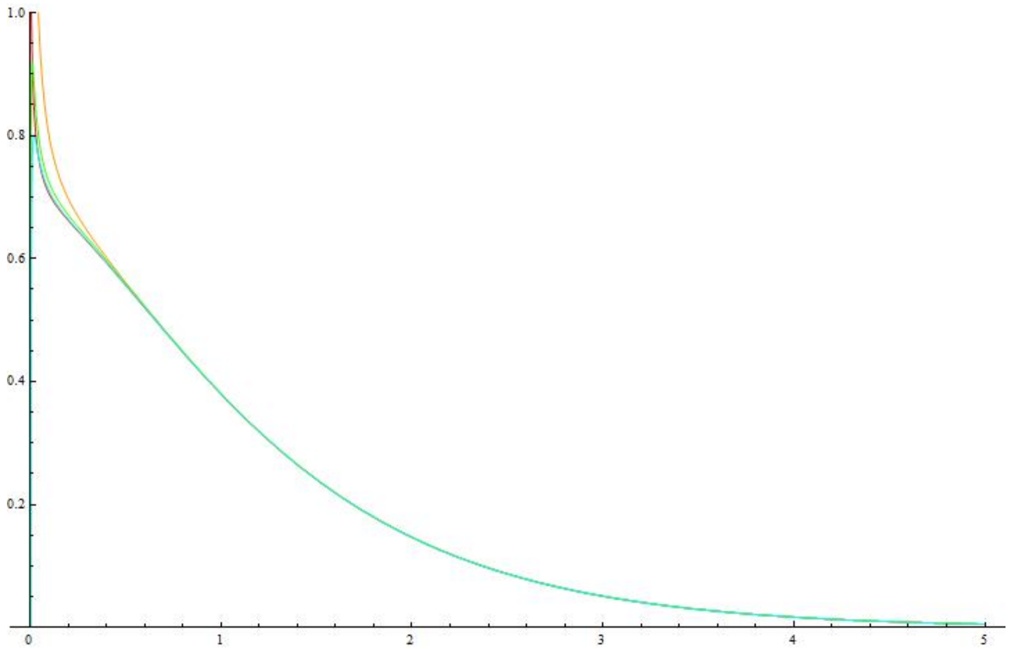

FIGURE 4. Here we plot p from equation (3.25) together with p_{HL} from equation (3.45) alongside our approximations in p_0 and p_1 from equation (3.42) for the parameter values $\alpha = 1$, $T = 10$, $\sigma = 0.3$, and $\beta = 0.6$.

We note that all the graphs are virtually indistinguishable for $S \in [0.5, 5]$. For small values $S \in [0.1, 0.5]$, our first and second order approximations are closer to the exact transition density function than p_{HL}. Now for very low values of S, all the approximations break down. This fact is brought out more clearly in Figure 5 where here in the middle exponential type graph we plot the exact solution, and in the greatest graph we again plot p_{HL}.

The other plots in decreasing order correspond to $p_1, p_2, p_3, p_5,$, and p_{10} respectively. Here we can see on the small scale $S \in [0, 0.1]$ that the approximation formulas are becoming increasingly inaccurate. We demonstrate the degeneration of the p_k in Figure 6.

Here, in increasing order of the graphs, we plot p_1, p_2, p_3, p_5, p_{10}, p_{20}, p_{30}, p_{40}, and p_{50}. Thus as n grows large, the associated p_n approximations degenerate. Also note that $p_{n+1} < p_n$ pointwise which can straightforwardly be deduced from the form of equation (3.42). In addition, it seems that whenever we find a point where the transition density is accurately approximated by some p_n, say at $S = S_0$, then the associated approximating function also gives an accurate approximation of the exact transition density for all $S > S_0$. One could exploit this fact in practice for instance by using a Monte Carlo method to check the accuracy of the approximation at some S value, and if he establishes that the approximation is indeed good, it can then be safely used for larger S values.

We now turn to one additional example where we keep all the model parameters fixed but let $\beta = 0.3$. In Figure 7, we plot the exact solution in the top red graph, and the second order approximation p_{HL} in green. Then the decreasing graphs are given in the plots $p_0, p_1, p_2, p_3, p_5, p_{10}$, and p_{20}.

Note again that the p_n degenerate as $n \to \infty$, and p_2 gives a more accurate second order approximation of the exact solution than p_{HL}.

Lastly, we provide a somewhat more global description of the error encountered in our CEV approximation in Figure 8. Here we again take $\alpha = 1$, $T = 10$, and $\sigma = 0.3$. We then

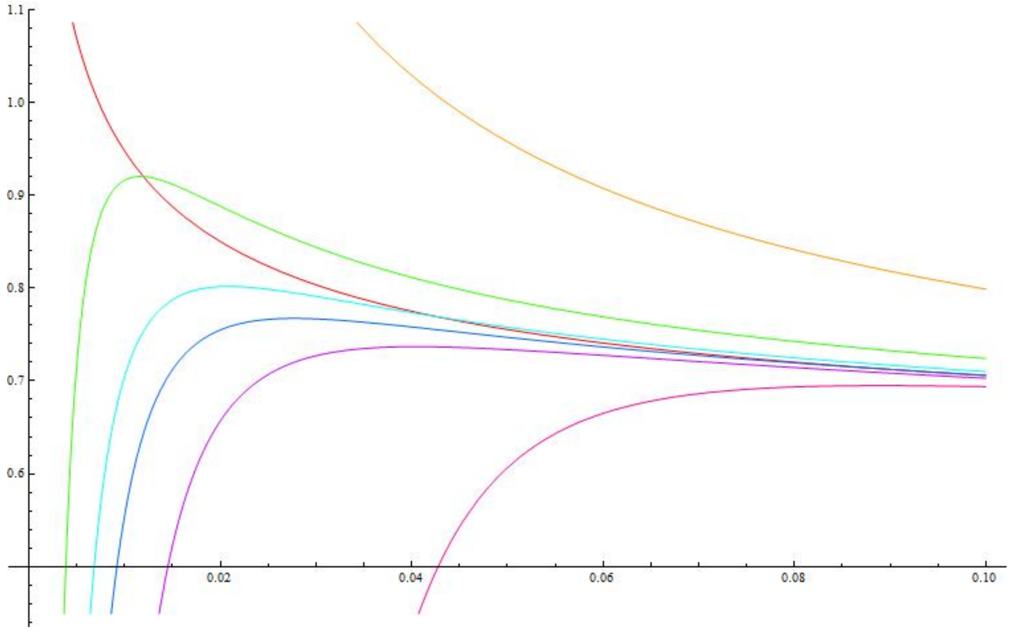

FIGURE 5. Here we plot p from equation (3.25) together with p_{HL} from equation (3.45) alongside our approximations in p_1, p_2, p_3, p_5, and p_{10} from equation (3.42) for the parameter values $\alpha = 1$, $T = 10$, $\sigma = 0.3$, and $\beta = 0.6$.

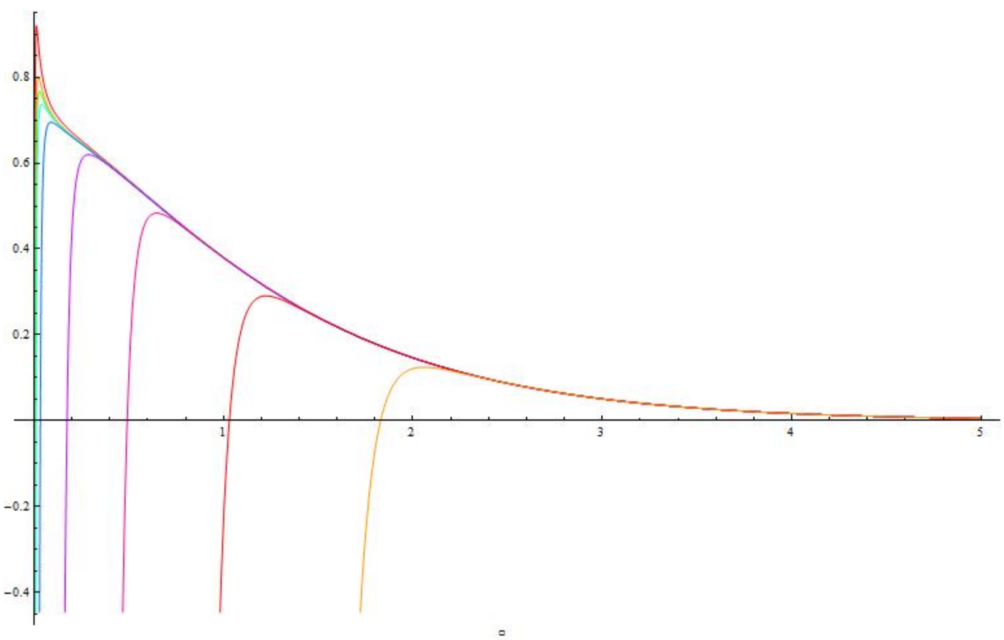

FIGURE 6. Here we plot p from equation (3.25) together with p_{HL} from equation (3.45) alongside our approximations in p_1,p_2, p_3, p_5, p_{10}, p_{20}, p_{40}, and p_{50} from equation (3.42) for the parameter values $\alpha = 1$, $T = 10$, $\sigma = 0.3$, and $\beta = 0.6$.

plot the difference between the exact solution and our second order approximation $p - p_2$ for $S \in [0, 0.5]$ and $\beta \in [0.001, 1]$.

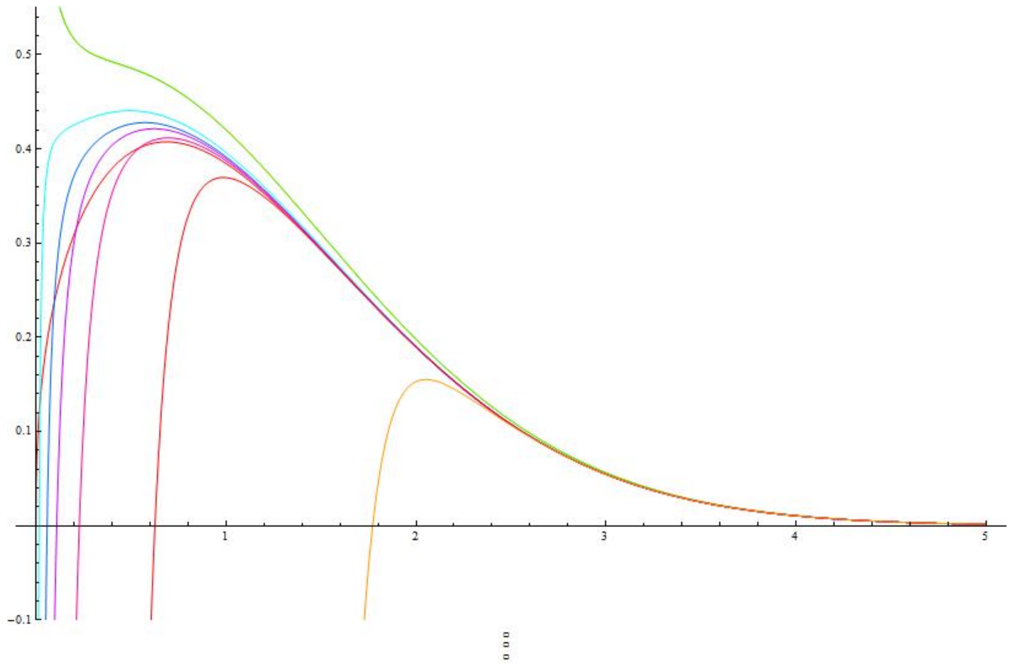

FIGURE 7. Here we plot p from equation (3.25) together with p_{HL} from equation (3.45) alongside our approximations in p_0, p_1, p_2, p_3, p_5, p_{10}, and p_{20} from equation (3.42) for the parameter values $\alpha = 1$, $T = 10$, $\sigma = 0.3$, and $\beta = 0.3$.

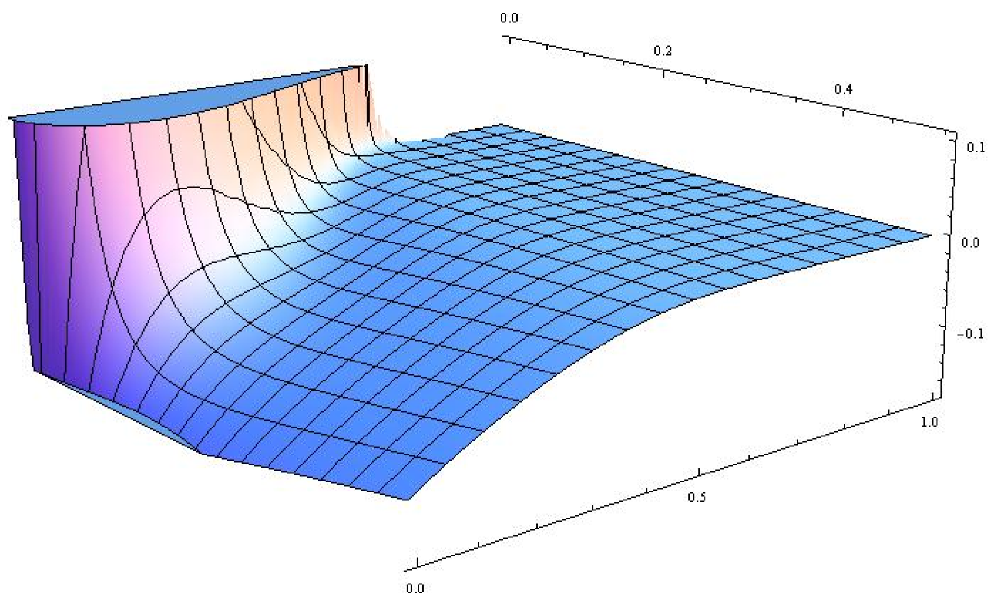

FIGURE 8. This is a second order absolute error plot of $p - p_2$, where p_2 is given by (3.42) and the exact transition density p is defined by equation (3.25). Here $S \in [0, 0.5]$ and $\beta \in [0.001, 1]$ with fixed model parameters $\alpha = 1$, $T = 10$, and $\sigma = 0.3$.

Note as we increase β, the graph tends to the zero plane for all S. In particular, the only places where trouble seems to arise is for small values of β (this appears to be independent of S) and for simultaneous small values of S and low values of β, e.g. $\beta \in [0, 0.5]$.

We finally note that it is possible to derive these results in a slightly simpler fashion. We can simplify the CEV SDE by changing variables. In particular, we can change variables by defining $x(t) = \bar{S}(t)^{2(1-\beta)}/\sigma^2(1-\beta)^2$. Then using Itō's lemma, we find that

$$dx = \frac{\partial x}{\partial \bar{S}} d\bar{S} + \frac{1}{2}\frac{\partial^2 x}{\partial \bar{S}^2} d\bar{S}^2 = \frac{1}{2}\frac{1-2\beta}{1-\beta}dt + 2\sqrt{x}dW \equiv \delta dt + 2\sqrt{x}dW. \tag{3.47}$$

The process x is known as a δ-Bessel process. Note for $\beta \in (-\infty, 1/2)$ then $\delta \in (0,1)$ and monotonically decreases as a function of β. For $\beta \in (1/2, 1)$, $\delta \in (0, -\infty)$ is monotonically decreasing. For $\beta \in (1, \infty)$, δ is decreasing from ∞ to 1.

The associated equation for the transition probability density is given by

$$p_\tau = 2\alpha p_{\alpha\alpha} + \delta p_\alpha. \tag{3.48}$$

where we again set $\alpha = x_0$ in order to simplify our notation. We will consider this processes for arbitrary δ.

We identify $\mu^\alpha = \delta$ and $\Sigma^{\alpha\alpha} = 4\alpha$. Thus $g^{\alpha\alpha} = 2\alpha$ and $g_{\alpha\alpha} = 1/2\alpha$. Again we change variables by defining an arclength coordinate $s(\alpha) = \frac{1}{\sqrt{2}}\int_\alpha^x \frac{du}{\sqrt{u}}$ which measures the distance we are away from x. This coordinate $g_{ss} = (\partial\alpha/\partial s)^2 g_{\alpha\alpha} = 1$, and thus we have $\Delta(\alpha, x) = 1$. Note that $s(x) = 0$ and $s(\alpha) = d(x, \alpha)$, so s parameterizes a geodesic by arclength starting from x and ending at 0, i.e. it parametrizes the curve backwards. We next compute $\Gamma^\alpha_{\alpha\alpha} = \frac{1}{2}g^{\alpha\alpha}\partial_\alpha g_{\alpha\alpha} = -1/2\alpha$, from which we see

$$A_\alpha = \frac{1}{2}g_{\alpha\alpha}\mu^\alpha + \frac{1}{2}\Gamma^\alpha_{\alpha\alpha} = \frac{(\delta-1)}{4\alpha} \equiv -\frac{\gamma}{\alpha}, \tag{3.49}$$

where we let $\gamma = -(\delta-1)/4$. Next we can compute

$$\mathcal{P}(\alpha, x) = \exp\left(-\int_\alpha^x A_\alpha d\alpha\right) = \left(\frac{x}{\alpha}\right)^\gamma. \tag{3.50}$$

so that $\mathcal{P}(x,\alpha) = (\alpha/x)^\gamma$. Now we compute

$$\mathcal{P}^{-1}(x,\alpha)\Delta_g\mathcal{P}(x,\alpha) = \frac{\mathcal{P}^{-1}}{\sqrt{g}}\partial_\alpha\left(g^{\alpha\alpha}\sqrt{g}\partial_\alpha\mathcal{P}\right) = \frac{\gamma(2\gamma-1)}{\alpha}, \tag{3.51}$$

$$\mathcal{P}^{-1}2g^{\alpha\alpha}A_\alpha\partial_\alpha\mathcal{P} = -\frac{4\gamma^2}{\alpha}, \tag{3.52}$$

From the above form of s, we note that the distance function is given by

$$d(x_1, x_2) = \sqrt{2}\left(\sqrt{x_2} - \sqrt{x_1}\right) \tag{3.53}$$

so

$$s = \sqrt{2}\left(\sqrt{x} - \sqrt{\alpha}\right) \to \alpha = \left(\sqrt{x} - \frac{s}{\sqrt{2}}\right)^2 \tag{3.54}$$

and $ds = -\sqrt{2}/(2\sqrt{\alpha})d\alpha$

Thus we can compute

$$a_1(x,\alpha) = \frac{1}{d}\int_{C(x,\alpha)} \mathcal{P}^{-1}(x,\alpha)L_\alpha\mathcal{P}(x,\alpha)ds = -\frac{2\gamma(1+2\gamma)}{\sqrt{x\alpha}\sigma}. \tag{3.55}$$

It is possible to proceed further with these computations to rederive the a_i in the CEV case.

We finally comment on the potential implications that these results may have for the SABR model. In [134], Paulot, has computed an explicit expression for a_0 and an analogous expression for a_1 which involves a numerical integration (although with a substantial computational effort, one may be able to construct an explicit formula for a_1). He then constructs three approximations for the implied volatility smile; to our knowledge, these are currently the best explicit approximation formulas for the implied volatilities of highly out of the money options in the SABR model. In an example, Paulot demonstrates that his second order formula degenerates for options with very low strike to a much greater degree than his first order approximation. Since the SABR model reduces to the CEV model as the volvol constant $\nu \to 0$, we suspect that this degeneration is an extension of the previously mentioned effect; namely, higher order correction terms of the CEV transition density cause the approximation to increasingly degenerate for low strikes. Thus it would not seem pertinent to compute a third order heat kernel correction for SABR (although such a computation probably is not realistic due to its complexity).

3.3. **Quadratic Local Volatility Model.** We now turn to the quadratic local volatility model which has been studied widely in the literature (c.f. Antonov and Misirpashaev [7]). Here we assume that the local volatility function is quadratic in the asset's price process. In particular, we assume that the dynamics are given by

$$dS_t = (a + bS_t + cS_t^2)dW, \quad S(0) = S_0, \tag{3.56}$$

where here a, b, and c, are taken to be constants. The properties of this SDE have been studied in Andersen [3]; the author also gives explicit formulas for European call prices when the underlying evolves with quadratic dynamics. For the sake of simplicity, we restrict to the case where the quadratic has positive distinct roots, which is a sufficient criteria for the existence of a real valued transition density function. The transition density PDE for this model is given by

$$\partial_\tau \phi = \frac{1}{2}(a + b\alpha + c\alpha^2)^2 \partial_\alpha^2 \phi, \quad \phi(0, \alpha, S) = \delta(S - \alpha). \tag{3.57}$$

It is possible to construct an exact solution of this equation using the heat kernel expansion method. To demonstrate this, we first note that the volatility function is given by $\Sigma = (a + b\alpha + c\alpha^2)^2$. The single component of the metric is $g_{\alpha\alpha} = 2/(a + b\alpha + c\alpha^2)^2$, and the Christoffel symbol is a rational function

$$\Gamma^\alpha_{\alpha\alpha} = -\frac{b + 2c\alpha}{a + b\alpha + c\alpha^2}. \tag{3.58}$$

Since the drift is trivial, the connection is just $A_\alpha = \Gamma^\alpha_{\alpha\alpha}/2$ and $Q = (b^2 - 4ac)/8$.

Using these expressions, we can calculate

$$\mathcal{P}(\alpha, S) = \sqrt{\frac{a + bS + cS^2}{a + b\alpha + c\alpha^2}}, \quad \mathcal{P}(S, \alpha) = \sqrt{\frac{a + b\alpha + c\alpha^2}{a + bS + cS^2}}. \tag{3.59}$$

The distance function is given by

$$d(\alpha, S) = \frac{2\sqrt{2}\left(\arctan\left(\frac{b+2cS}{\sqrt{4ac-b^2}}\right) - \arctan\left(\frac{b+2c\alpha}{\sqrt{4ac-b^2}}\right)\right)}{\sqrt{4ac - b^2}}. \tag{3.60}$$

Even though the distance function takes a somewhat complicated form, a_1 and a_2 can be computed exactly and can be expressed simply as

$$a_1(S, \alpha) = -\frac{1}{8}(b^2 - 4ac), \quad a_2(S, \alpha) = \frac{1}{128}(b^2 - 4ac)^2, \tag{3.61}$$

or more generally,

$$a_k = \frac{(-1)^k}{k! 2^{3k}}(b^2 - 4ac)^k. \tag{3.62}$$

Now, we can recognize the sum in the heat kernel expansion as an exponential function and invert the heat kernel coefficient series to find that

$$\sum_{k=0}^{\infty} a_k \tau^k = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{ac}{2} - \frac{b^2}{8}\right)^k \tau^k = \exp\left(\left[\frac{ac}{2} - \frac{b^2}{8}\right]\tau\right). \tag{3.63}$$

Therefore the exact transition density is given by

$$\phi(\tau, \alpha, S) = \frac{\sqrt{g(S)}}{\sqrt{4\pi\tau}} \mathcal{P}(S, \alpha) \exp\left(-\frac{d(\alpha, S)^2}{4\tau}\right) \sum_k a_k \tau^k \tag{3.64}$$

$$= \sqrt{\frac{a + b\alpha + c\alpha^2}{2\pi\tau(a + bS + cS^2)^3}} \exp\left(-\frac{d^2(\alpha, S)}{4\tau} + \left(\frac{ac}{2} - \frac{b^2}{8}\right)\tau\right), \tag{3.65}$$

which one can readily check solves equation (3.57). Moreover, we note that this solution satisfies the boundary values of equation (3.57). As $\tau \to \infty$, then $\phi \to 0$ if $\alpha \neq S$ since the exponential dominates the square root in the limit. If $\alpha = S$, then $d(\alpha, S) = 0$, and thus $\phi \approx 1/\sqrt{\tau}$ as $\tau \to 0$, so $\phi \to \infty$, which when formalized, shows that ϕ is a delta function initially.

The above is an example of the heat kernel method at its best. The heat kernel ansatz turned out to be a very good choice in the sense that we were able to determine a significant portion of the content of the functional form of the transition density from the zeroth order prefector. As a result, the a_k took particularly simple forms, and the perturbative formalism produced an exact solution.

3.4. **Cubic Local Volatility Model.** We now consider a local volatility model whose instantaneous volatility function is given by a cubic polynomial. We consider this model for two purposes. First, we wish to illustrate how the computation of the a_k becomes significantly more difficult in this polynomial model. Second, this model has applicability in computing a third order correction to a general local volatility model that depends on a small parameter, ν. Specifically, in certain cases, it could be useful to expand an instantaneous volatility function $\sigma(S; \nu)$ as

$$\sigma(S; \nu) = \sigma_0(S) + \nu\sigma_1(S) + \nu^2\sigma_2(S) + \nu^3\sigma_3(S) + O(\nu^4). \tag{3.66}$$

and then study the associated third order approximation of the local volatility model.

The model dynamics are given by

$$dS_t = (S_t - a)(S_t - b)(S_t - c)dW_t, \quad S(0) = \alpha. \tag{3.67}$$

The metric for this model is $g_{\alpha\alpha} = (\alpha - a)(\alpha - b)(\alpha - c)/2$, and the corresponding Christoffel symbol is

$$\Gamma^{\alpha}_{\alpha\alpha} = \frac{1}{2}\left(\frac{1}{a - \alpha} + \frac{1}{b - \alpha} + \frac{1}{c - \alpha}\right). \tag{3.68}$$

From this, we can deduce the form of $\mathcal{P}$,

$$\mathcal{P}(\alpha, S) = \left(\frac{(a - S)(b - S)(c - S)}{(a - \alpha)(b - \alpha)(c - \alpha)}\right)^{1/2}. \tag{3.69}$$

The distance function is given by

$$d(S, \alpha) = \frac{\sqrt{2}}{(a - b)(a - c)(b - c)} \ln\left(\left(\frac{S - a}{\alpha - a}\right)^{b-c} \left(\frac{S - b}{\alpha - b}\right)^{c-a} \left(\frac{S - c}{\alpha - c}\right)^{a-b}\right). \tag{3.70}$$

The a_1 integral can be computed exactly, but its expression is quite long and we do not write it here. Instead, we will compare the first order approximation formula which involves a_1 with Monte Carlo results below.

We now consider a numerical simulation which demonstrates the accuracy of this density formula for a specific choice of model parameters. In particular, we consider the cubic local volatility model with parameters $\tau = 1$, $a = 0.5$, $b = 0.75$, $c = 0.8$, $\sigma = 0.1$, and initial asset value $\alpha = 1$. We then simulate 10^5 paths using an Euler scheme discretization of equation (3.67) and organize and plot the results in a histogram in Figure 9. Note that the density

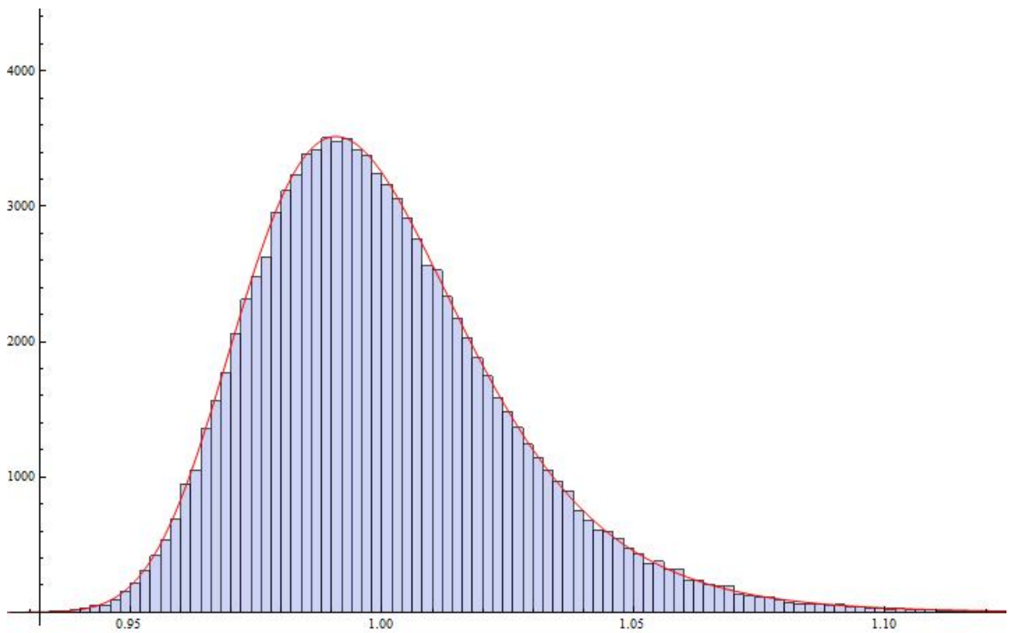

FIGURE 9. This is histogram for the transition density of the cubic local volatility model with parameters given by $\tau = 1$, $a = 0.5$, $b = 0.75$, $c = 0.8$, $\sigma = 0.1$, $\alpha = 1$. We plot the first order density approximation in red.

is slightly skewed to the left of the initial asset price. This is due to the fact that the cubic local volatility model has a tendency to pull the initial asset price towards the greatest root $c = 0.8$.

3.5. **Affine-Affine Short Rate Model.** We now turn attention to developing a first order approximation formula for the transition density of a model we refer to as the affine-affine short rate model. Let r_t represent a short rate whose evolution is governed by an SDE of the form

$$dr_t = (a + br_t)dt + (c + dr_t)dW_t, \quad r(0) = r_0, \tag{3.71}$$

where here a, b, c, and d are constants. We note that we could eliminate one of a or c by shifting r by an appropriate constant; however, we keep the above form for the sake of symmetry. In the case that $d = 0$, this reduces to the Vasicek model, (c.f. Andersen and Piterbarg [2], Brigo and Mercurio [23], and Brigo and Mercurio[24] for a discussion of short rate models). Although we will refer to this as a short rate model in light of this reduction, it can also be viewed as an asset model.

The drift is $\mu^\alpha = a + b\alpha$, and the instantaneous volatility function is $\sigma^\alpha_\alpha = (c + d\alpha)$. Hence $\Sigma^{\alpha\alpha} = (\sigma^\alpha_\alpha)^2 = (c + d\alpha)^2$, and thus the metric is given by $g = 2\Sigma^{-1} = 2/(c + d\alpha)^2$, whose Christoffel symbol is $\Gamma^\alpha_{\alpha\alpha} = \frac{1}{2} g^{\alpha\alpha} \partial_\alpha g_{\alpha\alpha} = -d/(c + d\alpha)$. We next need to determine the connection A_α as well as Q. These are given by

$$A_\alpha = \frac{2a - cd + (2b - d^2)\alpha}{2(c + d\alpha)^2}, \tag{3.72}$$

$$Q(\alpha) = \frac{4a^2 + c^2(4b + d^2) + 2cd^3\alpha + (d^2 - 2b)\alpha^2 - 8a(d(c + d\alpha) - b\alpha)}{8(c + d\alpha)^2}. \tag{3.73}$$

Using this, we find that

$$\mathcal{P}(\alpha, r) = \exp\left(-\int_\alpha^r A_r dr\right) = \left(\frac{c + dr}{c + d\alpha}\right)^{\frac{1}{2} - \frac{b}{d^2}} \exp\left(\frac{(bc - ad)(r - \alpha)}{d(c + dr)(c + d\alpha)}\right). \tag{3.74}$$

Again we can compute

$$2\mathcal{P}^{-1}(r, \alpha)\Delta_g \mathcal{P}(r, \alpha) + 2\mathcal{P}^{-1} g^{\alpha\alpha} A_\alpha \partial_\alpha \mathcal{P} = -Q. \tag{3.75}$$

Now, from the metric, we see that $ds = -\sqrt{2}/(c + \alpha d)d\alpha$. We finally need to compute the distance function for this metric in order to compute the a_i. To do this, we define an arc-length coordinate

$$s(\alpha) = \sqrt{2}\int_\alpha^r \frac{d\alpha}{c + \alpha d} = \frac{\sqrt{2}}{d} \ln\left(\frac{c + rd}{c + \alpha d}\right), \tag{3.76}$$

Again, in this coordinate the metric is just given by $g_{ss} = (\partial r/\partial s)^2 g_{rr} = 1$. Thus our distance function is

$$d(\alpha, r) = \frac{\sqrt{2}}{d} \ln\left(\frac{c + rd}{c + \alpha d}\right). \tag{3.77}$$

Using this, we can compute a somewhat involved explicit formula for a_1 given by

$$a_1(r, \alpha) = -\frac{1}{A}\Bigg[2d(bc - ad)(\alpha - r)\big(b(2c^2 + 4d^2 r\alpha + 3cd(r + \alpha)\big) + d\big(a(2c + d(r + \alpha)) \tag{3.78}$$

$$-4d(c + dr)(c + d\alpha))\big) + (d^2 - 2b)^2(c + dr)^2(c + d\alpha)^2 \ln\left(\frac{c + dr}{c + d\alpha}\right)\Bigg], \tag{3.79}$$

where

$$A = 8d^2(c + dr)^2(c + d\alpha)^2 \ln\left(\frac{c + dr}{c + d\alpha}\right). \tag{3.80}$$

We are able to compute an analytic expression for a_2 as well, but it is quite long and we refrain from writing it for this reason. We will use the second order formula for numerical experiments.

We first consider a Monte Carlo simulation in Figure 10. Here we fix model parameters $a = 0.5$, $b = -0.1$, $d = -0.2$, $\tau = 1$, with an initial asset price of $\alpha = 1$ and generate 10^5 paths via an Euler scheme and plot the resulting density approximation.

Here the drift term pushes the asset's distribution function to the right of its original value with a fair amount of dispersion. We plot our second order density approximation in red and note again we have strong agreement with the numerical result.

We next consider two examples that give us insight into how altering the model parameters effects the behavior of the transition density.

In Figure 11, we let $\alpha = 1$, $b = 0.05$ (we think of b as an interest rate r), $c = 0.0001$ (since our exact second order formula is not defined for $c = 0$), $d = 0.3$ (which we think of as a constant volatility σ), and $T = 1$. We then plot our first order approximation to the

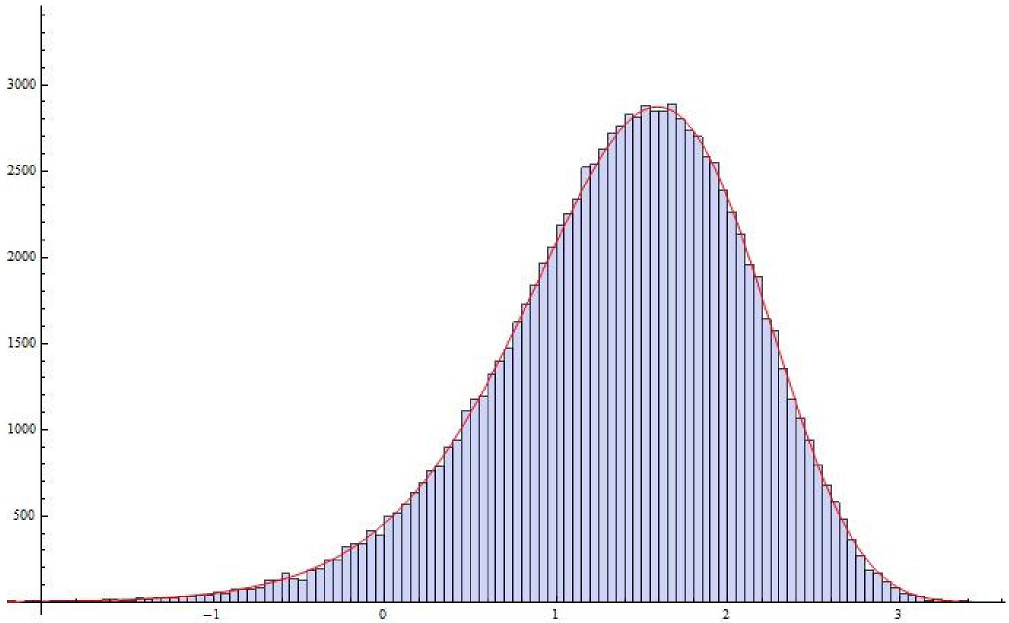

FIGURE 10. This is a histogram for the transition density of the Affine-Affine short rate model with parameters $\alpha = 1$, $a = 0.5$, $b = -0.1$, $d = -0.2$, and $\tau = 1$. In red, we plot our second order approximation formula.

transition density for the affine-affine short rate model over the parameter space $a \in [0, 0.5]$, $r \in [0, 2]$. We note that when $a \approx 0$, this model is close to the BSM case we previously considered. As we increase a, we can see that the associated density cross-section's peaks shift to the right. Also for the larger a values, we see that our transition densities begin to become negative and thus degenerate.

One way to measure how close the approximate density function is to the exact solution is to compute its integral over $\mathbb{R}^+$. Let p_a^1 denote the first order density function with parameter value a, and $I_a^1 = \int_{\mathbb{R}^+} p_a^1$. Then we compute $I_0^1 = 1$, $I_{0.1}^1 = 1.00117$, $I_{0.2}^1 = 0.996971$, $I_{0.3}^1 = 0.963378$, $I_{0.4}^1 = 0.826313$, and $I_{0.5} = 0.453194$ which reflects the degeneration we see in the plot as a increases.

Now in Figure 12, we plot our second order formula for the same parameter values. Note that as we increase the value of a, the associated density functions begin to become large. This shows that the approximation rapidly degenerates in this parameter regime.

We can again compute integrals similar to those we previously considered which have values $I_0^2 = 1.00000$, $I_{0.1}^2 = 1.00001$, $I_{0.2}^2 = 0.99939$, $I_{0.3}^2 = 0.99967$, $I_{0.4}^2 = 1.01966$, and $I_{0.5} = 1.13304$. Again we see that the approximations degenerate, but they do so fairly slowly relative to the first order formula.

Next in Figure 13, we plot cross sections of Figure 12 for $a = 0.0, a = 0.2, a = 0.5, a = 0.6$, and $a = 0.65$. Here we can see that the associated densities begin to grow without bound for increasing a values. Now in Figure 14, we fix $a = 0$ and let $c \in [0, 0.5]$ vary while keeping all other parameters fixed as well as keeping $S \in [0, 2]$ and plot our first order approximation. Note here that the initial BSM transition density disperses as we increase c. This is expected, as we are increasing the volatility as we increase c.

Next in Figure 16, we plot the second order version of our graph in Figure 15. Note here that there is very little difference from our previous plot. In fact, for our parameter restriction, the two plots differ pointwise by less than 0.03 over both surfaces' domains.

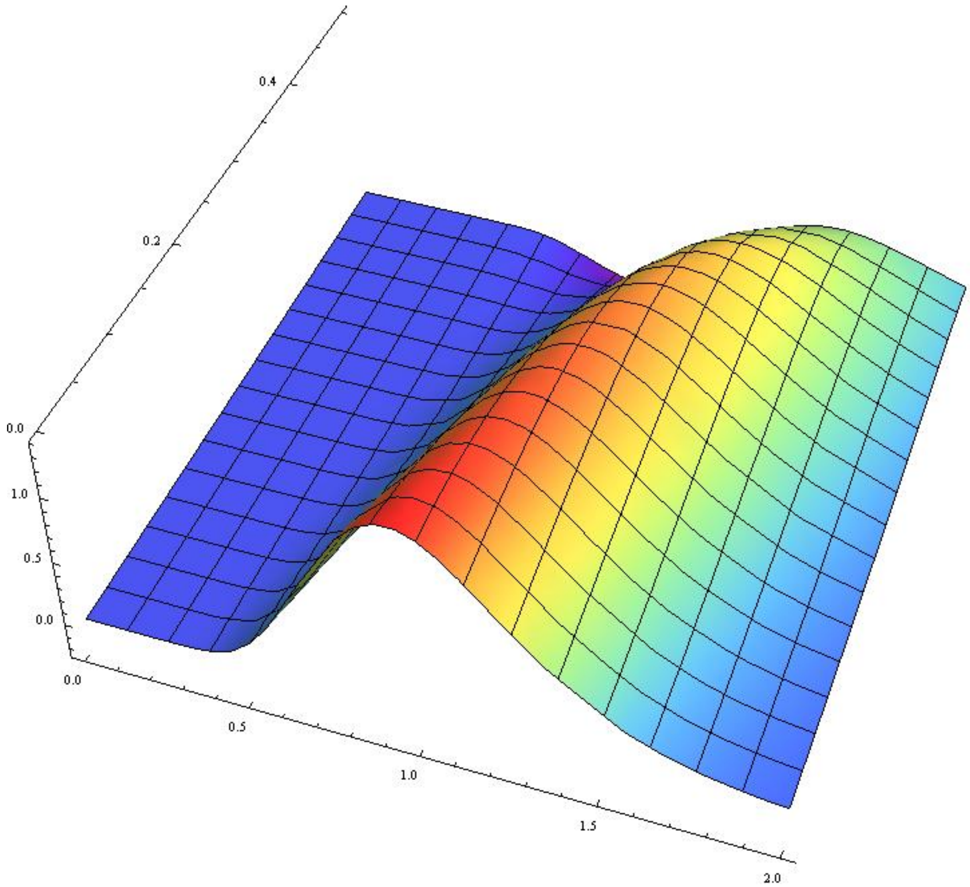

FIGURE 11. Plot of the first order density approximation for the affine-affine model where $a \in [0, 0.5]$ and $S \in [0, 2]$ with $b = 0.05$, $c = 0.0001$, $d = 0.3$, and $T = 1$.

We finally plot cross sections of this surface in Figure 16. Here, we plot the cross sections for $c = 0.0001$, $c = 0.1$, $c = 0.2$, $c = 0.3$, and $c = 0.5$. Note that for any $c \neq 0$ there is a nontrivial probability that $r < 0$ (we have cut off the negative region of the density function plot). Thus under these models, it is possible that interest rates can become negative (and in fact more probable with increasing c). For these parameters we compute $I_{0.0001} = 1$, $I_{0.1} = 0.999995$, $I_{0.2} = 0.998391$, $I_{0.3} = 0.985889$, $I_{0.4} = 0.959241$, and $I_{0.5} = 0.924588$.

3.6. **Generalized CEV Model.** We now consider another short rate model which can roughly be characterized as a generalized CEV model with generalized mean reversion. Specifically, the process reverts to the mean at a rate proportional to r_t^a. The dynamics for this model are given by

$$dr_t = k\left[\theta - r_t^a\right]dt + \sigma r_t^\beta dW_t, \quad r_0 = \alpha. \tag{3.81}$$

Just as in the CEV case, the volatility function, metric, and Christoffel symbol for this model are given by $\Sigma^{\alpha\alpha} = \sigma^2\alpha^{2\beta}$, $g_{\alpha\alpha} = 2/(\sigma^2\alpha^{2\beta})$, and $\Gamma^\alpha_{\alpha\alpha} = -\beta/\alpha$, respectively. The connection component takes the form

$$A_\alpha = \frac{1}{2}\left(\frac{2k\left(\theta - \alpha^a\right)}{\sigma^2\alpha^{2\beta}} - \frac{\beta}{\alpha}\right), \tag{3.82}$$

and

$$Q = \frac{1}{8\sigma^2\alpha^{2(1+\beta)}}\left(4k^2\alpha^2(\alpha^a - \theta)^2 - \sigma^4\alpha^{4\beta}(\beta - 2)\beta - 4k\sigma^2\alpha^{1+2\beta}(a\alpha^a + 2\beta(\theta - \alpha^a))\right). \tag{3.83}$$

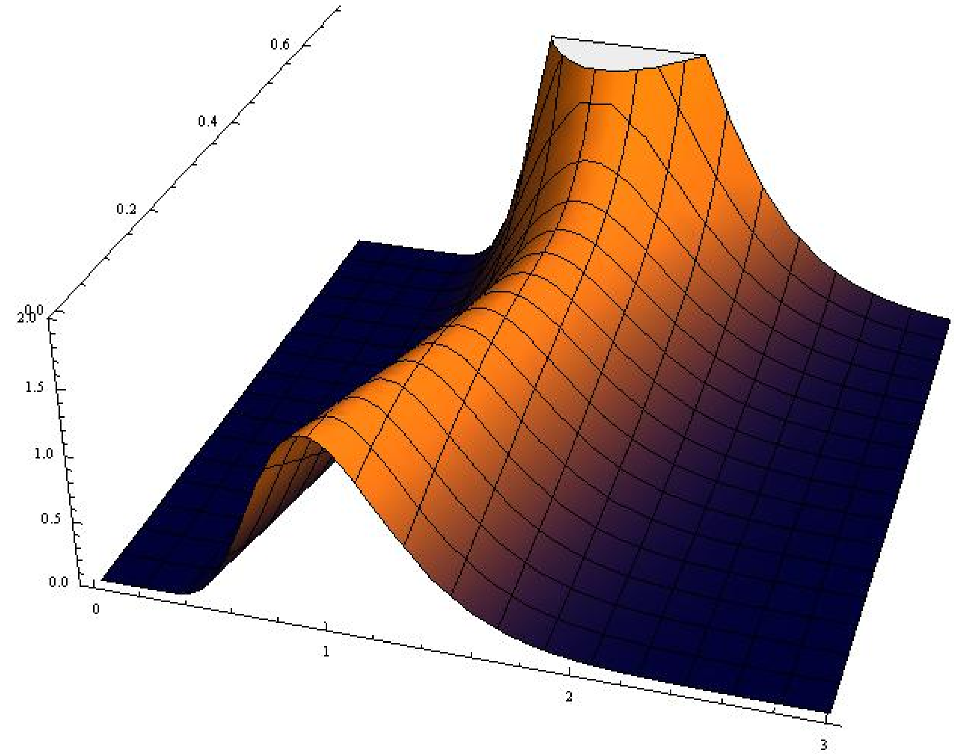

FIGURE 12. Plot of the second order density approximation for $a \in [0, 0.5]$ and $S \in [0, 2]$ with $b = 0.05$, $c = 0.0001$, $d = 0.3$, and $T = 1$.

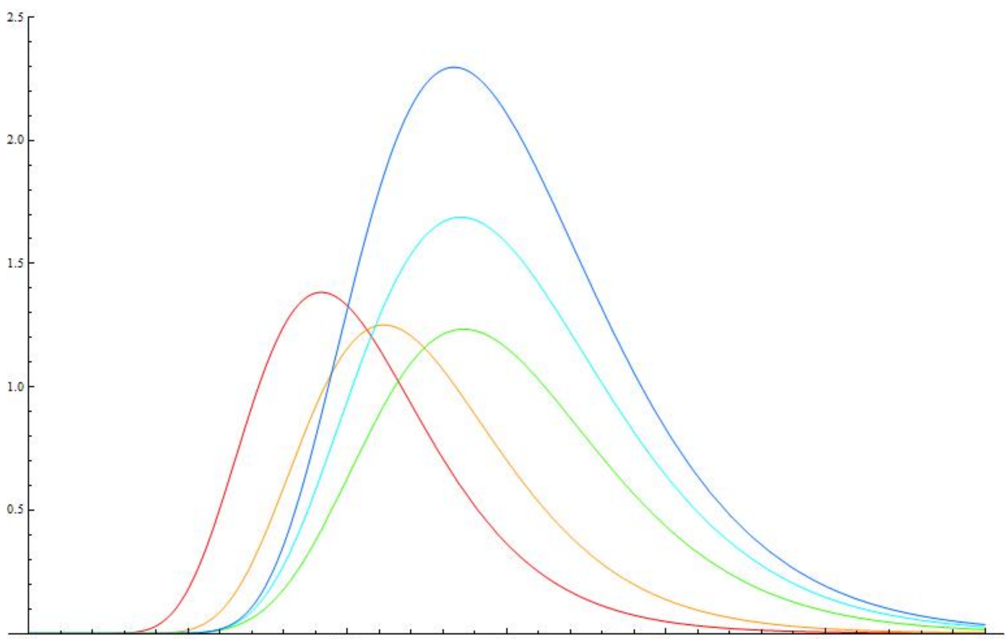

FIGURE 13. Plots for $S \in [0, 2]$ of cross sections of the second order density function for there affine-affine model where $a = \{0.0, 0.2, 0.5, 0.6, 0.65\}$

Next, we compute

$$\mathcal{P}(r, \alpha) = \left(\frac{\alpha}{r}\right)^{\beta/2} \exp\left(-\frac{k\left(r^{1+a-2\beta} - \alpha^{1+a-2\beta}\right)}{\sigma^2(1 + a - 2\beta)} + \frac{k\left(r^{1-2\beta} - \alpha^{1-2\beta}\right)}{\sigma^2(1 - 2\beta)}\right). \tag{3.84}$$

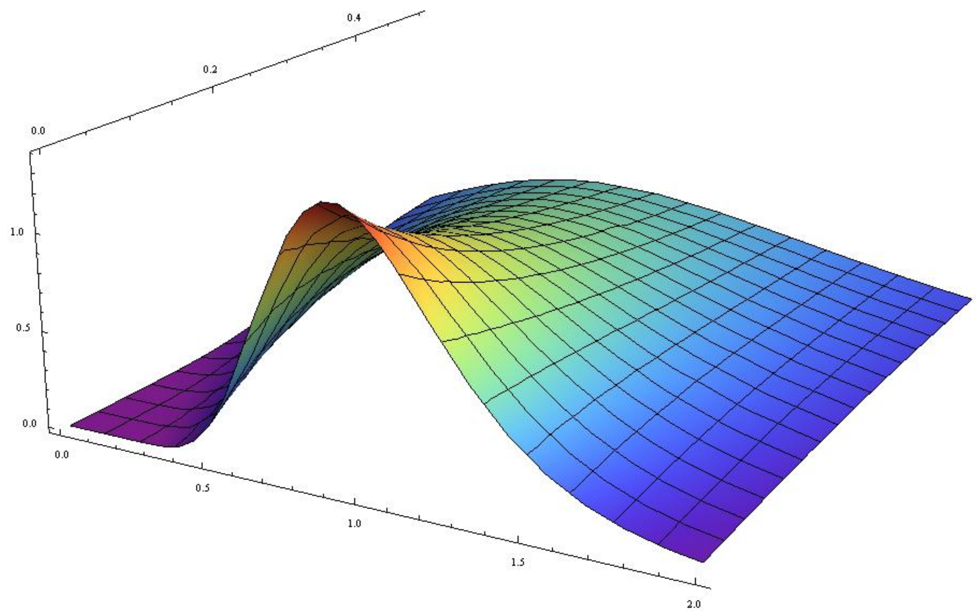

FIGURE 14. Plot of the first order density approximation for the affine-affine model where $c \in [0.0001, 0.5]$ and $S \in [0, 2]$ with $a = 0.0$, $b = 0.05$, $d = 0.3$, and $T = 1$.

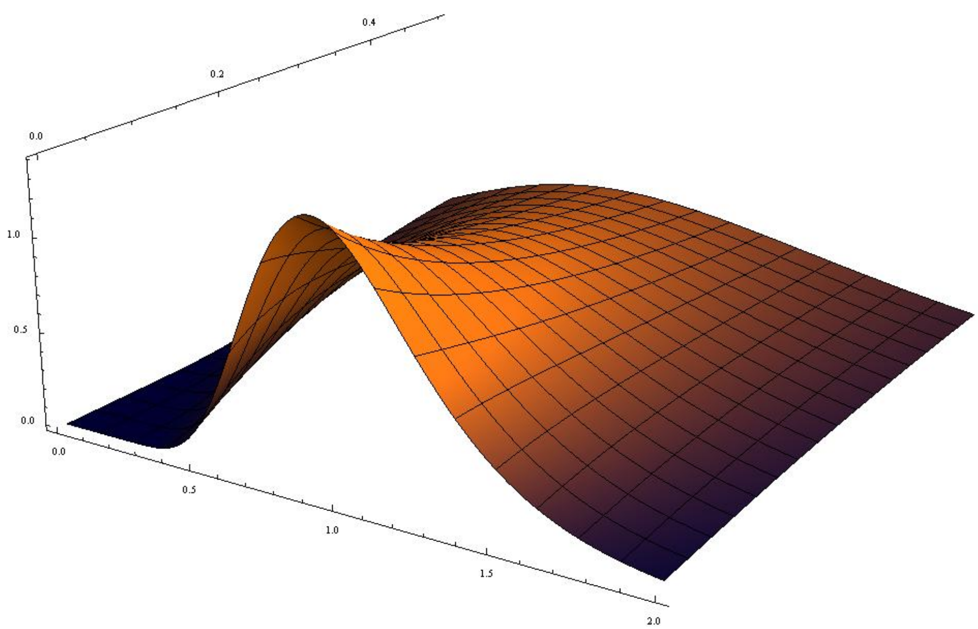

FIGURE 15. Plot of the second order density approximation for the affine-affine model where $c \in [0.0001, 0.5]$ and $S \in [0, 2]$ with $a = 0.0$, $b = 0.05$, $d = 0.3$, and $T = 1$.

The distance function for this metric is the same as in the CEV case. Thus we can use this to compute

$$a_1(r, \alpha) = \frac{(\beta - 1)(r\alpha)^\beta}{8\sigma^2(r^\beta \alpha - r\alpha^\beta)} \left[\frac{\sigma^4 \beta(\beta - 2)(r\alpha^\beta - r^\beta \alpha)}{r\alpha(\beta - 1)} \right. \tag{3.85}$$

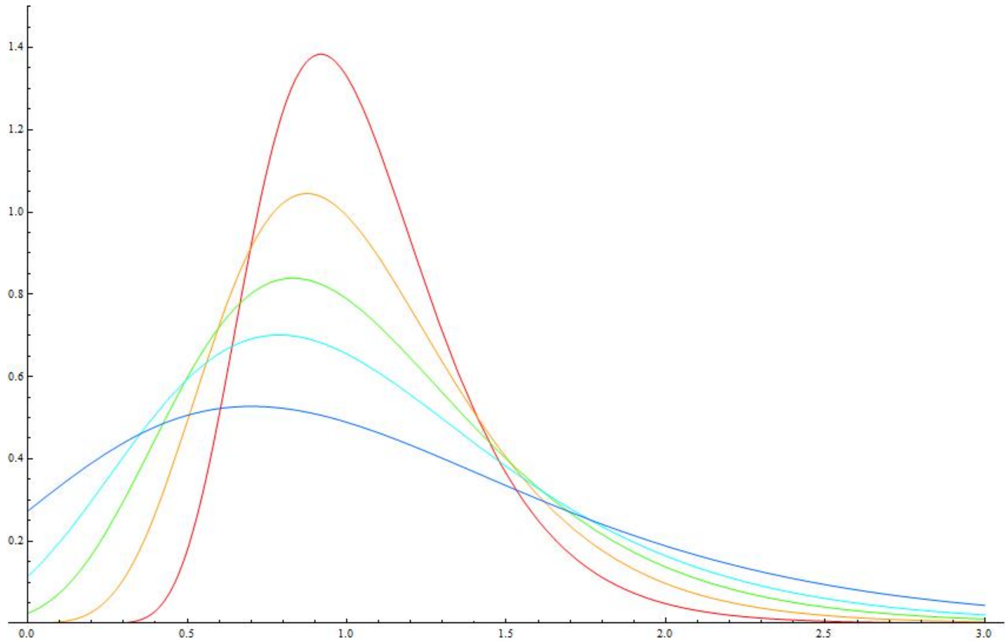

FIGURE 16. Plots for $S \in [0,2]$ of cross sections of the second order density function of the affine-affine model for $c = \{0.0001, 0.1, 0.2, 0.3, 0.5\}$

$$+\frac{4k\sigma^2}{(\beta-a)(r\alpha)^\beta}\left[a(r^a\alpha^\beta - r^\beta(\alpha^a - 2\theta) - 2\alpha^\beta\theta) + 2\beta(r^\beta(\alpha^a-\theta) - r^a\alpha^\beta + \alpha^\beta\theta))\right] \tag{3.86}$$

$$+4k^2\left(\frac{r^{1+2a-3\beta} - \alpha^{1+2a-3\beta}}{1+2a-3\beta} - \frac{2\theta(r^{1+a-3\beta} - \alpha^{1+a-3\beta})}{1+a-3\beta} + \frac{\theta^2(r^{1-3\beta} - \alpha^{1-3\beta})}{1-3\beta}\right)\Bigg]. \tag{3.87}$$

We are unable to compute a_2 exactly. We note that there are several values of the model parameters where the formula for a_1 is not defined; one should either restrict this class of models to preclude such parameters, or takes limits of the above expression to attempt to obtain approximation formulas.

We now turn to our first numerical experiment in Figure 17. Here, we fix model parameters $\theta = 1.1$, $k = 0.5$, $a = 0.5$, $\tau = 1$, $\sigma = 0.3$, $\beta = 0.6$, and initial asset price $\alpha = 1$ and simulate 10^5 paths form this model's SDE and plot the results in a histogram. The resulting distribution remains centered around the initial asset value, and is slightly skewed to the right. Again, we see that we have very good agreement between the Monte Carlo, and approximation densities.

We finally provide some evidence that the I_a are good approximations of the true transition density by denoting $I_a^1 = \int_{\mathbb{R}^+}\phi$, and computing via numerical integration, $I_0^1 = 1.00002$, $I_{0.5}^1 = 0.997325$, $I_1^1 = 0.996833$, $I_{1.5}^1 = 0.998367$, $I_2^1 = 1.00175$, $I_{2.5}^1 = 1.00683$, and $I_3^1 = 1.0134$.

We are unable to compute a_2 exactly. We note that there are several values of the model parameters where the formula for a_1 is not defined. In these cases, one needs to take limits of the relevant geometric quantities in order to construct a usable approximation for the transition density.

We consider one example which shows how the parameter a affects the associated transition densities of the process. In Figure 18, we plot our first order approximation formula for the transition density for different parameter values $a \in [0,3]$ and rates/asset values $r \in [0,2]$; we let r range out of reasonable interest rate values in order to consider how the approximation

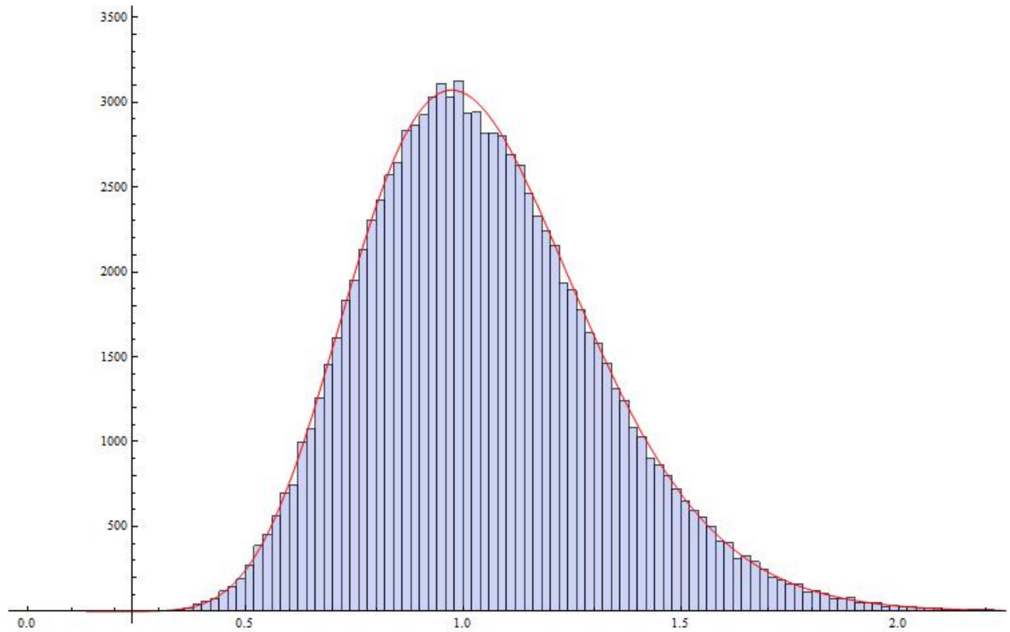

FIGURE 17. This is a histogram for the simulated transition density of the generalized CEV model with parameters $\alpha = 1$, $\theta = 1.1$, $k = 0.5$, $a = 0.5$, $\tau = 1$, $\sigma = 0.3$, $\beta = 0.6$. In red, we plot our first order approximation formula.

performs if the process is taken to be an exchange rate or asset. We note here that as one increases the value of a, the associated transition densities tend to spike near the value of α.

In Figure 19 we plot the increasing profiles of the previous three dimensional plot for values of a in the set $\{-0.5, 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0\}$. We note the graphs for positive a seem to be suitably regular. When we consider the $a = -0.5$ plot, we note that the graph hits the zero level around $r = 0.35$. In fact, the plot is negative for all r values before this. This behavior appears to be generic for all negative a, and it thus seems appropriate to preclude such numbers as valid model parameter values.

We finally provide some evidence that the I_a are good approximations of the true transition density by denoting $I_a^1 = \int_{\mathbb{R}+} p$, and computing via numerical integration, $I_0^1 = 1.00002$, $I_{0.5}^1 = 0.997325$, $I_1^1 = 0.996833$, $I_{1.5}^1 = 0.998367$, $I_2^1 = 1.00175$, $I_{2.5}^1 = 1.00683$, and $I_3^1 = 1.0134$, all of which are close to one which indicates that the associated psuedo-densities are approximately bona-fide transition densities.

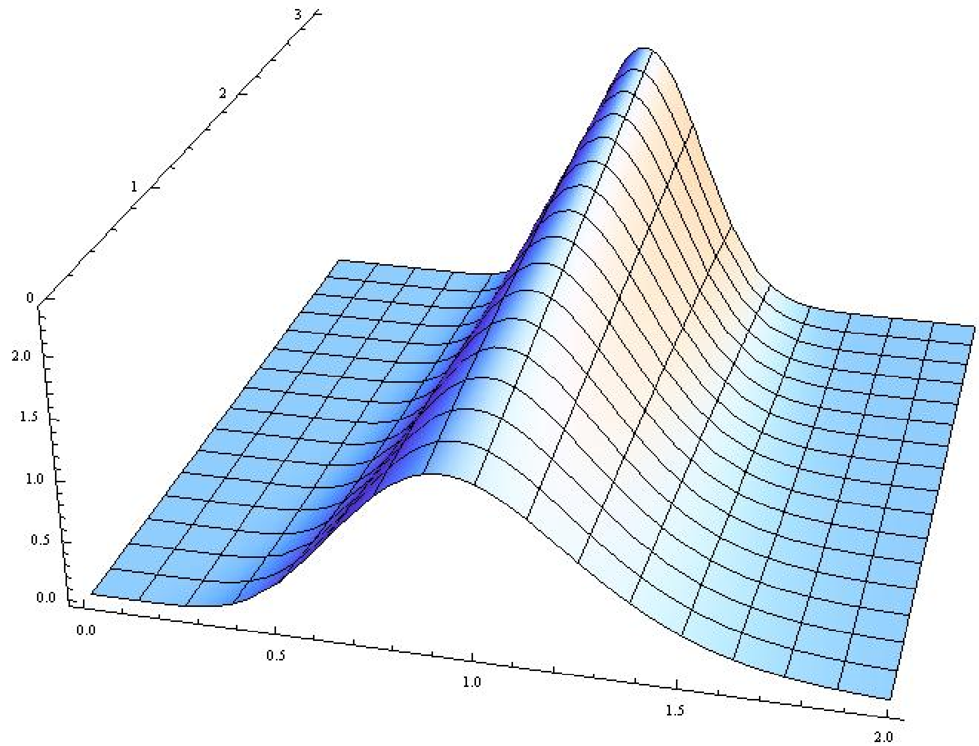

FIGURE 18. Plots for $r \in [0, 2]$ and $a \in [0, 3]$ of crossections of the first order density function of the short rate model where we take $\alpha = 1$, $\theta = 1$, $b = 0.5$, $T = 1$, $\sigma = 0.3$, and $\beta = 0.6$.

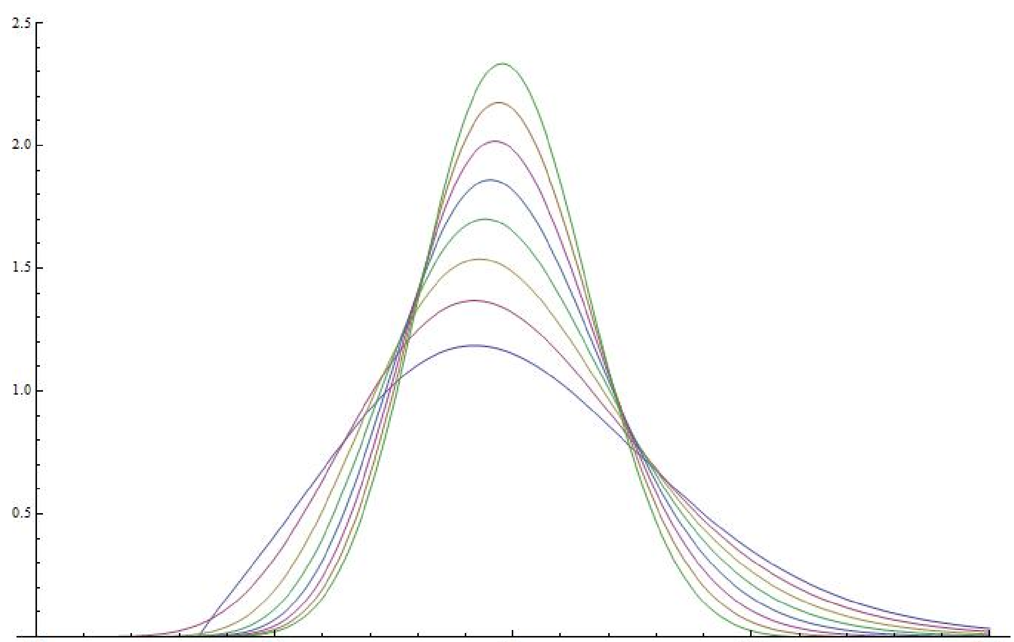

FIGURE 19. Plots for $r \in [0, 2]$ of cross sections of the second order density function of the affine-affine model for $a = \{-0.5, 0.0, 0.5, 1, 1.5, 2.0, 2, 5, 3.0\}$ where we take $\alpha = 1$, $\theta = 1$, $b = 0.5$, $T = 1$, $\sigma = 0.3$, and $\beta = 0.6$.

4. Two Dimensional Stochastic Volatility Models

We now consider several two dimensional stochastic volatility models. We first make a connection between a certain class of such models and associated two dimensional geometries. We then examine the history of implied volatility approximation formulas for the SABR model, and review and extend results related to constructing such a formula using heat kernel expansion techniques. We next briefly look at the Heston model followed by a new class of stochastic volatility models. We finally consider a different type of perturbation theory which allows one to build implied volatility formulas for stochastic volatility models and also show how geometric constructions arise in this method.

4.1. **Geometry of A Class of Two Dimensional Stochastic Volatility Models.** We now turn to two dimensional models which extend the class of one dimensional model by assuming that the instantaneous volatility is stochastic; the resulting models are known as stochastic volatility models. We will use the geometric fact that one can always express any two dimensional metric in form $g = e^{2\phi}\delta$ where $\delta = dx^2 + dy^2$ is the standard Euclidean metric and $\phi = \phi(x, y)$ is at least a C^2 function of the local coordinates (x, y) c.f. Alfhors [1], Do Carmo [44]. The coordinates in which the metric takes this form are known as isothermal coordinates, and they will greatly simplify many of the computations involved in constructing the geometric quantities needed for the heat kernel expansion. In two dimensions, one has more than an existence theorem for such coordinates, in fact, they are simple to construct. We now consider a generic two dimensional stochastic volatility model and show how to construct the model's corresponding isothermal coordinate system.

Consider a general driftless stochastic local volatility model in two dimensions that has dynamics given by

$$dS_t = \sigma_t S_t \gamma(S_t) dW_t^1, \quad d\sigma_t = \mu(\sigma_t)dt + b(\sigma_t)dW_t^0, \quad \mathbb{E}[dW_t^0, dW_t^1] = \rho dt \tag{4.1}$$

where here S_t represents the evolution of a financial observable, i.e. an asset price, interest rate, etc., which is driven by the product of a stochastic volatility process σ_t and a function $\gamma(S_t)$. The volatility σ_t evolves according to its own process and it is assumed to have some drift $\mu(\sigma_t)$ and volatility $b(\sigma_t)$, which we will refer to as the volvol function. In addition, we assume that the functions γ, μ, and b are all suitably regular such that all the subsequent calculations are well-defined; for instance, taking them to be C^2 is usually sufficient. We also suppose that the Brownian motions W^1 and W^0 are correlated, with correlation parameter $\rho > 0$. Note that by choosing the appropriate numeraire, we lose no generality by taking the drift of the S_t process to be zero.

We note that the covariance matrix for this class of stochastic volatility models is given by $\Sigma^{ij} = \sigma^i_k \sigma^j_l \rho^{kl}$, where we note that one should not confuse the indexed local volatility functions σ^i_j with the unindexed volatility process σ_t which we will just denote by σ for notational convenience. We can express this and its inverse in matrix form according to

$$\Sigma = \begin{pmatrix} \sigma^2 S^2 \gamma^2 & \rho\sigma S\gamma b \\ \rho\sigma S\gamma b & b^2 \end{pmatrix}, \quad \Sigma^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} S^{-2}\sigma^{-2}\gamma^{-2} & -\rho S^{-1}\sigma^{-1}\gamma^{-1}b^{-1} \\ -\rho S^{-1}\sigma^{-1}\gamma^{-1}b^{-1} & b^{-2} \end{pmatrix}. \tag{4.2}$$

Thus the line element for the inverse metric, in the coordinates (S, σ), is given by

$$ds^2_{g^{-1}} = \frac{1}{2}\Sigma = \frac{1}{2}\left(\sigma^2 S^2 \gamma^2 dS^2 + 2\rho\sigma S\gamma b dS d\sigma + b^2 d\sigma^2\right). \tag{4.3}$$

We wish to simplify this expression by defining new coordinates

$$q(S) = \int_{S_0}^{S} \frac{du}{u\gamma(u)}, \quad \xi(\sigma) = \int_{\sigma_0}^{\sigma} \frac{u du}{b(u)}, \tag{4.4}$$

and noting that

$$\frac{\partial q}{\partial S} = \frac{1}{S\gamma(S)}, \quad \frac{\partial \xi}{\partial \sigma} = \frac{\sigma}{b(\sigma)}, \tag{4.5}$$

which can be used to compute the inverse metric components in the new coordinates. To understand how this is done, note that given a $(0,2)$ tensor's local coordinate representation g_x^{ij} in a two dimensional coordinate system $x = (x_1, x_2)$, one can find its representation in a new coordinate system $\tilde{x} = (\tilde{x}_1, \tilde{x}_2)$ by using the transformation law formula (which follows from the multilinear mapping definition of a tensor, c.f. Nakahara [126]):

$$g_{\tilde{x}}^{kl} = \sum_{i,j=1}^{2} \frac{\partial \tilde{x}^k}{\partial x^i}\frac{\partial \tilde{x}^l}{\partial x^j} g_x^{ij}. \tag{4.6}$$

For our particular coordinate change, this takes the form

$$g^{qq} = \left(\frac{\partial q}{\partial S}\right)^2 g^{SS} + 2\frac{\partial q}{\partial S}\frac{\partial q}{\partial \sigma} g^{S\sigma} + \left(\frac{\partial q}{\partial \sigma}\right)^2 g^{\sigma\sigma} = \left(\frac{\partial q}{\partial S}\right)^2 g^{SS} = \frac{\sigma^2}{2}, \tag{4.7}$$

$$g^{q\xi} = \frac{\partial q}{\partial S}\frac{\partial \xi}{\partial S} g^{SS} + \frac{\partial q}{\partial S}\frac{\partial \xi}{\partial \sigma} g^{S\sigma} + \frac{\partial q}{\partial \sigma}\frac{\partial \xi}{\partial S} g^{\sigma S} + \frac{\partial q}{\partial \sigma}\frac{\partial \xi}{\partial \sigma} g^{\sigma\sigma} = \frac{\rho\sigma^2}{2}, \tag{4.8}$$

$$g^{\xi\xi} = \left(\frac{\partial \xi}{\partial S}\right)^2 g^{SS} + 2\frac{\partial \xi}{\partial S}\frac{\partial \xi}{\partial \sigma} g^{S\sigma} + \left(\frac{\partial \xi}{\partial \sigma}\right)^2 g^{\sigma\sigma} = \left(\frac{\partial \xi}{\partial \sigma}\right)^2 g^{\sigma\sigma} = \frac{\sigma^2}{2}, \tag{4.9}$$

so that in these new coordinates, the metric takes the simplified form,

$$ds_g^2 = \frac{\sigma^2}{2}\left(dq^2 + 2\rho dq d\xi + d\xi^2\right). \tag{4.10}$$

Now we change coordinates once again, in order to reduce the metric even further. Let $w = q - \rho\xi$ so that we find

$$g^{ww} = \left(\frac{\partial w}{\partial q}\right)^2 g^{qq} + 2\frac{\partial w}{\partial q}\frac{\partial \xi}{\partial \xi} g^{q\xi} + \left(\frac{\partial w}{\partial \xi}\right)^2, \quad g^{\sigma\sigma} = \frac{\sigma^2}{2}(1-\rho^2), \tag{4.11}$$

$$g^{w\xi} = \frac{\partial w}{\partial q}\frac{\partial \xi}{\partial q} g^{qq} + \frac{\partial w}{\partial q}\frac{\partial \xi}{\partial \xi} g^{q\xi} + \frac{\partial w}{\partial \xi}\frac{\partial \xi}{\partial q} g^{\xi q} + \frac{\partial w}{\partial \xi}\frac{\partial \xi}{\partial \xi} g^{\xi\xi} = 0, \tag{4.12}$$

where we also note that $g^{\xi w} = 0$ by symmetry. Thus the metric is diagonal in this coordinate system,

$$ds_g^2 = \frac{\sigma^2}{2}\left((1-\rho^2)dw^2 + d\xi^2\right). \tag{4.13}$$

We finally can make the coordinate scalings uniform by defining $x = \sqrt{2}/[(1-\rho^2)]w$ and $y = \sqrt{2}\xi$ so that

$$g^{xx} = \left(\frac{\partial x}{\partial w}\right)^2 g^{ww} = \sigma^2, \quad g^{yy} = \left(\frac{\partial y}{\partial \xi}\right)^2 g^{\xi\xi} = \sigma^2 \tag{4.14}$$

which implies that the inverse metric takes the conformally flat form

$$g^{-1} = \sigma^2(y)(dx^2 + dy^2). \tag{4.15}$$

Note here that $\sigma = \sigma(y)$, which is a simplification of the generic conformally flat where $\sigma = \sigma(x,y)$. The reason we do not need to consider conformal factors with x dependence is due to the fact that we only assumed b was a function of the volatility process σ, rather than a function of both S and σ. This simplification is key; in the upcoming computations, it will allow us to find that the distance function associated to this metric can be represented exactly modulo computing antiderivatives and inverting an algebraic equation.

The metric is constructed by taking the inverse of the inverse metric which is

$$g = \frac{1}{\sigma^2(y)}(dx^2 + dy^2) \equiv e^{2\phi(y)}(dx^2 + dy^2), \tag{4.16}$$

where we let $\phi = -\ln \sigma$ be our conformal factor in order to stick with standard geometric conventions.

We then can compute the Christoffel symbols and associated curvature tensors of this conformally flat metric, which take the relatively simple forms

$$\Gamma^x_{xy} = \Gamma^y_{yy} = -\Gamma^y_{xx} = \phi'(y), \tag{4.17}$$

$$Ric = -\phi''(y)(dx^2 + dy^2), \quad R = -2e^{-2\phi(y)}\phi''(y). \tag{4.18}$$

In two dimensions, the Ricci and Riemann tensors each only have one independent component which is proportional to the Gauss curvature K of g; also, the scalar curvature is given by $R = 2K$ (c.f. Hamilton [81]), hence all the curvature tensors encode the same information. Moreover, the Gauss curvature can not be arbitrarily specified; its form is constrained by the prescribed Gauss curvature equation (see Mateljevic et. al. [112] for more detail). These curvature computations can be used to construct new transition density approximations for stochastic volatility models. In fact, they are particularly useful in the case of the SABR model whose metric g corresponds to the hyperbolic metric. We will also need the following computations for the norms of the curvatures that are used in approximating the a_i heat kernel coefficients.

$$|Riem|^2 = g^{ia}g^{jb}g^{kc}g^{ld}R_{ijkl}R_{abcd} = R^{ijkl}R_{ijkl} = 4e^{-4\phi}(\phi'')^2, \quad |Ric|^2 = g^{ia}g^{jb}R_{ij}R_{ab} = 2e^{-4\phi}(\phi'')^2. \tag{4.19}$$

Our next aim is to try to solve the geodesic equations for this class of stochastic volatility models whose solutions can be used to construct the distance function associated to g. An explicit (or nearly explicit, that is to say, modulo relatively fast numerical analysis) expression for the distance function is key for the viability of heat kernel expansion methods.

Our aim now is to try to solve the geodesic equations in as general a context as possible for this model. After solving these equations for a specific model, we will then use the form of the resulting geodesics to compute the associated distance function.

Let $\gamma(t) = (x(t), y(t))$ be a geodesic joining two points $p_1, p_2 \in \mathbb{R}^2$ (again we assume uniqueness of this geodesic which can be proven under certain hypotheses in Forde [60]) which have representations $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ in isothermal coordinates. We want the geodesic to be parametrized by arclength, so that we can eventually construct a distance function from it. This choice of parameterization requires $|\dot{\gamma}|^2 = 1$, or equivalently

$$1 = g_{ij}\dot{\gamma}^i\dot{\gamma}^j = e^{2\phi}(\dot{x}^2 + \dot{y}^2). \tag{4.20}$$

The geodesic equations are given by

$$0 = \ddot{x} + \Gamma^x_{xy}\dot{x}\dot{y} + \Gamma^x_{yx}\dot{y}\dot{x} = \ddot{x} + 2\phi'(y)\dot{x}\dot{y}, \tag{4.21}$$

$$0 = \ddot{y} + \Gamma^y_{yy}(\dot{y})^2 + \Gamma^y_{xx}(\dot{x})^2 = \ddot{y} + \phi'(y)[(\dot{y})^2 - (\dot{x})^2]. \tag{4.22}$$

We can view the first equation as a first order equation for $\dot{x}$. In particular, if we let $z = \dot{x}$, then we can write

$$\dot{z} = -2z\phi'(y)\dot{y} \to \frac{d}{dt}\ln z = -2\phi'(y)\dot{y} = -2\frac{\partial \phi}{\partial t}, \tag{4.23}$$

so that

$$\dot{x} = c_1 e^{-2\phi}, \tag{4.24}$$

for some constant c_1. We now can use the arc length parameterization constraint in equation (4.20) in order to compute

$$\dot{y}^2 = e^{-2\phi} - \dot{x}^2 = e^{-2\phi} - c_1^2 e^{-4\phi}, \tag{4.25}$$

from which we find

$$\frac{dx}{dy} = \frac{\dot{x}}{\dot{y}} = \frac{c_1 e^{-2\phi}}{\sqrt{e^{-2\phi} - c_1^2 e^{-4\phi}}} = \frac{c_1}{\sqrt{e^{2\phi} - c_1^2}}. \tag{4.26}$$

Integrating this equation, we find

$$x_2 - x_1 = \int_{y_1}^{y_2} \frac{c_1 dy}{\sqrt{e^{2\phi(y)} - c_1^2}}, \tag{4.27}$$

where we note that the chain rule requires

$$\int_{y_1}^{y_2} \frac{dx}{dy} dy = \int_{y_1}^{y_2} \frac{dx}{dt}\frac{dt}{dy} dy = \int_{t_1}^{t_2} \frac{dx}{dt} dt = x(t_2) - x(t_1) = x_2 - x_1. \tag{4.28}$$

One can determine the constant $c_1 = c_1(x_1, y_1, x_2, y_2)$ by solving this equation. We note that this can seldom be done exactly, and one often needs to use a numerical algebraic equation solver in order to estimate c_1, which we will see occurs for the Heston model.

Now from the form of the metric, we know that the infinitesimal line element is given by

$$ds_g = e^{\phi}\sqrt{dx^2 + dy^2}. \tag{4.29}$$

Thus if we wish to find the length of the geodesic γ which joins the point $p_1 = (x_1, y_1)$ to $p_2 = (x_2, y_2)$, we need to evaluate the arclength functional

$$d(p_1, p_2) = l(\gamma) = \int_{p_1}^{p_2} ds = \int_{t_1}^{t_2} e^{\phi(y_1(t))}\sqrt{dx(t)^2 + dy(t)^2} = \int_{y_1}^{y_2} \frac{e^{2\phi(y)}}{\sqrt{e^{2\phi(y)} - c_1^2}} dy, \tag{4.30}$$

where we used equation (4.26) for the last step. Thus the distance function can be solved up to quadrature and inversion of an algebraic equation in any stochastic volatility model in the form of equation (4.1). In particular, it admits and explicit algebraic representation for the class of geometries where one can evaluate the two integrals in equations (4.26) and (4.30). We will now give a few examples where one indeed can do this.

For a toy example, we first consider the case of Euclidean space where $g = dx^2 + dy^2$, which corresponds to choosing $\phi = 0$. In this case, the integral for c_1 can be evaluated, which produces

$$x_2 - x_1 = \frac{c_1}{\sqrt{1 - c_1^2}}(y_2 - y_1) \to c_1 = \pm \frac{x_1 - x_2}{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}}, \tag{4.31}$$

from which we see that we can evaluate the distance integral to find

$$d(p_1, p_2) = \frac{y_2 - y_1}{\sqrt{1 - c_1^2}} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \tag{4.32}$$

as expected.

We now turn to the more involved case of the hyperbolic metric where the conformal factor is given by $\phi = -\ln y$. We again can evaluate the c_1 integral to find that

$$x_2 - x_1 = \frac{\sqrt{1 - c_1^2 y_1^2}}{c_1} - \frac{\sqrt{1 - c_1^2 y_2^2}}{c_1}, \tag{4.33}$$

which has solution

$$c_1 = \pm\frac{2(x_1 - x_2)}{\sqrt{((x_1 - x_2)^2 + y_1^2)^2 + 2((x_1 - x_2)^2 - y_1^2)y_2^2 + y_2^4}}. \tag{4.34}$$

We can also perform the integral involved in the distance function to find that

$$d(p_1, p_2) = \operatorname{arccoth}\left(\sqrt{1 - c_1^2 y_1^2}\right) - \operatorname{arccoth}\left(\sqrt{1 - c_1^2 y_2^2}\right) \tag{4.35}$$

$$= \frac{1}{2}\ln\left[\frac{(\sqrt{1 - c_1^2 y_1^2} + 1)(\sqrt{1 - c_1^2 y_2^2} - 1)}{(\sqrt{1 - c_1^2 y_1^2} - 1)(\sqrt{1 - c_1^2 y_2^2} + 1)}\right], \tag{4.36}$$

where we used the definition of the inverse hyperbolic tangent

$$\operatorname{arccoth}(x) = \frac{1}{2}\log\left(\frac{x + 1}{x - 1}\right). \tag{4.37}$$

After a fair amount of algebra, one can show that

$$\sqrt{\ln\left[\frac{(\sqrt{1 - c_1^2 y_1^2} + 1)(\sqrt{1 - c_1^2 y_2^2} - 1)}{(\sqrt{1 - c_1^2 y_1^2} - 1)(\sqrt{1 - c_1^2 y_2^2} + 1)}\right]} = \frac{c_1\left((x_1 - x_2)^2 + y_1^2 + y_2^2\right) + 2(x_1 - x_2)}{2y_1 y_2 c_1}, \tag{4.38}$$

which can be used alongside the definition of the inverse hyperbolic cosine

$$\operatorname{arccosh}(x) = \ln\left(x + \sqrt{x^2 - 1}\right), \tag{4.39}$$

to represent the two dimensional hyperbolic distance function in the standard form

$$d(p_1, p_2) = \operatorname{arccosh}\left(1 + \frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2y_1 y_2}\right). \tag{4.40}$$

We finally note that the above is a brute force approach to determine the distance function for the two dimensional hyperbolic geometry. In fact, there are a variety of simpler ways of achieving this classical result (see e.g. Goldman [74]) and even performing the analogue of the above calculation in the complex plane will greatly reduce the basic algebra involved in arriving at equation (4.40).

We now turn to the case where $e^{2\phi} = y^a$ or $\phi = a(\ln y)/2$. Evaluating the distance integral, we find

$$x_2 - x_1 = iy_2F_1\left(\frac{1}{a}, \frac{1}{2}, 1 + \frac{1}{a}, \frac{y^a}{c_1^2}\right)\Bigg|_{y_1}^{y_2}, \tag{4.41}$$

where here $a \neq -1$, $i = \sqrt{-1}$, and ${}_2F_1$ is the ordinary hypergeometric function

$${}_2F_1(a, b, c, x) = \sum_{k=0}^{\infty}\frac{(a)_k(b)_k}{(c)_k}\frac{x^k}{k!}, \quad |x| < 1, \tag{4.42}$$

which uses the Pochhammer symbol defined by

$$(x)_k \equiv \frac{x!}{(x - k)!} = x(x - 1)(x - 2)\cdots(x - k + 1). \tag{4.43}$$

One can extend the definition of $(x)_k$ to values of x outside the unit disc by analytic continuation on the complex plane (see e.g. Arfken [9], Hassani [82]). We also note that one should not be worried that the above expression is imaginary, as solving for c_1 and substituting will result in a final purely real formula.

In the case when $a = -1$, the above integral can be evaluated in terms of elementary functions, and the equation for c_1 becomes

$$c_1^2(x_2 - x_1) = c_1\left[\sqrt{y_1 - c_1^2 y_1^2} - \sqrt{y_2 - c_1^2 y_2^2}\right] - \arcsin(c_1\sqrt{y_1}) + \arcsin(c_1\sqrt{y_2}), \tag{4.44}$$

which gives an algebraic equation that defines $c_1(x_1, y_1, x_2, y_2)$ implicitly in terms of the x_i and y_i. We now turn to evaluating the distance function for this class of models. In the case $a = -1$, this simplifies to

$$d(p_1, p_2) = \frac{2}{c_1}\left(\arcsin[c_1\sqrt{y_2}] - \arcsin[c_1\sqrt{y_1}]\right). \tag{4.45}$$

If $a = -2$, then the distance function integral has value

$$d = i \arctan\left(\frac{1}{\sqrt{c_1^2 y^2 - 1}}\right)\Bigg|_{y=y_1}^{y_2}. \tag{4.46}$$

More generically when $a \neq 0, -1, -2$, we have that

$$d(p_1, p_2) = \frac{2y}{(a+2)\sqrt{y^a - c_1^2}}\left[y^a - c_1\sqrt{c_1^2 - y^a}\,{}_2F_1\left(\frac{1}{a}, \frac{1}{2}, 1 + \frac{1}{a}, \frac{y^a}{c_1^2}\right)\right]\Bigg|_{y_1}^{y_2}. \tag{4.47}$$

There are also a variety of other cases where ${}_2F_1$ simplifies to a composition of elementary functions (c.f. Daalhuis [41]).

There are three two dimensional geometries where there are well-known formulas for the distance function, namely, the geometries of constant curvature: which are locally the sphere, plane, or hyperbolic space. The spherical case does not seem relevant to finance since financial variables tend to be noncompact.

There are a variety of other choices for $e^{2\phi}$ which are different from those above that we considered which admit analytic distance functions. For example, if we take $e^{2\phi} = e^a$, or $\phi = a/2$. In this case, the relevant integrals are given by

$$\int \frac{e^{ay}}{\sqrt{e^{ay} - c_1^2}}dy = \frac{2}{a}\sqrt{e^{ay} - c_1^2}, \quad \int \frac{c_1 dy}{\sqrt{e^{ay} + c_1^2}} = -\frac{2}{a}\operatorname{arctanh}\left(\frac{\sqrt{c_1^2 + e^{ay}}}{c_1}\right). \tag{4.48}$$

This choice of ϕ has an asset SDE of the form $dS = \sigma S e^{-aS} dW^1$. In particular, for smaller values of the asset price, the local volatility will be small, and thus the asset value will not change dramatically.

A second example which could be further explored where both integrals can be expressed as elliptic functions if one takes $e^{2\phi} = \sin^2(y)$, or if one exchanges $\sin y$ with $\cos y$. If we choose $e^{2\phi} = \log y$, then these integrals can be evaluated up to an Efri function (which is a real valued function that is defined by a certain complex transform of the error function). There are indeed a wide variety of functions one can choose for the conformal factor in order to force the distance function to admit an exact representation. In fact, one could use such functions as a starting point to study novel stochastic volatility models, since we will see that knowledge of the distance function will allow one to construct at least a zeroth order approximation formula for the implied volatility of a vanilla call option on the model's asset.

4.2. **Geometry of Extended Stochastic Volatility Models.** We now wish to consider stochastic volatility models of the form

$$dS_t = \sigma_t S_t \gamma(S_t) dW_t^1, \quad d\sigma_t = \mu(\sigma_t)dt + b(\sigma_t, S_t)dW_t^0, \quad \mathbb{E}\left[dW_t^0, dW_t^1\right] = \rho dt, \tag{4.49}$$

where now the volvol function b depends on the asset process as well as σ. Our main goal is to give an indication of the difficulties involved in computing exact expressions for the distance function for such models.

The inverse metric for this class of models takes the same form as before,

$$g^{-1} = \frac{1}{2}\left(\sigma^2 S^2\gamma^2(S)dS^2 + 2\sigma\rho\gamma(S)Sb(S,\sigma)dSd\sigma + b(S,\sigma)^2 d\sigma^2\right). \tag{4.50}$$

(Note that $g^{-1} : T^*M \otimes T^*M \to \mathbb{R}$, so dS^2 really is vector products as opposed to a one form product). We wish to attempt to constructively diagonalize g^{-1} by finding the appropriate isothermal coordinates. Let's attempt to proceed in the same manner as in our previous case by defining a coordinate $q(S) = \int_{S_0}^{S} du/(u\gamma(u))$, and letting $\xi = \xi(S,\sigma)$ be a function which we will obtain by solving a differential equation. As in our previous example, we want the inverse metric to take the form $g^{qq} = \sigma^2/2$, $g^{q\xi} = \rho\sigma^2/2$, and $g^{\xi\xi} = \sigma^2/2$. The first equation is identically satisfied. Using the $(0,2)$ tensor transformation law that we utilized before, we find that the second and third equations require that ξ must be determined by

$$g^{q\xi} = \frac{1}{2}\xi_S\sigma^2 S\gamma + \frac{1}{2}\xi_\sigma\rho\sigma b = \frac{\rho}{2}\sigma^2, \tag{4.51}$$

$$g^{\xi\xi} = \frac{1}{2}\xi_S^2\sigma^2 S^2\gamma^2 + \xi_\sigma\xi_S\rho\sigma S\gamma b + \frac{1}{2}\xi_\sigma^2 b^2 = \frac{1}{2}\sigma^2, \tag{4.52}$$

which are algebraic equations in ξ_σ and ξ_S.

The system has two solutions. Either,

$$\xi_\sigma = \frac{\sigma}{b}, \quad \xi_S = 0, \quad \text{or} \quad \xi_\sigma = -\frac{\sigma}{b}, \quad \xi_S = \frac{2\rho}{S\gamma}. \tag{4.53}$$

The first solution corresponds to our choice of ξ in the previous section and requires that $b = b(\sigma)$ only. This is not acceptable for our current model consideration. Moreover, the second system imposes constraints on the model's functions. Equating mixed partials, we must have

$$\frac{\sigma\partial_S b}{b^2} = \xi_{S\sigma} = \xi_{\sigma S} = 0, \tag{4.54}$$

since we have assumed that γ is not a function of σ. Thus we would need b to be independent of S in order to make such a coordinate transformation, which contradicts our model assumption. Thus we cannot proceed as before. The calculation indicates that the isothermal coordinates for this class of metrics depend on both of the local coordinates, i.e. $x = x(S,\sigma)$ and $y = y(S,\sigma)$. This significantly complicates the task of constructing these coordinates (and in "most" cases no explicit form exists). It also shows that no stochastic volatility model in this class corresponds to hyperbolic geometry. To construct these coordinates, one needs to solve the nonlinear first order system

$$f(S,\sigma) = g^{xx} = x_S^2 g^{SS} + 2x_S x_\sigma g^{S\sigma} + x_\sigma^2 g^{\sigma\sigma}, \tag{4.55}$$

$$0 = g^{xy} = x_S y_S g^{SS} + x_S y_\sigma g^{S\sigma} + x_\sigma y_S g^{\sigma y} + x_\sigma y_\sigma g^{\sigma\sigma}, \tag{4.56}$$

$$f(S,\sigma) = g^{yy} = y_S^2 g^{yy} + 2y_S y_\sigma g^{S\sigma} + y_\sigma^2 g^{\sigma\sigma}, \tag{4.57}$$

(where f is the isothermal conformal factor, and where for our class of models $g^{SS} = \sigma^2 S^2\gamma^2/2$, $g^{S\sigma} = \sigma\rho\gamma Sb/2$, and $g^{\sigma\sigma} = b^2/2$). Although the existence/uniqueness theory shows this can always be done, exact solution can only be achieved in relatively simple cases.

Even if one is able to construct the (x,y) coordinate system where g^{-1} is diagonal, i.e.,

$$g^{-1} = f^2(x,y)(dx^2 + dy^2), \tag{4.58}$$

he still has a long way to go in order to use these techniques to produce an exact approximation formula for the implied volatility. The next step in the process is to attempt to construct an exact formula for the distance function. The main point that we wish to emphasise now is that $f = f(x,y)$ as opposed to $f = f(y)$ as was the case in the previous class of stochastic volatility models that we considered. This increases the complexity of the geometry since the conformal factor $f(x,y) \equiv e^{\phi(x,y)}$ depends on both local coordinates. In particular, we now write, $g = e^{2\phi(x,y)}\delta$ where here δ is the Euclidean metric in the x, y coordinates .

All the Christoffel symbols for this metric are now nontrivial and take the form

$$\Gamma^x_{xx} = -\Gamma^x_{yy} = \Gamma^y_{xy} = \phi_x, \tag{4.59}$$

$$\Gamma^x_{xy} = -\Gamma^y_{xx} = \Gamma^y_{yy} = \phi_y, \tag{4.60}$$

and the Ricci tensor's components and scalar curvature are given by

$$R_{xx} = R_{yy} = -\phi_{xx} - \phi_{yy} = -\Delta_\delta \phi, \quad R = -2e^{-2\phi}(\phi_{yy} + \phi_{xx}), \tag{4.61}$$

(see e.g. Chow [33]). The geodesic equations coupled to the arclength parameterization constraint are now given by

$$\ddot{x} + \phi_x(\dot{x})^2 + 2\phi_y\dot{x}\dot{y} - \phi_x(\dot{y})^2 = 0, \tag{4.62}$$

$$\ddot{y} - \phi_y(\dot{x})^2 + 2\phi_x\dot{x}\dot{y} + \phi_y(\dot{y})^2 = 0, \tag{4.63}$$

$$1 = e^{2\phi}(\dot{x}^2 + \dot{y}^2). \tag{4.64}$$

These equations cannot be solved up to quadrature and an implicit function inversion like in our previous example. There is a large literature across many different fields that addresses the numerical simulation of the geodesics equations, c.f. Bronstein et. al. [25], Chen et. al. [38], Memoli [116], Ying and Candes [170], as well as higher dimensional Lorentzian considerations in the numerical classical gravity literature Sharp [143]. One is also free to attempt to simply compute the Christoffel symbols and write the geodesic equation in (S,σ) coordinates, but this typically leads to an even more complicated representation of the geodesic equations.

We now illustrate some of the above complexities with a model which extends the SABR model. The dynamics are given by

$$dS = \sigma(t)S(t)^\beta dW_1(t), \quad d\sigma(t) = \mu(\sigma(t))dt + S^\zeta\sigma(t)dW_0, \quad \mathbb{E}\left[dW^0, dW^1\right] = \rho dt. \tag{4.65}$$

where all parameters are the usual SABR parameter and now ζ is a fixed constant which we take to be non-zero (since this reduces to our previous consideration). The inverse metric for this class of models takes the form

$$ds^2_{g^{-1}} = \frac{\sigma^2}{2}\left(S^{2\beta}dS^2 + 2\rho S^{\beta+\zeta}dSd\sigma + S^{2\zeta}d\sigma^2\right). \tag{4.66}$$

We first note that the isothermal coordinates for this metric each depend on both initial local coordinates (S,σ) and we are not able to construct them for generic model parameters. If we simplify this class of models by choosing $\zeta = \beta$ which allows us to write the inverse metric as

$$ds^2_{g^{-1}} = \frac{\sigma^2 S^{2\beta}}{2}\left(dS^2 + 2\rho dSd\sigma + d\sigma^2\right), \tag{4.67}$$

then we can diagonalize the inverse metric by choosing $x = \sqrt{2}/(1-\rho^2)(S-\rho\sigma)$, $y = \sqrt{2}\sigma$, so that it now takes the form

$$ds^2_{g^{-1}} = \frac{y^2}{2\sqrt{2}}\left[(1-\rho^2)x + \rho y\right]^{2\beta}(dx^2 + dy^2). \tag{4.68}$$

The Christoffel symbols for this metric are computable and not overly complicated; however, the corresponding geodesic equations are still quite complicated and cannot be solved exactly. We thus make the further simplifying assumption of zero correlation, i.e. $\rho = 0$. In this case, the Christoffel symbols are given by

$$\Gamma^x_{xx} = \Gamma^y_{xx} = -\Gamma^x_{yy} = \frac{\beta}{x}, \quad \Gamma^x_{xy} = \Gamma^y_{yy} = -\Gamma^y_{xx} = -\frac{1}{y}, \tag{4.69}$$

and the corresponding geodesic equations are given by

$$\ddot{x} + \frac{\beta}{x}(\dot{x})^2 + \frac{2}{y}\dot{x}\dot{y} - \frac{\beta}{x}(\dot{y})^2 = 0, \tag{4.70}$$

$$\ddot{y} + \frac{(\dot{y})^2}{y} + \frac{2}{y}\dot{x}\dot{y} + \frac{\beta}{x}(\dot{x})^2 = 0. \tag{4.71}$$

which are again coupled to the arclength parametrization constraint. There is little hope in attempting to obtain an exact solution of these equations, i.e. even after all of the simplifying assumptions that we have made, we are still unable to compute an exact expression for the model's distance function.

Thus is appears that one will need to resort to numerical methods in order to construct distance functions for extended stochastic volatility models. Even after one has produced these distance functions, there is still a considerable amount of work to be done in order to construct and asymptotic representation formula using heat kernel methods. One may be better off considering extensions of other asymptotic methods, (e.g. Andersen and Piterbarg [2]).

We now turn attention to using the exact hyperbolic distance function formula in order to compute approximations for the transition densities for two dimensional stochastic volatility models. These methods have been quite useful when applied to the SABR model Hagan [77]. Thus we first turn to reviewing the history of SABR, summarizing the known literature, and then provide additional approximation techniques.

4.3. **The History of SABR Approximations with Examples.** We now want to consider one of the most widely used stochastic volatility models, introduced in Hagan [77], called the stochastic alpha-beta-rho model (SABR). European call options cannot be priced exactly in this model. We will also consider several smile approximation functions for this model which have been constructed using a wide variety of methods. We start by specifying the model dynamics. Fix constants $\beta \in (0,1)$ and $\alpha, \rho, \nu, f > 0$ and let F_t and V_t represent asset/interest rate and volatility processes, respectively, whose time evolution is determined by

$$dF_t = V_t F_t^\beta dW_t^1, \quad dV_t = \nu V_t dW_t^2, \quad \mathbb{E}[dW_t^1, dW_t^2] = \rho dt, \tag{4.72}$$

where $F_0 = f > 0$ is the initial forward rate, $V_0 = \alpha$ is the initial volatility, and $t \in [0,T]$ for some maturity time $T > 0$. When applicable, we take the process $F(t)$ to be strongly absorbing at the zero level, i.e. if $F_t = 0$ for some $t \in [0,T)$, then $F_t = 0$ for every $t > \bar{t}$.

We note that if one takes into account market data, then there is a redundancy in the specification of the SABR parameters. In particular, note that we can write

$$dF_t = V_t F_t^\beta dW_t = V_t F_t F_t^{\beta-1} dW_t. \tag{4.73}$$

Fixing $t = 0$, this becomes

$$dF_0 = (\alpha F_0^{\beta-1})F_0 dW_0 \equiv \sigma_{IV} F_0 dW_0, \tag{4.74}$$

where here σ_{IV} is a market implied volatility (typically associated to a call with a fixed strike on a forward rate F) which can be solved for $\alpha = \sigma_{IV} F_0^{1-\beta}$; thus α is determined by the initial asset/rate value and the market implied volatility.

Note that the above SABR dynamics can be viewed as a CEV process with stochastic lognormal volatility. Alternatively, it can be viewed as the natural interpolation of the normal and lognormal stochastic volatility models. The SABR model has many nice properties, one of which is that the moments of the asset/forward process are bounded for all time (this is a property that is not shared by other popular stochastic volatility models, including the Heston Model).

In Hagan et. al. [77], the authors approximated the transition density of the random variable corresponding to the above dynamics using singular perturbation theory techniques. Essentially, these boil down to making a series of function definitions, variable changes, and simplifying assumptions that allows one to reduce the SABR transition density PDE to a sourced heat equation with delta function initial data, which can be solved explicitly. Then, one uses this solution to construct an approximate price for a European call under SABR dynamics which is then converted to an implied volatility function, i.e. a smile, by equating option prices and only keeping terms that are linear in maturity time.

The resulting well known approximation formula for the price of a European call with strike K is given by

$$\sigma_H = \frac{\alpha \left(1 + \left[\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(fK)^{1-\beta}} + \frac{1}{4}\frac{\rho\beta\nu\alpha}{(fK)^{(1-\beta)/2}} + \frac{2-3\rho^2}{24}\nu^2\right] T + \cdots\right)}{(fK)^{(1-\beta)/2}\left\{1 + \frac{(1-\beta)^2}{24}\ln^2 f/K + \frac{(1-\beta)^4}{1920}\ln^4 f/K + \cdots\right\}} \left(\frac{z}{\chi(z)}\right), \tag{4.75}$$

where we set

$$z = \frac{\nu}{\alpha}(fK)^{(1-\beta)/2}\ln f/K, \quad \chi(z) = \ln\left(\frac{\sqrt{1-2\rho z + z^2} + z - \rho}{1-\rho}\right), \tag{4.76}$$

which is valid for all strikes except when $f = K$, i.e. when we are at the money. In this case, one can take the limit $f \to K$ of equation (4.75) to recover the simpler formula West [168],

$$\sigma_H^{ATM}(f) = \frac{\alpha}{f^{1-\beta}}\left(1 + \left(\frac{(1-\beta)^2}{24}\frac{\alpha^2}{f^{2-2\beta}} + \frac{\rho\beta v\alpha}{4f^{1-\beta}} + \frac{2-3\rho^2}{24}v^2\right)T\right). \tag{4.77}$$

One key point to note about the construction of this formula is that is was derived using methods that are perturbative in both the maturity time T and the strike K; in addition, its construction assumes that $\nu << 1$. The effects of these conditions can be seen by noting that when the formula is compared against Monte Carlo results, σ_H becomes increasingly inaccurate as T or $|K - K_{ATM}|$ increase and especially as both simultaneously become large. The approximation σ_H is known to overestimate the implied volatility for options that are far out of the money.

Another issue with this formula is that it does not reproduce the correct zeroth order limit for the implied volatility in the case that $\beta \to 1$. In Obloj [129], the author used methods similar to those in Berestychi et. al. [16], [17] to construct a new implied volatility formula that generally only slightly differs from σ_H. The approximate implied volatility formula given in Obloj [129], which reproduces the correct implied volatility in the $\beta \to 1$ limit is similar to those given for $\beta < 1$ by

$$\sigma_{Ob} = \sigma_0(1 + \sigma_1 T), \tag{4.78}$$

where the zeroth order term is then given by

$$\sigma_0 = \frac{\nu x}{\ln\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right)}, \quad z = \frac{\nu}{\alpha}\frac{f^{1-\beta}-K^{1-\beta}}{1-\beta}, \tag{4.79}$$

and the first order factor is

$$\sigma_1 = \frac{(\beta-1)^2}{24}\frac{\alpha^2}{(fK)^{1-\beta}} + \frac{1}{4}\frac{\rho\nu\alpha\beta}{(fK)^{(1-\beta)/2}} + \frac{2-3\rho^2}{24}\nu^2, \tag{4.80}$$

where we have set $x = \ln(f/K)$. We now consider two smile plots to demonstrate the differences between these two formulas. We first take the initial rate value to be $f = 0.08$, along with initial volatility $\alpha = 0.01$, volvol $\nu = 0.2$, and model parameters $\beta = 0.6$, $T = 1$, $\rho = -0.8$ and plot the result in Figure 20. Here we plot σ_H in the lower graph and σ_{Ob}

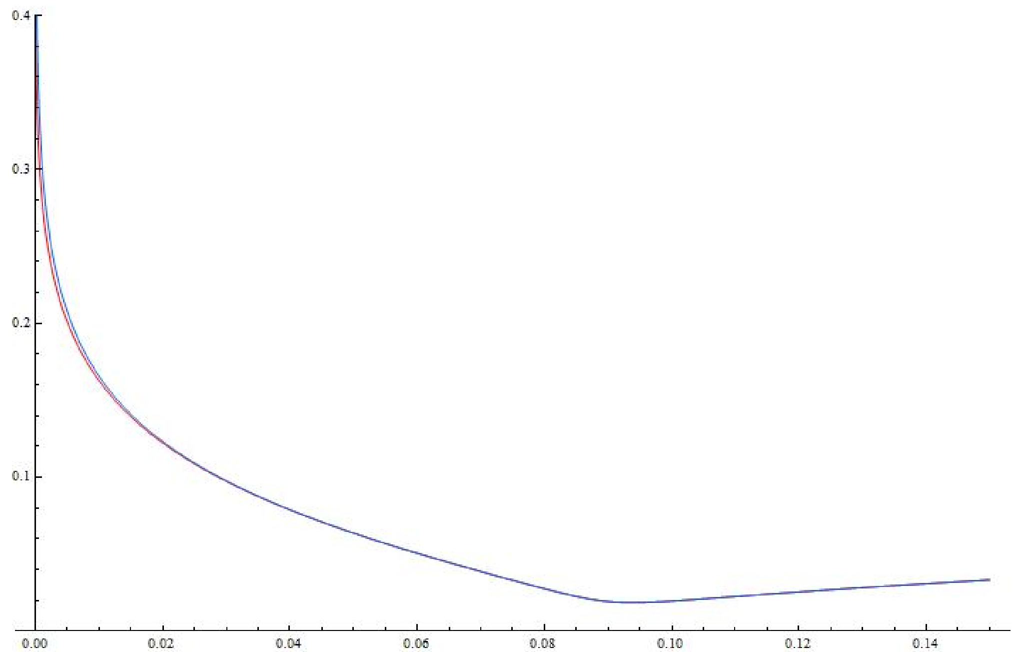

FIGURE 20. σ_H vs σ_{Ob}, $f = 0.08$, $\alpha = 0.01$, $\beta = 0.6$, $\nu = 0.2$, $\rho = -0.8$, and $T = 1$.

in the other plot. We see here that the two formulas have a high level of agreement for strikes near the money; in particular, they are indistinguishable in this range in both plots. It is only when we look at small K (i.e. the extreme left wing) when we see that the two approximations begin to noticeably differ. This is to be expected generically, as the two formulas only differ when the variables

$$x_1(S) = \frac{S^{1-\beta}-K^{1-\beta}}{1-\beta}, \quad x_2(S) = \frac{S-K}{(SK)^{\beta/2}}, \tag{4.81}$$

have a numerically significant difference. This only happens when $S - K$ is large, i.e. for strikes highly out of the money. Since σ_{Ob} produces the correct $\beta \to 1$ limit and it has analytic complexity comparable to Hagan's formula (i.e. it is not significantly more computationally expensive to evaluate), one should always prefer σ_{Ob} over σ_H.

We also note that when we now decrease β, we enlarge the difference between the two approximations. We demonstrate this fact in Figure 21, where we keep all the same parameters as in the previous except that we decrease β to $\beta = 0.4$. Now note that on the left

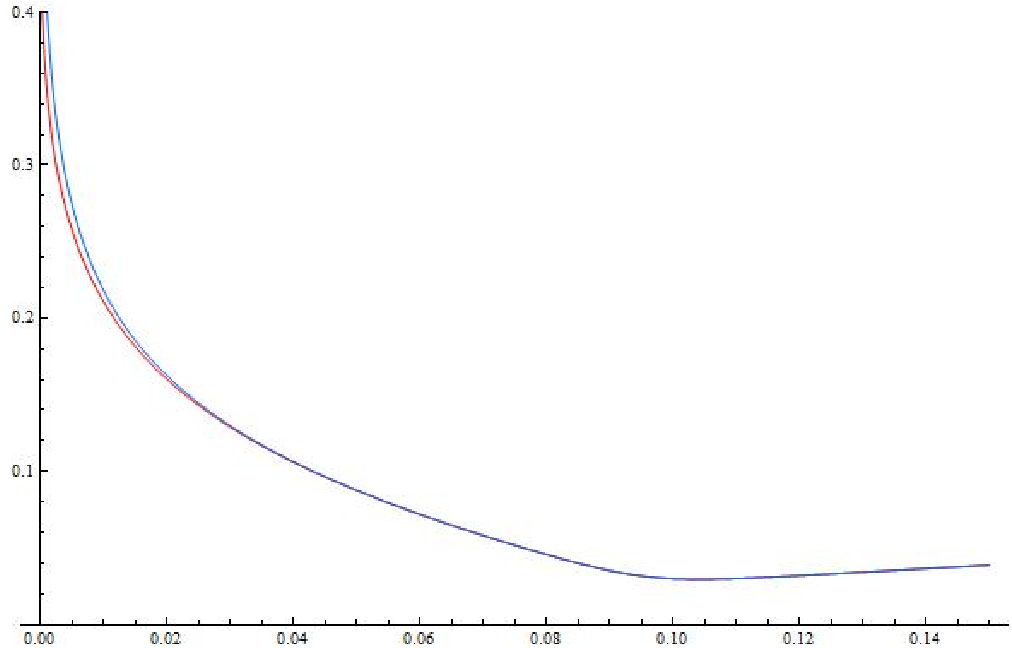

FIGURE 21. σ_H vs σ_{Ob}, $f = 0.08$, $\alpha = 0.01$, $\beta = 0.4$, $\nu = 0.2$, $\rho = -0.8$, and $T = 1$.

wing the error is slightly larger than in the previous case where $\beta = 0.6$. We have yet to see an example with reasonable model parameter values where these two formulas differ significantly and would always recommend using σ_{Ob} over σ_H since it produces the correct limiting implied volatility in the $\beta \to 1$ case as the two formulas are comparable in complexity and hence can be computed numerically in similar times.

Unfortunately, σ_{Ob} still suffers from asymptotic strike estimation degeneration. This leads us to consider another smile approximation formula constructed by different, and arguably generically superior, perturbation techniques. We start by noting that the seminal work of Hagan and Lesniewski [79] was the first to point out that there is a correspondence between the SABR model and hyperbolic geometry. In this paper, the authors study the Green's function transition density PDE (which is equivalent to the call PDE but where the initial data is replaced with a delta function). If one can obtain Green's function for a given linear operator with some specified boundary data, then it is possible to write down the price of any European option by integrating the Green function against the payoff. This PDE technique is widely employed in classical electrodynamics (c.f. Jackson [88]), amongst other subjects, to construct solution representation formulas for PDE with arbitrary initial data. The authors also point out the very interesting and important fact that SABR dynamics are just those of a two dimensional Brownian motion on the hyperbolic half-plane. We will not consider any of the formulas given in Hagan et. al. [79] and now turn attention to two recent extensions of this work.

There are several extensions of [79] (Hagan et. al.) where different authors have used the nice geometric properties of the SABR model in order to construct more accurate implied volatility formulas (e.g. Avramidi [11], Berestycki [15], Bourgade and Croissant [27], Fouque [61]). More recently Rebonato and Barjaktarevic [139] and Doust [46] have constructed improved version of Hagan's original SABR approximation formula. First, we note that in Labordere [99], the author developed an application of the heat kernel method that could be used to approximate the smile in SABR. This method involved computing the distance function and $\mathcal{P}$ exactly, but approximating the a_i heat kernel coefficients with their diagonal formulas. The a_i approximation introduces additional error for strikes on the wings. This was improved in Paulot [134] where the author is able to compute a_1 exactly (at least in principle) and, to the best of our knowledge, constructs the most explicit first order SABR approximation formula in the current literature. We outline this in more detail in the subsequent section and consider some possible improvements. For now, we state this first order in time approximation formula for SABR:

$$\sigma_P = \sigma_0 + \sigma_1 T, \tag{4.82}$$

$$\sigma_0 = \frac{\nu \ln(K/f)}{\ln\left(\frac{\sqrt{\alpha^2+2\rho\alpha\nu q+\nu^2 q^2}+\rho\alpha+q\nu}{(1+\rho)\alpha}\right)}. \tag{4.83}$$

with first order term given by

$$\sigma_1 = -\nu^2 \sigma_0 \frac{\tilde{C} + \ln\left(\sigma_0 \nu^{-1}\sqrt{Kf}\right)}{\ln\left(\frac{\sqrt{\alpha^2+2\rho\alpha\nu q+\nu^2 q^2}+\rho\alpha+q\nu}{(1+\rho)\alpha}\right)^2}, \tag{4.84}$$

where

$$\tilde{C} = -\frac{1}{2}\ln\left(\alpha\nu^{-1} f^{\beta} V_{min} K^{\beta}\right) - \frac{\rho\beta}{(1-\beta)\sqrt{1-\rho^2}}[G(t_2) - G(t_1)], \tag{4.85}$$

for $\beta < 1$, where we have set

$$q = \frac{K^{1-\beta} - f^{1-\beta}}{1-\beta}, \quad V_{min} = \sqrt{\alpha^2\nu^{-2} + 2\rho\alpha\nu^{-1}q + q^2}, \tag{4.86}$$

and when $(a + bX)^2 \neq (1-\beta)^2 R^2$,

$$G(t) = \arctan(t) - \frac{a+bX}{\sqrt{(a+bX)^2 - (1-\beta)^2R^2}} \arctan\left(\frac{cR + t(a + b(X-R))}{\sqrt{(a+bX)^2 - (1-\beta)^2R^2}}\right) \tag{4.87}$$

$$t_1 = \sqrt{\frac{R - x_1 + X}{R + x_1 - X}}, \quad t_2 = \sqrt{\frac{R - x_2 + X}{R + x_2 - X}}, \tag{4.88}$$

where we have used

$$X = \frac{x_2^2 - x_1^2 + y_2^2 - y_1^2}{2(x_2 - x_1)}, \quad R = \sqrt{y_1^2 + (x_1 - X)^2}, \tag{4.89}$$

and

$$x_1 = -\frac{\rho\alpha}{\nu\sqrt{1-\rho^2}}, \quad y_1 = \frac{\alpha}{\nu}, \quad x_2 = \frac{q - \rho V_{min}}{\sqrt{1-\rho^2}}, \quad y_2 = V_{min}. \tag{4.90}$$

$$a = f^{1-\beta}, \quad b = (1-\beta)\sqrt{1-\rho^2}, \quad c = (1-\beta)\rho. \tag{4.91}$$

Note if the argument under the square roots in $G(t)$ is negative, the formula for G is still real valued. We give a formula for G in the case of equality in another section, and note that it is not really useful here since the chance of equality occurring is nearly negligible. The main advantage of this formula is that its construction does not involve perturbation theory about the at the money strike; hence, it has better strike asymptotic properties than the previous formulas that we have considered.

All the approximation formulas tend to be nearly numerically equivalent with Monte Carlo simulations for short maturity times. In particular, preliminary tests seem to indicate that σ_P works quite well in approximating Monte Carlo results for mid to long maturity options for "small" initial asset/forward values when the other model parameters take "reasonable" values. Before probing the full parameter space, we offer one example that demonstrates the advantages of σ_P with σ_H. This is demonstrated in Figure 28 where here we consider fixed income parameter values $S = 0.0801$, $\beta = 0.4$, $\alpha = 0.1$, $\rho = -0.33$, $\nu = 0.25$, and let the maturity times be taken from the set $T \in \{20, 25, 30, 35, 40, 45\}$. We plot results from Monte Carlo simulations with $N = 10^8$ runs. We took the total number of time steps in each case to be $\{1500, 1750, 2000, 2250, 2500, 2750\}$, e.g. for $T = 20$ we took 1500 time steps, for $T = 25$ we took 1750, etc. We do this to keep error due to path step size roughly uniform. In our simulations, the standard Monte Carlo error (c.f Brigo and Mercurio [23]) was less than one volatility basis point for each of the curves we generated, i.e. the graphs represent the true smiles. We plot Hagan's formula σ_H directly above Paulot's formula σ_P. One should note how well Paulot's formulas approximate the Monte Carlo results, even in the $T = 30$ and $T = 35$ cases (which are some of the longest tenors one finds for fixed income instruments). The price one pays for this increased accuracy is that Paulot's formula takes slightly longer than Hagan's to evaluate. Paulot's formula appears to be preferable to Hagan's, and generally speaking, σ_P is a "better" smile approximation than σ_H.

We now consider a more recent approximation from Wu [169]. We first summarize the result. The author gives a second order approximation for the transition density which was created using a different series expansion method than those we have considered thus far. Let $p_2(T-t, f, \alpha, F, A)$ be the probability transition density of a forward f with initial stochastic volatility α at time t to evolve to a future value F with volatility A at time T. The formula we wish to consider is given by

$$p_2(T-t, f, \alpha, F, A) = \frac{1}{\nu T F^{\beta} A^2}\left(1 + \frac{\nu\sqrt{T}}{2(\rho^2-1)}\left(a_{11} + \frac{a_{10}}{\tau}\right)\right) \tag{4.92}$$

$$+\frac{\nu^2 T}{24(1-\rho^2)^2}\left(a_{23}\tau + a_{22} + \frac{a_{21}}{\tau} + \frac{a_{20}}{\tau^2}\right)\hat{p}_0, \tag{4.93}$$

where we define

(4.94)

$$\hat{p}_0 = \frac{1}{2\pi\sqrt{1-\rho^2}(1-t/T)}\exp\left(-\frac{\left(\frac{f^{1-\beta}-F^{1-\beta}}{\alpha(1-\beta)\sqrt{T}}\right)^2 - 2\rho\left(\frac{f^{1-\beta}-F^{1-\beta}}{\alpha(1-\beta)\sqrt{T}}\right)\left(\frac{\ln(\alpha/A)}{\nu\sqrt{T}}\right) + \left(\frac{\ln(\alpha/A)}{\nu\sqrt{T}}\right)^2}{2(1-\rho^2)(1-t/T)}\right),$$

where the p_0 variables are defined by

$$\tau = \frac{T-t}{T}, \quad u = \frac{f^{1-\beta} - F^{1-\beta}}{\alpha(1-\beta)\sqrt{T}}, \quad v = \frac{\ln(\alpha/A)}{\nu\sqrt{T}}, \tag{4.95}$$

and the remaining parameters are given by

$$a_{11} = -\beta F^{\beta-1} A \nu^{-1}(u - \rho v), \tag{4.96}$$

$$a_{10} = u^2 v - \rho u v^2, \tag{4.97}$$

$$a_{23} = \rho^4(20-6b) - 12\rho^3 a - \rho^2(28 - 3a^2 - 12b) + 12a\rho + 8 - 3a^2 - 6b, \tag{4.98}$$

$$a_{22} = u^2[3a^2 - 12a\rho + 6b(\rho^2-1) + 2\rho^2 + 10] - 2uv[\rho^3(2+3b) + \rho^2(3c - 9a) \tag{4.99}$$

$$+\rho(10 + 3a^2 - 3b) - (3a + 3c)] + v^2[(2 + 3(a-2\rho)^2)\rho^2 + 6c\rho(\rho^2-1) - 2], \tag{4.100}$$

$$a_{21} = u^4 + v^4\rho^2 + u^3 v(8\rho - 6a) + uv^3(8\rho^3 - 6a\rho^2) + u^2v^2(12a\rho - 14\rho^2 - 4), \tag{4.101}$$

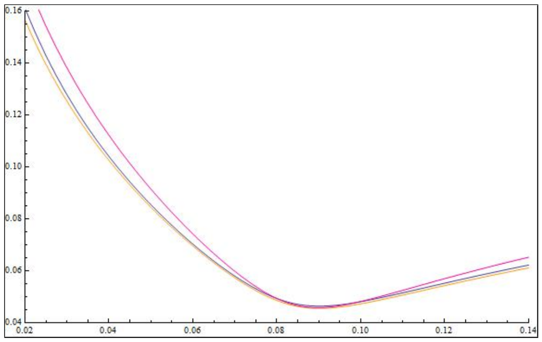

FIGURE 22. $T = 20$

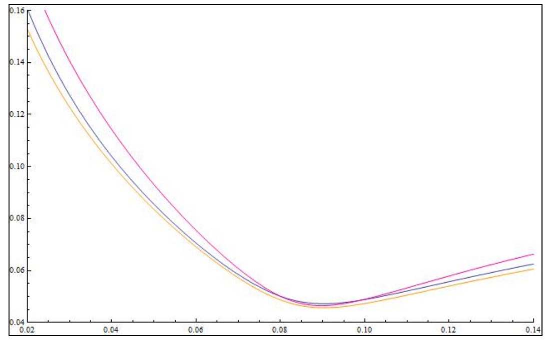

FIGURE 23. $T = 25$

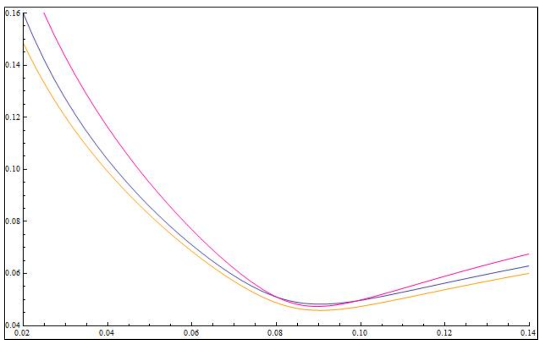

FIGURE 24. $T = 30$

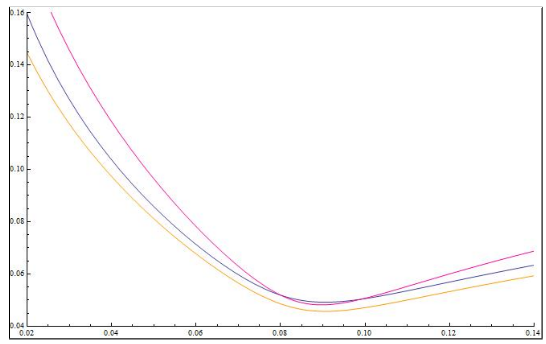

FIGURE 25. $T = 35$

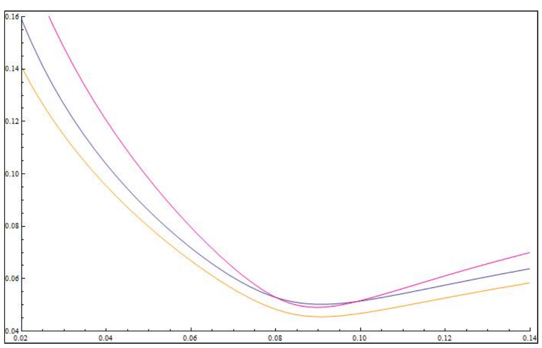

FIGURE 26. $T = 40$

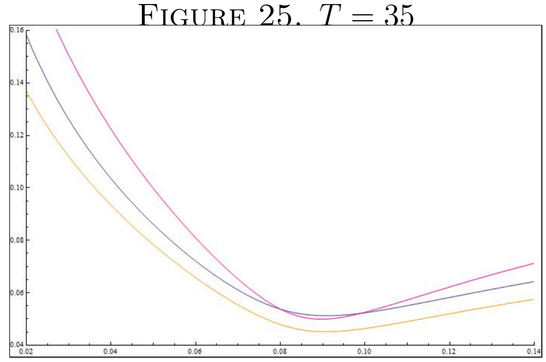

FIGURE 27. $T = 45$

FIGURE 28. For SABR model parameters $S = 0.081$, $\beta = 0.4$, $\alpha = 0.1$, $\rho = -0.33$, $\nu = 0.25$ and six different maturities mentioned above, we plot Monte Carlo results in orange, σ_P in blue and σ_H in purple.

$$a_{20} = 3u^4v^2 - 6u^3v^3\rho + 3u^2v^4\rho^2, \tag{4.102}$$

$$a = 2\rho + \beta F^{\beta-1}A\nu^{-1}, \quad b = 2 + \beta(1-\beta)F^{2(\beta-1)}A^2\nu^{-2}, \quad c = \beta F^{\beta-1}A\nu^{-1}, \tag{4.103}$$

We note that this formula can be very useful for gaining qualitative intuition on how varying the SABR parameters effects the joint transition density. For instance in Figure 29, we fix parameter values $\rho = -0.7, f = 100, \alpha = 0.6, \beta = 0.5, \nu = 0.4, T = 0.25$, $t = 0$, and let $\beta \in \{0.4, 0.5, 0.6, 0.7\}$. Note that the graphs of the joint transition density simultaneously

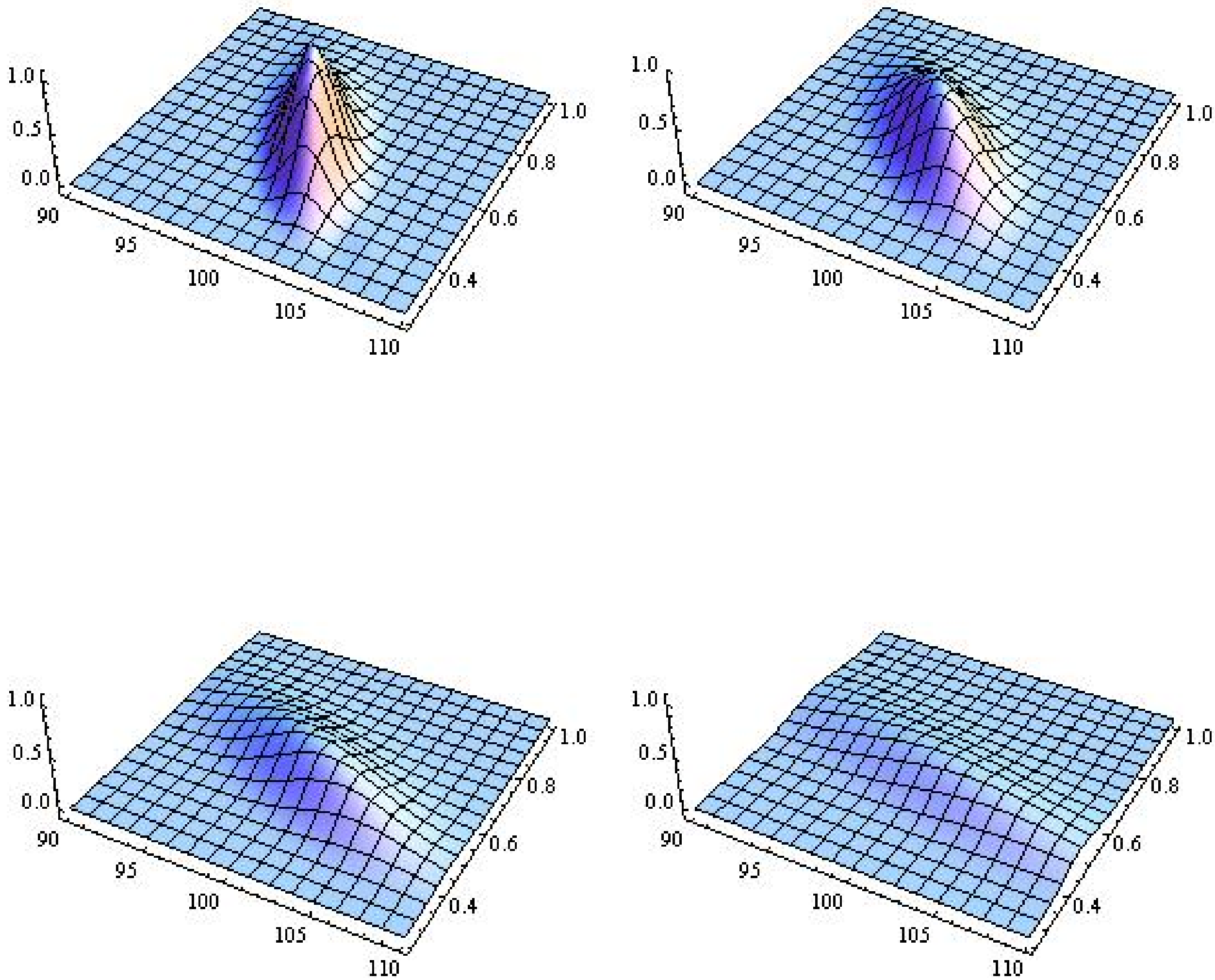

FIGURE 29. Here we plot σ_{QX} for $\rho = -0.7, f = 100, \alpha = 0.6, \beta = 0.5, \nu = 0.4, T = 0.25$, and $t = 0$ for four values of $\beta \in \{0.4, 0.5, 0.6, 0.7\}$.

rotate and disperse as we increase β. We note that there seem to be several limitations of this approximation formula, many of which are accounted for by the author.

One of the main disadvantages of σ_{QX} is that it is computationally more intensive than the other approximations that we have considered. In particular, in order to extract the implied volatility σ_{QX}, formula from the density p_2, one needs to solve the following equation

$$C_{BSM}(\sigma_{QX}) = \int [f - K]^+ p_2, \tag{4.104}$$

for each strike where an implied volatility is needed. To do this one, needs to numerically evaluate the right-hand integral, and then apply a root finder to find the associated implied volatility σ_{QX}. This takes significantly longer to compute relative to the other approximations we have considered. One could use additional techniques to convert the transition density approximation formula to an explicit formula for the implied volatility. In particular, one could expand both sides of the above equation to second order in T and equate coefficient functions in order to construct an implied volatility formula. This eliminates the need for root finding and numerical integration at the possible expense of making the formula slightly more inaccurate.

In addition to the computational expense of this formula, we have often found in many examples that we have considered that it under-performs when compared with σ_H and σ_P relative to Monte Carlo. We consider one such example now in Figure 30.

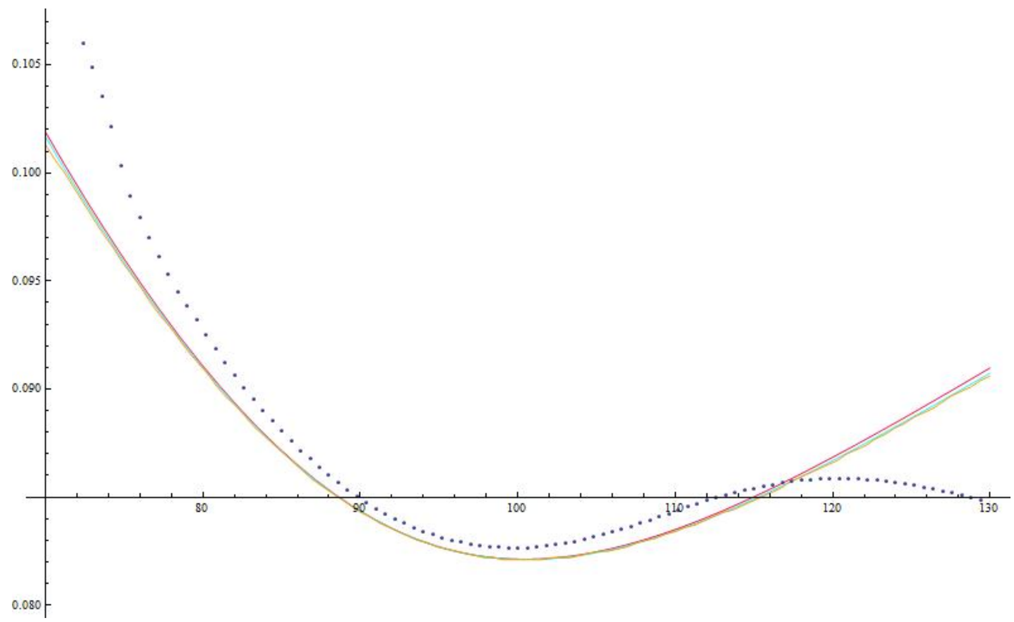

FIGURE 30. Comparison of σ_H, σ_P and Monte Carlo results with 5×10^7 paths using the Milstein with 300 time steps and model parameters $f = 95$, $\beta = 0.5$, $\alpha = 0.8$, $\rho = 0$, $\nu = 0.3$, and $T = 1$.

Here we plot σ_H and σ_{Ob} (which are indistinguishable) and σ_P which is closest to the dotted curve. We plot σ_{QX} in the dotted curve. We plot our Monte Carlo simulation results. Here we have chosen parameter values $f = 95$, $\beta = 0.5$, $\alpha = 0.8$, $\rho = 0$, $\nu = 0.3$, and $T = 1$ following an example in Wu [169]. For our Monte Carlo, we used the Milstein scheme with 300 time steps and 5×10^7 paths. The Monte Carlo standard error was less than one basis point for values of strike greater than $K = 90$. The error becomes larger for smaller strike values. We see from the graph that σ_{QX} is not as accurate as any of the other approximations for strikes near the money, which tend to work quite well for the relatively low maturity time $T = 1$ independent of reasonable choices for the other model parameters, and significantly degenerate for strikes on the wings. In other examples we have considered, increasing the maturity time tends to exacerbate this issue. Keeping these issues in mind, we have not been able to find a single example where this approximation outperforms the other approximations we have mentioned, although we do not claim the nonexistence of such a case.

We now consider a two long maturity examples where we compare σ_H and σ_P to Monte Carlo results in order to better understand the differences between these formulas in the context of different asset classes. We first consider an equity example in Figure 31 where we plot Monte Carlo results, σ_P, σ_{Ob} and σ_H (which are not distinguishable). Note here that globally σ_P approximates Monte Carlo results better than σ_H although there are some regions near the money where this is not the case. We finally consider another example in

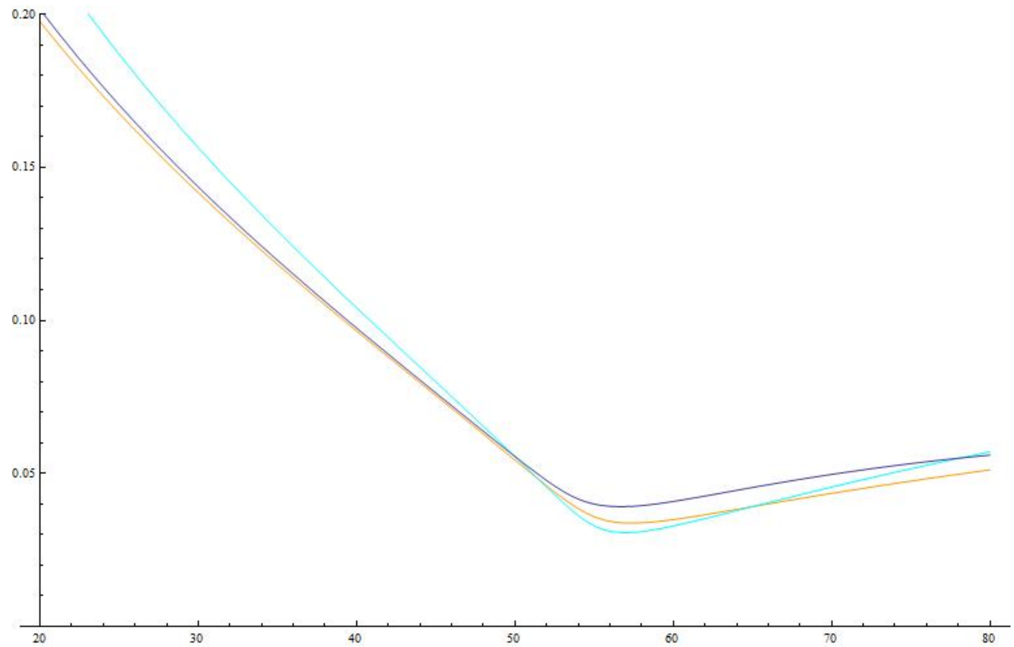

FIGURE 31. Implied Volatility plots with color conventions specified above and model parameters $f = 50$, $\beta = 0.6$, $\alpha = 0.3$, $\rho = -0.9$, $\nu = 0.6$, $T = 10$.

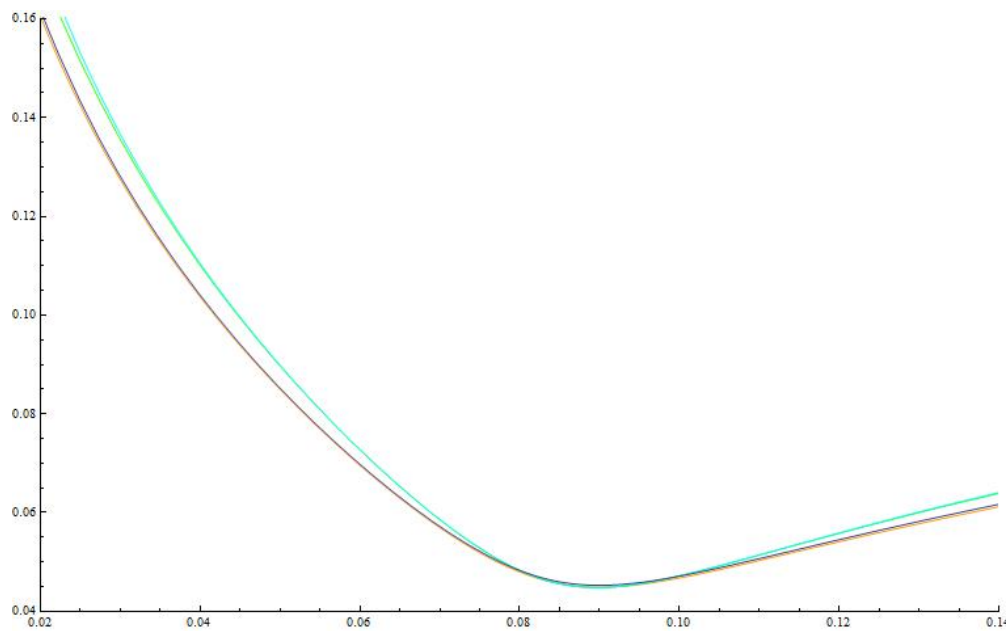

FIGURE 32. Implied volatility plots. $f = 0.0801$, $\beta = 0.4$, $\alpha = 0.01$, $\rho = -0.33$, $\nu = 0.25$, $T = 15$.

Figure 32 Note how well σ_P agrees with the Monte Carlo result in this case. We now review the construction of σ_P.

4.4. **Heat Kernel Methods for the SABR Model.** We now show how the heat kernel method can be applied to the stochastic alpha-beta-rho (SABR) model.

Recall that the dynamics for this model are given by

$$dF_t = V_t F_t^{\beta} dW_t^1, \quad dV_t = \nu V_t dW_t^2, \quad \mathbb{E}\left[dW_1, dW_2\right] = \rho dt. \tag{4.105}$$

This fact implies that Δ is now non-trivial, which complicates computations for the heat kernel coefficients a_k. Moreover, the geodesics are no longer subsets of $\mathbb{R}^1$, and one actually needs to solve the geodesic equations in order to construct them. Moreover, the line integrals relevant to the construction of an approximate smile formula will now have to be evaluated over $\mathbb{R}$, which will complicate certain calculations.

This is a small literature on heat kernel methods applied to the SABR model. The first such paper along these lines was Labordere [98]. The main limitation of this work is that the author uses methods that are both perturbative in maturity time T and strike K. A subsequent work Paulot [134] resolves these issues and the author indicates how one can construct an exact formula for $a_1(x, y)$ instead of approximate it by its diagonal element $a_1(x, x)$ (as was done in Paulot [134]). A recent preprint Forde [60] provides rigorous arguments for many of the constructions in Paulot [134] for a general class of stochastic volatility models in the case of zero correlation.

Our aim now is to use techniques from these works to construct a second order approximation for the transition density. We will closely follow portions of Paulot [134]. We fill in details as well as provide extensions of the author's results. We start by considering the parabolic equation for the SABR transition density function.

The SABR covariance matrix is given by $\Sigma^{ij} = \sigma^i_k \sigma^j_l \rho^{kl}$ where $\rho^{FF} = \rho^{VV} = 1$, $\rho^{FV} = \rho^{VF} = \rho$ for some $\rho > 0$. and takes the explicit form

$$\Sigma^{ij} = \begin{pmatrix} \sigma^{FF} & \sigma^{FV} \\ \sigma^{VF} & \sigma^{VV} \end{pmatrix} = \begin{pmatrix} (\sigma^F_F)^2 & \sigma^F_F \sigma^V_V \rho \\ \sigma^V_V \sigma^F_F \rho & (\sigma^V_V)^2 \end{pmatrix} = \begin{pmatrix} V^2 F^{2\beta} & \nu V^2 F^\beta \rho \\ \nu V^2 F^\beta \rho & \nu^2 V^2 \end{pmatrix}. \tag{4.106}$$

The associated second order parabolic equation in terms of our initial data coordinates $\alpha = V(0)$ and $f = F(0)$ is given by

$$\partial_\tau p = \frac{1}{2}\Sigma^{ij} p_{ij} = \frac{1}{2}\alpha^2 f^{2\beta} p_{ff} + \nu\rho\alpha^2 f^\beta p_{f\alpha} + \frac{1}{2}\nu^2\alpha^2 p_{\alpha\alpha}, \tag{4.107}$$

where $p = p(\tau, \alpha, f, V, F)$ and $p(0, \alpha, f, V, F) = \delta(V - \alpha)\delta(F - f)$.

We will show via changing coordinates that the metric associated to Σ is just the hyperbolic metric in disguise. Define a forward coordinate $q = \int_f^F \frac{du}{u^\beta}$. Then in the q coordinate, Σ takes the form

$$\Sigma^{ij}_{(q,\alpha)} = \frac{\partial q^i}{\partial x^k}\frac{\partial q^j}{\partial x^l}\Sigma^{kl} = \alpha^2 \begin{pmatrix} 1 & -\rho\nu \\ -\rho\nu & \nu^2 \end{pmatrix}, \tag{4.108}$$

where here we denote the new coordinates by $q^i = (q, \alpha)$ and the old ones by $x^i = (f, \alpha)$. We now make another change of coordinates. Define $x = \sqrt{2}(\nu q + \rho\alpha)/(\nu\sqrt{1-\rho^2})$ and $y = \sqrt{2}\alpha/\nu$, and note in these coordinates that

$$\Sigma^{ij}_{(x,y)} = 2y^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \to g^{ij} = \frac{1}{2}\Sigma^{ij} = y^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{4.109}$$

so that

$$g_{ij} = \frac{1}{y^2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{4.110}$$

which is the two-dimensional metric for the Poincare model of hyperbolic geometry.

We now work out some intrinsic geometric quantities associated with this geometry, in particular, the distance function, geodesics, van-Vleck determinant, and curvature. A few of these constructions follow from previous general geometric results that we have considered, but we repeat them hear for the sake of cohesion.

We start by finding the geodesic equations. The only non-trivial Christoffel symbols associated to g are $\Gamma^x_{xy} = \Gamma^y_{yy} = -\Gamma^y_{xx} = -\frac{1}{y}$. Thus if we let a generic geodesic be given by $\gamma(t) = (x(t), y(t))$, then the geodesic equation becomes

$$0 = \ddot{x} - \frac{2}{y}\dot{x}\dot{y}, \quad 0 = \ddot{y} + \frac{1}{y}\left[(\dot{x})^2 + (\dot{y})^2\right]. \tag{4.111}$$

One can solve this equation, to find that

$$(x-a)^2 + y^2 = r^2, \tag{4.112}$$

for some constants of integration a and r. In particular, x and y are coordinates that parameterize a semi-circle with radius r that is centered at the point $(a, 0)$ on the horizontal axis of the plane. If $a, r \to \infty$, then the semi-circles degenerate to straight vertical lines (i.e. semi-circles of infinite radius centered at infinity) which are the other class of geodesics of two dimensional hyperbolic space.

Now if we let $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ be two points on a geodesic, we can compute the distance between them exactly

$$d(p_1, p_2) = \int_{p_1}^{p_2} ds = \int_{p_1}^{p_2} \frac{1}{y}\sqrt{dx^2 + dy^2} = \operatorname{arccosh}\left(1 + \frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2y_1 y_2}\right). \tag{4.113}$$

The metric has constant sectional curvature $K = -1$, from which if follows that the van-Vleck determinant is $\Delta = \sinh(d)/d$ (which can either be computed directly or derived in a parallel manner using the fact that hyperbolic geometry is a symmetric space c.f. Avramidi [13] , [14]).

The metric, expressed in (f, α) coordinates, and its inverse are given by

$$g = \frac{2}{\alpha^2(1-\rho^2)\nu^2}\begin{pmatrix} \nu^2 f^{-2\beta} & -\rho\nu f^{-\beta} \\ -\rho\nu f^{-\beta} & 1 \end{pmatrix}, \quad g^{-1} = \frac{\alpha^2}{2}\begin{pmatrix} f^{2\beta} & \nu\rho f^{\beta} \\ \nu\rho f^{\beta} & \nu^2 \end{pmatrix}, \tag{4.114}$$

Since both processes have zero drift, the connection coefficients are given by

$$A_j = \frac{1}{2} g_{ij} g^{lk} \Gamma^i_{lk}, \tag{4.115}$$

from which we can compute

$$A_f = -\frac{\beta}{2f(1-\rho^2)} = -\frac{1}{2(1-\rho^2)} d\ln C(f), \quad A_\alpha = \frac{\beta\rho f^{\beta-1}}{2\nu(1-\rho^2)} = \frac{\rho C'(f)}{2\nu(1-\rho^2)}, \tag{4.116}$$

where here $C(f) = f^\beta$, which can be expressed in the x coordinates as

$$A = \frac{C'(f)}{2\sqrt{2}\sqrt{1-\rho^2}} dx. \tag{4.117}$$

Here, we use the fact that

$$dx = -\frac{\sqrt{2} f^{-\beta}}{\sqrt{1-\rho^2}} dq + \frac{\sqrt{2}\rho}{\nu\sqrt{1-\rho^2}} d\alpha. \tag{4.118}$$

We also find that

$$Q = g^{ij}\left(\partial_i A_j - \Gamma^k_{ij} A_k + A_i A_j\right) \tag{4.119}$$

$$= g^{xx}\partial_x A_x + g^{xx} A_x^2 = \frac{\beta}{4}\left[1 - \beta + \frac{\beta}{2(1-\rho^2)}\right]\alpha^2 f^{2(\beta-1)} \equiv \delta\alpha^2 f^{2(\beta-1)}, \tag{4.120}$$

noting that

$$\partial_x A_x = \frac{\beta}{2\sqrt{2}\sqrt{1-\rho^2}}(\beta-1)f^{\beta-2}\frac{\partial f}{\partial x}, \quad \frac{\partial x}{\partial f} = -\frac{\sqrt{2}}{\sqrt{1-\rho^2}}\frac{1}{f^\beta}. \tag{4.121}$$

One can also verify the equation for Q by performing the computation in the initial coordinates, f, α although all the Christoffel symbols are nontrivial in this case as well as both A_i (hence the change of variables to the standard form of hyperbolic space). Thus we wish to evaluate

$$\mathcal{P}((\alpha,f),(V,F)) = \exp\left(-\int_{(\alpha,f)}^{(V,F)} A_x dx\right) = \exp\left(-\frac{\beta}{2\sqrt{2}\sqrt{1-\rho^2}}\int_{(\alpha,f)}^{(V,F)} f^{\beta-1}\right) dx. \tag{4.122}$$

Inverting the above coordinate transformations, we note that

$$f^{\beta-1} = \left[F^{1-\beta} - \frac{(1-\beta)}{\sqrt{2}}[\sqrt{1-\rho^2}x - \rho y]\right]^{-1}. \tag{4.123}$$

If we parametrize the semi-circle geodesic by $x = a + r\cos\theta$, $y = r\sin\theta$, then we find

$$I = \int_C f^{\beta-1}dx = -r\int_{\theta_1}^{\theta_2}\frac{\sin\theta d\theta}{c_1 - c_2\cos\theta - c_3\sin\theta}, \tag{4.124}$$

where

$$c_1 = F^{1-\beta} - \frac{(1-\beta)}{\sqrt{2}}a\sqrt{1-\rho^2}, \quad c_2 = \frac{(1-\beta)}{\sqrt{2}}\sqrt{1-\rho^2}r, \quad c_3 = -\frac{(1-\beta)}{\sqrt{2}}\rho r. \tag{4.125}$$

In the case that $c_1^2 \neq c_2^2 + c_3^2$, this integral can be evaluated exactly and is given by

$$I = -\frac{r}{c_2^2+c_3^2}\Bigg[- c_3\theta + c_2\log\left[c_2\cos\theta + c_3\sin\theta - c_1\right] \tag{4.126}$$

$$+\frac{2c_1c_3}{\sqrt{c_2^2+c_3^2-c_1^2}}\text{arctanh}\left(\frac{c_3 - (c_1+c_2)\tan(\theta/2)}{\sqrt{c_2^2+c_3^2-c_1^2}}\right)\Bigg]\Bigg|_{\theta_1}^{\theta_2}. \tag{4.127}$$

If $c_1^2 = c_2^2 + c_3^2$, then for suitable choices of model parameters, the integral is given by

$$-\frac{I}{r} = \Bigg[\frac{1}{c_3\cot\theta - c_2} + \frac{c_3}{\sqrt{c_2^2+c_3^2}(c_3\cos\theta - c_2\sin\theta)} \tag{4.128}$$

$$+\frac{c_2\log\left[c_2\sin\theta - c_3\cos\theta\right] - c_3\theta - 2c_2\text{arctanh}\left[\frac{c_2+c_3\tan(\theta/2)}{\sqrt{c_2^2+c_3^2}}\right]}{c_2^2+c_3^2}\Bigg]\Bigg|_{\theta_1}^{\theta_2} \tag{4.129}$$

where by suitable choices of model parameters, we mean those such that all denominators in the above are finite. Also note that if $c_1^2 > c_2^2 + c_3^2$, then one can use the identity $\text{arctanh}(i\theta) = i\arctan(\theta)$ in tandem with the above result to produce a real valued formula for I.

We now give a partial computation of a_1 in order to give an indication of the complexity involved in achieving an exact form for this function. It is defined by

$$a_1 = \frac{1}{d}\int_C \mathcal{P}\Delta^{-1/2}\Delta^A(\Delta^{1/2}\mathcal{P}) - Qds. \tag{4.130}$$

We first compute $a_1^Q \equiv -\frac{1}{d}\int_C Q ds$. To do this, first note that

$$ds = \sqrt{\frac{dx^2 + dy^2}{y^2}} = \frac{r}{y} d\theta, \tag{4.131}$$

so that

$$I = \int_C Q ds = \frac{\delta \nu^2 r}{2} \int_C y f^{2(\beta-1)} d\theta = \frac{\delta\nu^2 r^2}{2} \int_{\theta_1}^{\theta_2} \frac{\sin\theta d\theta}{(c_1 - c_2\cos\theta - c_3\sin\theta)^2}. \tag{4.132}$$

This integral can be evaluated exactly if $c_1^2 \neq c_2^2 + c_3^2$, and it is given by

$$I = \frac{c_3^2 - c_1^2 + c_1 c_2 \sin\theta}{c_3(c_3^2 + c_2^2 - c_1^2)(c_3\cos\theta + c_2\sin\theta - c_1)} - \left.\frac{2c_2 \operatorname{arctanh}\left(\frac{c_2 - (c_1+c_3)\tan(\theta/2)}{c_2^2 + c_3^2 - c_1^2}\right)}{(c_2^2 + c_3^2 - c_1^2)^{3/2}}\right|_{\theta_1}^{\theta_2}. \tag{4.133}$$

If $c_1^2 = c_2^2 + c_3^2$, then we find

$$I = \frac{c_2c_1 - 3c_2c_1\cos(2\theta) - 2c_2c_3\cos 3\theta + 2c_3^2(2\cos(2\theta))\sin\theta + 4c_2^2\sin^2\theta^3 + 3c_3c_1\sin(2\theta)}{6c_1(c_2\cos t - c_3\sin\theta)^3}. \tag{4.134}$$

Now there are a few line integrals that need to be evaluated in order to construct a formula for a_1. In particular, we note that since $\mathcal{P}$ is parallel along C that

$$\mathcal{P}\Delta^{-1/2}\Delta_A(\mathcal{P}\Delta^{1/2}) = \mathcal{P}^{-1}\Delta_A\mathcal{P} + \Delta^{-1/2}\Delta_g\Delta^{1/2}. \tag{4.135}$$

Now the first heat kernel coefficient for the hyperbolic metric in two dimensions with a trivial connection is known, c.f. Avramidi [12], in particular

$$\frac{1}{d}\int_C \Delta^{-1/2}\Delta_g\Delta^{1/2} = -\frac{1}{8}\left[1 + \frac{1}{d}\left(\coth(d) - \frac{1}{d}\right)\right], \tag{4.136}$$

so it just remains to compute the integral of $\mathcal{P}^{-1}\Delta^i\Delta_i\mathcal{P}$ (which is more complicated than the above calculations we have considered). This was partially done in Paulot [134] up to a numerical integration; however, with some work it appears one can construct an exact formula for this integral (although the final expression will almost surely be fairly complicated).

One alternative to these complicated exact computations (along the lines of Labordere [99]) is to attempt to estimate the a_i by using their diagonal coefficients. We can compute $a_1(x,x)$ and $a_2(x,x)$ exactly:

$$a_1(x,x) = \frac{R}{6} - Q(x,x), \tag{4.137}$$

$$a_2(x,x) = \frac{1}{2}a_1(x,x)^2 - \frac{1}{6}\Delta_g Q + \frac{1}{30}\Delta_g R + \frac{1}{12}\mathcal{F}_{ij}\mathcal{F}^{ij} + \frac{1}{80}\left(R_{ijkl}R^{ijkl} - R_{ij}R^{ij}\right). \tag{4.138}$$

Now for the case of the hyperbolic metric, we note that

$$R_{ijkl}R^{ijkl} = 2R_{ij}R^{ij} = -R = 2. \tag{4.139}$$

Also, using the antisymmetry property of $\mathcal{F}_{ij}$, we have that

$$\mathcal{F}_{ij}\mathcal{F}^{ij} = g^{ik}g^{jl}(\partial_i A_j - \partial_j A_i)(\partial_k A_l - \partial_l A_k) = 2(g^{yy})^2\partial_y^2 A_x. \tag{4.140}$$

where one can directly compute

$$\partial_y^2 A_x = \frac{\rho^2\beta(1-\beta)^2}{2\sqrt{2}\sqrt{1-\rho^2}}\left(F^{1-\beta} - \frac{(1-\beta)}{\sqrt{2}}[\sqrt{1-\rho^2}x - \rho y]\right)^{-3}. \tag{4.141}$$

Now, $\Delta_g R = 0$ since the scalar curvature is constant for this model so it suffices to compute
(4.142)
$$\Delta_g Q = \frac{1}{\sqrt{g}}\partial_i(\sqrt{g}g^{ij}\partial_j Q) = f^{-4+2\beta}\alpha^2\left(f^2\nu^2 + 4f^{1+\beta}\rho\nu\alpha(\beta-1) + 3f^{2\beta}\alpha^2(\beta-1)^2\right)\delta.$$

From this, we can then use the diagonal elements to approximate the transition density by evaluating at $a_1((X_0+X)/2,(X_0+X)/2)$, where here X represents our coordinates and X_0 their initial values.

4.5. **Paulot's Conversion Technique.** We now wish to give a short discussion of a technique in Paulot [134] which can be used to transform an n-th order time approximation formula for a transition density into an $(n+1)$-st order approximation for the implied volatility. This technique is a variation of the saddle-point method (or Laplace method) for integral estimation which is adapted to the heat kernel ansatz. A rigorous version of this technique is given in Forde [60].

Although we are interested in the case of SABR, this formula works for any stochastic volatility model of the form

$$dF_t = \sigma_F(F_t,V_t)dW_t^1, \quad dV_t = \mu_V(V_t)dt + \sigma_V(V_t)dW_t^2 \tag{4.143}$$

where the Brownian motions W_t^1 and W_t^2 have correlation ρ, the initial forward has value F_0, and the volatility has initial value V_0. Moreover, it can be extended to any model for which one can approximate the joint transition density of the model's processes.

The goal is to convert a given expression for a transition density $p(F_0,F,V_0,V)$ to a corresponding approximation for the stochastic volatility model's Black implied volatility. To do this, we first note that in the case of zero interest rates, the call price under the stochastic volatility model can be written as

$$C_{SV} = \int_K^\infty \int_{\mathbb{R}} (F-K)^+ p(F_0,F,V_0,V,t,T)dVdF = \int_K^\infty (F-K)^+ p(F_0,F,t,T)dVdF \tag{4.144}$$

where we define the marginal forward density

$$p(F_0,F) = \int_{\mathbb{R}} p(F_0,F,V_0,V,t,T)dV. \tag{4.145}$$

(From now on, we suppress the dependence of all functions on their initial data F_0, V_0, and t). Following the argument in Hagan et. al. [77], this can be rewritten in the form

$$C_{SV} = (F_0-K)^+ + \int_K^\infty \int_0^T (F-K)^+ \partial_t p(K,T)dTdF. \tag{4.146}$$

The PDE satisfied by the marginal density in F is given by

$$\partial_t p = \frac{1}{2}\partial_F^2(\sigma_F^2(F,t)p(F,t)). \tag{4.147}$$

Now note that we can write
(4.148)
$$\mathbb{E}[\sigma_F^2(F,V)\delta(F,K)] = \int_{\mathbb{R}}\int_K^\infty \sigma_F^2(F,V)\delta(F-K)p(V,F)dVdF = \int_{\mathbb{R}} \sigma_F^2(F,V)p(F,T)dV,$$

so that

$$\mathbb{E}[\sigma_F^2(F,V)\delta(F,K)] = \mathbb{E}[\sigma_F^2(F,V)|F=K] = \sigma_F^2(F,t), \tag{4.149}$$

where here we note that $\sigma^2(F,t)$ is the local volatility function for the asset process where, the V dependence has been integrated out. One can then substitute this expression into the equation for the call and integrate by parts by twice to find that

$$C_{SV} = (F_0 - K)^+ + \frac{1}{2}\int_0^T \mathbb{E}[\sigma_F^2(K,V)\delta(F-K)dt]. \tag{4.150}$$

Note that everything up to this point is very similar to what was done in Appendix A of Hagan et. al. [77]. We now wish to estimate the expected variance integral

$$\mathbb{E}[\sigma_F^2(K,V)\delta(F-K)] = \int \sigma_F^2 p(K,V)dV, \tag{4.151}$$

and will use the form of the heat kernel ansatz in order to simplify this process. In particular, if we let $X_0 = (F_0, V_0)$ and $X = (F, V)$ we write,

$$p(X,t) = \frac{\sqrt{g(X)}}{(2\pi t)^{n/2}}\sqrt{\Delta(X_0,X)}\mathcal{P}(X_0,X)\exp\left(-\frac{d^2(X_0,X)}{2t}\right)\sum_k a_k t^k, \tag{4.152}$$

where $p(X,0) = \delta(X - X_0)$ with $a_0 = 0$ with all other functions are the same as in the previous section. Now note that to first order in t, we can represent

$$\sigma_F^2(K,V)p(K,V,t) = \frac{1}{2\pi t}\exp\left(-\frac{B}{t} - C - Dt + o(t)\right), \tag{4.153}$$

where

$$B = \frac{1}{2}d(F,V)^2, \quad D = -a_1(K,V), \tag{4.154}$$

$$C = -2\ln\left(\sigma_F(K,V)\right) - \frac{1}{2}\left[\ln(g(K,V)) + \ln(\Delta(F,V))\right] - \ln\mathcal{P}(K,V). \tag{4.155}$$

Here we use the "little o" convention which is defined by the following: For two functions $f(x)$ and $h(x)$, we say that $f(x) = o(h(x))$ as $x \to \infty$ if

$$\lim_{x\to\infty}\frac{f(x)}{h(x)} = 0. \tag{4.156}$$

Now we apply the approximation technique, by first noting the leading order term $B = d^2/2$ will dominate the integrand for small T values. In particular, as a function of V, $B(V)$ will be maximal near the V values that are close to the minimizers V_{min} of V (which equivalently is the V value that minimizes the distance function). Thus we define $\delta V = V - V_{min}$ and write the integrand as a Taylor series in δV. To do this, note that

$$\exp\left(-\frac{B}{t} - C\right) \approx \exp\Bigg(-\frac{1}{t}\left[B + \frac{1}{2}B''\delta V^2 + \frac{1}{6}B^{(3)}\delta V^3 + \frac{1}{24}B^{(4)}\delta V^4\right] \tag{4.157}$$

$$-\left(C + C'\delta V + \frac{1}{2}C''\delta V^2 + \frac{1}{6}C^{(3)}\delta V^3 + \cdots\right)\Bigg), \tag{4.158}$$

where all functions are evaluated at V_{min}, i.e. $B'' = B''(V_{min})$, and we note that $B'(V_{min}) = 0$ since V_{min} is chosen to be a minimizer of B by construction. As we have already seen, for a general stochastic volatility model, we cannot evaluate the distance function explicitly. However, in the case of hyperbolic geometry, (which arises in SABR as well as an additional class of stochastic volatility models which we will consider) this can be computed exactly and in fact has a relatively simple form. We can continue with the approximation to note

$$\approx \exp\left(-\frac{1}{t}B - C - \frac{1}{2t}B''\delta V^2\right)\exp\left(-\frac{1}{t}\left[\frac{1}{6}B^{(3)}\delta V^3 + \frac{1}{24}B^{(4)}\delta V^4 + \cdots\right]\right) \tag{4.159}$$

$$\times \exp\left(-C'\delta V - \frac{1}{2}C''\delta V^2 - \frac{1}{6}C^{(3)}\delta V^3 + \cdots\right) \tag{4.160}$$

$$= \exp\left(-\frac{B}{t} - C - \frac{1}{2t}B''\delta V^2\right)\left[1 - \frac{1}{6t}B^{(3)}\delta V^3 - \frac{1}{24t}B^{(4)}\delta V^4 + \frac{1}{72t^2}(B^{(3)})^2\delta V^6 + \cdots\right] \tag{4.161}$$

$$\times \left[1 - C'\delta V - \frac{1}{2}C''\delta V^2 + \frac{1}{2}(C')^2\delta V^2 + \cdots\right] \tag{4.162}$$

$$= \exp\left(-\frac{B}{t} - C - \frac{1}{2t}B''\delta V^2\right)\left[1 - \frac{1}{2}C''\delta V^2\right. \tag{4.163}$$

$$\left. + \frac{1}{2}(C')^2\delta V^2 + \frac{1}{6t}C'B^{(3)}\delta V^4 - \frac{1}{24t}B^{(4)}\delta V^4 + \frac{1}{72t^2}(B^{(3)})^2\delta V^6 + \cdots\right] \tag{4.164}$$

$$= \exp\left(-\frac{B}{t} - C - \frac{1}{2t}B''\delta V^2\right)\left[1 - \frac{1}{2}\left[C'' - (C')^2\right]\delta V^2\right. \tag{4.165}$$

$$\left. - \left[\frac{1}{24}B^{(4)} - \frac{1}{6}C'B^{(3)}\right]\frac{\delta V^4}{t} + \frac{(B^{(3)})^2}{72}\frac{\delta V^6}{t^3} + \text{odd} + \cdots\right], \tag{4.166}$$

where we ignore terms that are odd in $\delta V/t$ since they will not contribute to the integration which we will preform with this approximation. We also note that $t \approx \delta V^2$, so here we just keep terms of order δV^2. We now can use this approximation to evaluate the integral. In particular,

$$\mathbb{E}\left[\sigma_F^2(K,V)\delta(F-K)\right] = \frac{1}{2\pi t}\exp\left(-\frac{B}{t} - C - Dt\right)\left[\int_{\mathbb{R}} e^{-\frac{B''}{2t}\delta V^2} - \frac{1}{2}\left(C'' - (C')^2\right)\int_{\mathbb{R}} e^{-\frac{B''}{2t}\delta V^2}\delta V^2\right. \tag{4.167}$$

$$\left. - \frac{1}{t}\left(\frac{B^{(4)}}{24} - \frac{B^{(3)}C'}{6}\right)\int_{\mathbb{R}} e^{-\frac{B''}{2t}\delta V^2}\delta V^4 + \frac{(B^{(3)})^2}{72t^2}\int_{\mathbb{R}} e^{-\frac{B''}{2t}\delta V^2}\delta V^6 + o(t)\right], \tag{4.168}$$

again where all functions are evaluated at $V = V_{min}$ and δV is the integration variable in each integral. Computing the integrals and simplifying yields

$$= \frac{1}{\sqrt{2\pi t B''}}\exp\left(-\frac{B}{t} - C - Dt\right)\left[1 - \frac{1}{2}\left(C'' - (C')^2\right)\frac{t}{B''}\right. \tag{4.169}$$

$$\left. - \left(\frac{B^{(4)}}{24} - \frac{B^{(3)}C'}{6}\right)\frac{3t}{(B'')^2} + \frac{(B^{(3)})^2}{72}\frac{15t}{(B'')^3} + o(t)\right]. \tag{4.170}$$

Finally, to simplify the task of equating this to a similar formula for the Black implied volatility. We note that it can be written in the form

$$\mathbb{E}\left[\sigma_F^2(K,V)\delta(F-K)\right] = \frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{B}{t} - \tilde{C} - \tilde{D}t + o(t)\right), \tag{4.171}$$

where we set

$$\tilde{C} = C + \frac{1}{2}\ln(B''), \tag{4.172}$$

$$\tilde{D} = D + \frac{1}{2B''}\left[C'' - C^2 + \frac{1}{4}\frac{B^(4)}{B''} - \frac{B^{(3)}}{B''}C' - \frac{5}{12}\left(\frac{B^{(3)}}{B''}\right)^2\right], \tag{4.173}$$

which can be verified by differentiating, substituting, and expanding exponentials. Now, we can substitute this formula into our expectation formula and compute the time integral to find

$$\frac{1}{2}\int_0^T \mathbb{E}\left[\sigma_F^2(K,V)\delta(K-V)\right]dt = \frac{1}{2\sqrt{2\pi T}}\int_0^T \exp\left(-\frac{B}{t} - \tilde{C} - \tilde{D}t + o(t)\right)dt \tag{4.174}$$

(4.175)

$$= \frac{e^{-\tilde{C}}}{\sqrt{2}}\left[\sqrt{\frac{T}{\pi}}e^{-\frac{B}{T}} - \sqrt{B}\mathrm{erfc}\left(\frac{\sqrt{B}}{T}\right) - \frac{\tilde{D}}{3}\left(\sqrt{\frac{T}{\pi}}(T-2B)e^{-\frac{B}{T}} + 2B^{3/2}\mathrm{erfc}\left(\sqrt{\frac{B}{T}}\right) + o\left(T^{5/2}e^{-B/T}\right)\right)\right],$$

where here we have defined

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-y^2}dy = 2\Phi(-\sqrt{2}x). \tag{4.176}$$

To see that this is indeed true, note that if we differentiate the right hand side with respect to T that the result is just $e^{-B/T}(1-\tilde{D}T)/(2\sqrt{\pi T})$, so the above holds if

$$\frac{1}{2\sqrt{\pi T}}\frac{\partial}{\partial T}\int_0^T \exp\left(-\frac{B}{t} - \tilde{D}t + o(t)\right)dt = \frac{e^{-\frac{B}{T}}(1-\tilde{D}T)}{2\sqrt{\pi T}} + o\left(T^{3/2}e^{-B/T}\right). \tag{4.177}$$

But the left hand side can be approximated as

$$\frac{1}{2\sqrt{\pi T}}\exp\left(-\frac{B}{T} - \tilde{D}T + o(T)\right) = \frac{1}{2\sqrt{\pi T}}\exp\left(-\frac{B}{T}\right)\left[1 + \tilde{D}T + o(T)\right]. \tag{4.178}$$

Now note that

$$\frac{1}{\sqrt{T}}\exp\left(-\frac{B}{T}o(T)\right) \approx o(T^{3k}e^{-B/T}), \tag{4.179}$$

and thus we have equality up to $o(T^{5/2}e^{-B/T})$.

Now we are only interested in the case of short maturities, so we will expand erfc about $t = \infty$. To do this, just expand $\mathrm{erfc}(1/x)$ about $x = 0$ and note that the resulting Taylor series is given by

$$\mathrm{erfc}(x) = \frac{e^{-x^2}}{x\sqrt{\pi}}\left(1 - \frac{1}{2x^2} + \frac{3}{4x^4} + o\left(\frac{1}{x^4}\right)\right), \tag{4.180}$$

which implies

$$\mathrm{erfc}\left(\sqrt{\frac{B}{T}}\right) = \sqrt{\frac{T}{B\pi}}e^{-\frac{B}{T}}\left(1 - \frac{T}{2B} + \frac{3T^2}{4B^2} + o\left(\frac{t^2}{B^2}\right)\right). \tag{4.181}$$

Substituting this into our previous formula, we find that

(4.182)

$$= \frac{e^{-\tilde{C}}}{\sqrt{2}}\left[\sqrt{\frac{T}{\pi}}e^{-\frac{B}{T}} - \sqrt{B}\mathrm{erfc}\left(\frac{\sqrt{B}}{T}\right) - \frac{\tilde{D}}{3}\left(\sqrt{\frac{T}{\pi}}(T-2B)e^{-\frac{B}{T}} + 2B^{3/2}\mathrm{erfc}\left(\sqrt{\frac{B}{T}}\right) + o\left(T^{5/2}e^{-B/T}\right)\right)\right]$$

(4.183)
$$= \frac{e^{-\tilde{C}}e^{-B/T}\sqrt{T}}{\sqrt{2\pi}}\left[1 - \left(1 - \frac{T}{2B} + \frac{3T^2}{4B} + \cdots\right) - \frac{\tilde{D}}{3}\left((T-2B) + 2B\left(1 - \frac{T}{2B} + \frac{3T^2}{4B} + \cdots\right)\right)\right]$$

(4.184)
$$= \frac{e^{-\tilde{C}-\frac{B}{T}}T^{3/2}}{\sqrt{2\pi}}\left[\frac{1}{2B} - \frac{3T}{4B} - \frac{\tilde{D}T}{2} + \cdots\right] = \frac{T^{3/2}}{2\sqrt{2\pi}}\exp\left(-\frac{B}{T} - \tilde{C} - \ln B - \tilde{D}T - \frac{3T}{2B} + o(T)\right),$$

which completes the call price estimate for the stochastic volatility model.

Next, we want to derive a similar expression for the call prices of the standard Black-Scholes log-normal model, and equate the two prices in order to construct an exact formula for an implied volatility approximation.

In particular, we have previously seen that the transition density for BSM (appropriately changed to fix Paulot's convention's) is given by

$$p(K,t) = \frac{\sqrt{g(K)}}{\sqrt{2\pi T}}\sqrt{\Delta(F_0,K)}\mathcal{P}(F_0,K)e^{-d^2(F_0,K)/2T}(1 + a_1T + \cdots), \tag{4.185}$$

where here $a_1 = R/6 - Q = -\sigma^2/8$. Like in the previous case, this can be written as
(4.186)
$$\mathbb{E}[(\sigma_F^2)_{BSM}\delta(F-K)] = \sigma^2K^2p(K,T) = \frac{1}{2\pi T}\exp\left(-\frac{B_{BSM}}{T} - \tilde{C}_{BSM} - \tilde{D}_{BSM}T + o(T)\right),$$

where we have defined

$$B_{BSM} = \frac{1}{\sigma^2}\ln^2\frac{K}{F_0}, \quad \tilde{C}_{BSM} = -\ln\sigma - \frac{1}{2}\log(KF_0), \quad \tilde{D}_{BSM} = \frac{\sigma^2}{8}, \tag{4.187}$$

which can be verified simply by substitution and expanding exponential as in the SABR case. Now to find the Black implied volatility, we need to determine the unique constant σ such that when the BSM call is evaluated at σ, the result is the call price in our SABR approximation formula. Now note that the two call prices are equal iff

$$\frac{B}{T} + \tilde{C} + \ln B + \tilde{D}T + \frac{3}{2B}T = \frac{B_{BSM}}{T} + \tilde{C}_{BSM} + \ln(B_{BSM}) + \tilde{D}_{BSM} + \frac{3}{2B_{BSM}}T + o(T), \tag{4.188}$$

which should be viewed as an equation for σ that can now be solved order by order. To do this, we first assume that the implied volatility takes the form

$$\sigma(K,T) = \sigma_0(K) + \sigma_1(K)T + \sigma_2(K)T^2 + \cdots, \tag{4.189}$$

and note that if one uses the expansion formulas

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_0^2}\left[1 - 2\frac{\sigma_1}{\sigma_0}T + 3\frac{\sigma_1^2}{\sigma_0^2}T^2 - \frac{\sigma_2}{\sigma_0}T^3\right], \tag{4.190}$$

$$\ln\sigma \approx \ln\sigma_0 + \frac{\sigma_1}{\sigma_0}T, \quad \frac{\sigma^2}{8} \approx \frac{\sigma_0^2}{8} + \frac{\sigma_0\sigma_1}{4}T, \tag{4.191}$$

(4.192)
$$\ln\left(\frac{1}{2\sigma^2}\ln^2\frac{K}{F_0}\right) = \ln\left(\frac{1}{2\sigma_0^2}\ln^2\frac{K}{F_0}\right) - \frac{2\sigma_1}{\sigma_0}T, \quad \frac{3}{2}T\left(\frac{1}{2\sigma^2}\ln^2\frac{K}{F_0}\right)^{-1} = \frac{3\sigma_0^2T}{\ln^2(K/F_0)} + o(T^2),$$

the call equation can be written in terms of the coefficients of the implied volatility as

$$\frac{1}{2\sigma_0^2T}\ln^2\left(\frac{K}{F_0}\right)\left(1 - 2\frac{\sigma_1}{\sigma_0}T - 2\frac{\sigma_1}{\sigma_0}T^2 + 3\left(\frac{\sigma_1}{\sigma_0}\right)^2T^2\right) - \ln(\sigma_0) - \frac{\sigma_1}{\sigma_0}T \tag{4.193}$$

$$-\frac{1}{2}\ln(KF_0) + \ln\left(\frac{1}{2\sigma_0^2}\ln^2\frac{K}{F_0}\right) - 2\frac{\sigma_1}{\sigma_0}T + \frac{\sigma_0^2}{8}T + \frac{3\sigma_0^2}{\ln^2 K/F_0}T \tag{4.194}$$

$$= \frac{B}{T} + \tilde{C} + \ln B + \tilde{D}T + \frac{3}{2B}T + o(T). \tag{4.195}$$

Terms can then be equated order by order in T, which yields three equations for the coefficients of the implied volatility. The order -1 implied equation gives

$$\sigma_0 = \frac{1}{\sqrt{2B}}\ln\left(\frac{K}{F_0}\right) = \frac{\ln(K/F_0)}{d(F_0, V_0, K, V_{min})}. \tag{4.196}$$

The zeroth order equation gives

$$\frac{\sigma_1}{\sigma_0} = -\frac{\tilde{C} + \ln(\sigma_0\sqrt{KF_0})}{2B}, \tag{4.197}$$

and the first order term requires that

$$\frac{\sigma_2}{\sigma_0} = \frac{3}{2}B\left(\frac{\sigma_1}{\sigma_0}\right)^2 - \frac{1}{2B}\left[\tilde{D} + 3\frac{\sigma_1}{\sigma_0} - \frac{\sigma_0^2}{8}\right]. \tag{4.198}$$

We now make a note regarding the explicit computation of the different σ terms. We first note that σ_0 and σ_1 are significantly easier to compute than σ_2. This is because σ_2 involves the first heat kernel coefficient a_1, which is generally very difficult to compute in stochastic volatility models. Paulot provides a method for possibly constructing a_1 in the case of SABR.

4.6. **Lee's Moment Formula.** We now show that Paulot's first order approximation satisfies Lee's moment formula. In Lee [104], [105] the author showed under very general conditions that the Black implied volatility $I(x)$ for any stochastic volatility model must satisfy the following: There exists an $x^* > 0$ such that for all $x > x^*$,

$$I(x) < \sqrt{2|x|/T}, \tag{4.199}$$

where $x = \log(K/S_0)$ is the log moneyness. Any reasonable asymptotic approximation formula for a stochastic volatility model should satisfy this inequality. We now want to show that Paulot's first order approximation indeed satisfies the asymptotic condition.

Recall that Paulot's implied volatility takes the form $\sigma = \sigma_0 + \sigma_1 T$. We will show that σ and σ_1 individually satisfy this constraint, and thus any linear superposition of the does as well. First, note that σ_0 can be written as

$$\sigma_0 = \frac{\nu x}{\ln\left(\frac{\sqrt{\alpha^2+2\rho\alpha\nu q+\nu^2q^2}+\rho\alpha+q\nu}{(1+\rho)\alpha}\right)}, \tag{4.200}$$

where in the case that $\beta \neq 1$ we have $q = (K^{1-\beta} - F^{1-\beta})/(1-\beta)$. Thus for large strike values, we see that

$$\ln\left(\frac{\sqrt{\alpha^2+2\rho\alpha\nu q+\nu^2q^2}+\rho\alpha+q\nu}{(1+\rho)\alpha}\right) \approx \ln(c_1 q) \approx \ln\left(c_2K^{2-2\beta}\right) \approx c_3\log K/F = c_3 x, \tag{4.201}$$

for some constants c_i. Thus we see that $\sigma_0 \approx C$ for some constant C for large strike values. Similarly from the expression for σ_1/σ_0, one can note that $\tilde{C} \approx c_4 x$ and hence the numerator goes like x. The denominator goes like x as well. Hence the ratio σ_1/σ_0 goes like $1/x$, and thus the bound is satisfied.

4.7. **The Heston Model.** We now consider the Heston model. The form of the SDEs for this model (in particular the volvol function) prohibit us from constructing an exact implied volatility approximation. However, the below analysis is instructive in the sense that it shows the limitations of the heat kernel method for non hyperbolic stochastic volatility models.

The Heston model dynamics are given by

$$dF = \mu F dt + \sqrt{V} F dW_1, \tag{4.202}$$

$$dV = \kappa(\theta - V)dt + \xi\sqrt{V} dW_2, \tag{4.203}$$

where $\langle dW_1, dW_2\rangle = \rho dt$, (c.f. Gatheral [63, Chp.2]). Thus for this model, we identify the drift components as $\mu^F = \mu F$ and $\mu^V = \kappa(\theta - V)$. The covariance matrix is given by

$$\Sigma = \begin{pmatrix} \Sigma^{FF} & \Sigma^{FV} \\ \Sigma^{VF} & \Sigma^{VV} \end{pmatrix} = V\begin{pmatrix} F^2 & \rho F\xi \\ \rho F\xi & \xi^2 \end{pmatrix} \to \alpha \begin{pmatrix} f^2 & \rho f\xi \\ \rho f\xi & \xi^2 \end{pmatrix}, \tag{4.204}$$

in our usual initial value coordinates. We first note that if we define a new forward coordinate $q = \int_f^F du/u = \ln(F/f)$, then in this new coordinate, the volatility matrix becomes

$$\Sigma^{ij} = \alpha \begin{pmatrix} 1 & -\rho\xi \\ -\rho\xi & \xi^2 \end{pmatrix}. \tag{4.205}$$

Thus taking the same pair of (x, y) coordinates as were used in the SABR model. We see that

$$\Sigma_{(x,y)} = 2y \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \to g = \frac{1}{2}\Sigma^{-1} = \frac{1}{y}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.206}$$

Moreover the Ricci and scalar curvatures associated to this metric are given by

$$Ric = \frac{1}{2y^2}\delta, \quad R = -\frac{1}{y}, \tag{4.207}$$

so we see that this geometry is asymptotically flat. Also,

$$|Ric|^2 = 2e^{-4\phi}(\phi'')^2 = \frac{1}{2y^2}, \tag{4.208}$$

and we find that the distance integral can be integrated exactly, and is given by

$$d(p_1, p_2) = \int_{y_1}^{y_2} \frac{1/u du}{\sqrt{1/u - C^2}} = -\frac{2}{C}\left(\operatorname{arcsinh}\sqrt{C^2 y_1} - \operatorname{arcsinh}(\sqrt{C^2 y_2})\right). \tag{4.209}$$

To determine the constant $C(x_1, y_1, x_2, y_2)$, one needs to solve

$$x_2 - x_1 = \int_{y_1}^{y_2} \frac{du}{\sqrt{1/u - c^2}} \tag{4.210}$$

$$= \frac{1}{C^3}\left(C\sqrt{y_1 - C^2 y_1^2} - C\sqrt{y_2 - C^2 y_2^2} - \arcsin(C\sqrt{y_1}) + \arcsin(C\sqrt{y_2})\right). \tag{4.211}$$

We make the assumption that $C^2 y_i < 1$ to guarantee that the result is in fact real. This equations cannot be inverted exactly for C and one would needs to apply a rootsolver to numerically approximate $C(x_1, y_1, x_2, y_2)$. Since the calls in the Heston model can be priced with a formula that involves an inverse integral transform (see e.g. Gatheral [63], Carr and Madan [29]), this fact limits the use of heat kernel methods for the Heston model. In particular, there are many approximation methods for the standard call pricing formula for the Heston model which are numerically more efficient than any result that one can derive with heat kernel methods.

However, there is another class of stochastic volatility models where no such exact call pricing formula is known. For these models, the heat kernel method can be used in tandem with numerical methods to create a faster pricing formula than say direct Monte Carlo/finite difference techniques. We now examine this class of models.

4.8. **General Class of Stochastic Volatility Models.** We now characterize another class of models that correspond to hyperbolic geometry. We first note that if we reconsider our generic class of stochastic volatility models where now instead of taking $b(\sigma)$ to be generic, we fix $b(\sigma) = \sigma$. In particular, we consider stochastic volatility models of the form

$$dS = \sigma S\gamma(S)dW^1, \quad d\sigma = \mu(\sigma)dt + \nu\sigma dW^0, \tag{4.212}$$

that is, we have assumed that the volatility σ has lognormal volatility with arbitrary drift. Note that SABR is an example of such a model. Now using the same formalism that we previously considered in our stochastic volatility section, we recall that the inverse metric associated with this model is given by

$$g^{-1} = \frac{1}{2}\Sigma = \frac{\sigma^2}{2}\begin{pmatrix} S^2\gamma^2 & \rho\nu S\gamma \\ \rho\nu S\gamma & \nu^2\gamma^2 \end{pmatrix}. \tag{4.213}$$

We now can define new coordinates $x, \bar{y}$ given by

$$x = \frac{\sqrt{2}}{1-\rho^2}\left(\int_{S_0}^{S}\frac{du}{u\gamma(u)} - \rho\int_{\sigma_0}^{\sigma}\frac{udu}{b(u)}\right) = \frac{\sqrt{2}}{1-\rho^2}\left(\int_{S_0}^{S}\frac{du}{u\gamma(u)} - \frac{\rho}{\nu}(\sigma-\sigma_0)\right), \tag{4.214}$$

$$\bar{y} = \sqrt{2}\int_{\sigma_0}^{\sigma}\frac{udu}{b(u)} = \frac{\sqrt{2}}{\nu}(\sigma-\sigma_0), \tag{4.215}$$

from which we note that

$$\sigma = \frac{\nu\bar{y}}{\sqrt{2}} + \sigma_0. \tag{4.216}$$

Now in these coordinates the metric takes the form

$$g = \frac{1}{\sigma(\bar{y})^2}(dx^2 + d\bar{y}^2) = \frac{1}{\left(\frac{\nu\bar{y}}{\sqrt{2}} + \sigma_0\right)^2}(dx^2 + d\bar{y}^2), \tag{4.217}$$

where the constant in the denominator can be removed by defining $y = \bar{y} + \frac{\sigma_0\sqrt{2}}{\nu}$. In these final coordinates the metric is simply given by

$$g = \frac{1}{y^2}(dy^2 + dx^2) \tag{4.218}$$

and thus the distance function is just given by the hyperbolic distance that we previously mentioned. We finally note that although one can write an explicit form for the distance function for this class of models, it would be desirable also to be able to compute the integral involving γ in the x variable explicitly (which can be done for a vary wide class of functions γ). In addition, if one uses heat kernel methods to approximate the transition density, there will be restrictions on μ (although standard mean reversion does seem to present any issues) as μ in involved in the definition of A which in turn defines $\mathcal{P}$ which is needed in the heat kernel approximation formula. Now the distance function is again given by

$$d(p_1, p_2) = \operatorname{arccosh}\left(1 + \frac{(x_2-x_1)^2 + (y_2-y_1)^2}{2y_1y_2}\right), \tag{4.219}$$

where here $x_1 = x(S_0, \sigma_0)$, $y_1 = y(\sigma_0)$, $x_2 = x(S, \sigma)$, and $y_2 = y(\sigma)$.

Now that we have a distance function for the model, we can use the previously mentioned method to convert this into an implied volatility. To do this, we note that one can think of $d(p_1, p_2)$ as a function of the volatility σ. One then needs to minimize the distance function $d(\sigma)$ with respect to σ, and call the minimizer σ_{min}. Next, the distance function should be evaluated at σ_{min} in order to establish a zeroth order implied volatility approximation.

Constructing the first order correction is more complicated, since it involves computing a line integral along the geodesics of hyperbolic space which involves both functions γ and μ.

We finally note that one can extend the assumption that the volvol is of the form $\nu\sigma$ to being of the form $\nu\sigma^\gamma$ for some $\gamma > 0$. In this case, (as we previously saw with the analogous class of distance functions), one would only be able to write down an expression for the transition density up to solving an algebraic equation numerically which involves hypergeometric functions.

4.9. **Higher Dimensions and Limitations.** We now give a few comments on potential higher dimensional applications of these methods. As one increases the dimension of the problem, the generic geodesic equations become increasingly difficult to solve. Moreover, in dimension three and higher, not all metrics can be represented in some local coordinate system as conformally flat metrics. In particular, in dimension three, a metric is locally conformally flat iff the Cotton tensor

$$C_{ijk} = \nabla_i R_{jk} - \nabla_j R_{ik} - \frac{1}{4}(\nabla_i R g_{jk} - \nabla_j R g_{ik}), \tag{4.220}$$

is trivial. In dimension $n \geq 4$, the Weyl tensor

$$W_{ijkl} = R_{ijkl} + \frac{R}{(n-1)(n-2)}(g_{il}g_{jk} - g_{ik}g_{jl}) + \frac{1}{n-2}(R_{il}g_{jk} + g_{il}R_{jk} - R_{ik}g_{jl} - g_{ik}R_{jl}) \tag{4.221}$$

is the obstruction for a metric to admit a locally conformally flat coordinate system. If one wanted to attempt to extend heat kernel methods to higher dimensional models, it would be a good idea to start with models where $C_{ijk} = 0$ or $W_{ijkl} = 0$. However, even in this case, the geodesic equations are still quite complex and one may need to restrict to particular classes of conformal factors in order to be able to make progress towards constructing and explicit approximation formula.

It seems that the prospects of general extensions of heat kernel methods to construct implied volatility formulas for higher dimensional models seem quite grim. However, there is at least one class of higher dimensional financially relevant models where this becomes possible. In the case of SABR-LMM dynamics, the underlying geometry is hyperbolic geometry as pointed out it Labordere [100] and the analysis becomes somewhat tractable. One may be able to consider the generalized class of higher dimensional stochastic volatility Libor Market Models Brace et. al. [21] whose associated metrics are higher dimensional hyperbolic metrics. There is a known connection with these models and the two dimensional SABR model Memoli [117] which one could use to construct an approximation formula for the LMM. Alternatively, one could attempt to mirror the lower dimensional analysis in order to construct corresponding first order implied volatility approximation formulas.

4.10. **An Alternative Perturbation Theory for Stochastic Volatility.** We finally consider a more straightforward type of perturbation theory that one can use to construct an implied volatility which was pioneered in Berestycki [16] and applied to a family of stochastic volatility models in Medvedev [114]. We want to make a few connections with this work to geometry as well as some of the previous perturbation theory that we have considered.

We now consider an alternative (and technically simpler) form of perturbation theory that will give us another way to construct implied volatilities for stochastic volatility models. We follow the set up given in Medvedev [114].

We first restrict ourselves to two dimensions. Assume that the model takes the form

$$dS = \sigma S\phi(S)dW^1, \tag{4.222}$$

$$d\sigma = a(\sigma)dt + b(\sigma)dW^0, \tag{4.223}$$

where W^1 and W^0 are two Brownian motions with correlation ρ. In the variables (σ, S), a call price must satisfy the PDE

$$C_\tau = a\partial_\sigma C + \frac{b^2}{2}\partial_\sigma^2 C + \frac{\sigma^2\phi^2 S^2}{2}\partial_S^2 C + S\rho b\sigma\phi\partial_\sigma\partial_S C, \tag{4.224}$$

which has initial data $C(0, S, \sigma) = [S - K]^+$. We can change coordinates by defining a stochastic process $X = \log(S/K)$ and applying Ito's lemma to note that the above dynamics become

$$dX = -\frac{\sigma^2\phi^2}{2}dt + \sigma\phi dW^1, \tag{4.225}$$

$$d\sigma = adt + bdW^0. \tag{4.226}$$

In these coordinates, the associated PDE is given by

$$C_\tau = aC_\sigma + \rho\sigma b\phi C_{\sigma X} + \frac{b^2}{2}C_{\sigma\sigma} + \frac{\sigma^2\phi^2}{2}(C_{XX} - C_X), \tag{4.227}$$

where now the initial data takes the form

$$C(0, X, \sigma) = K[e^X - 1]^+. \tag{4.228}$$

Now if we let

$$C^{BSM}(X, \sigma, \tau, K) = K[e^X\Phi(d_1) - \Phi(d_2)], \tag{4.229}$$

where

$$d_{1,2} = \frac{X}{\sigma\sqrt{\tau}} \pm \frac{\sigma}{2}\sqrt{\tau}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}du, \tag{4.230}$$

be the price of a call under the lognormal BSM model, we can define the implied volatility function $I(X, \sigma, \tau, K)$ of the stochastic volatility model implicitly by

$$C = C^{BSM}(X, I(X, \sigma, \tau, K), \tau, K). \tag{4.231}$$

One can substitute this into the equation for C and after a bit of algebra find that it is equivalent to solving the implied volatility equation

$$[-2\tau I_\tau - I] + \sigma^2\phi^2\left[\tau I_{XX} + d_X(1 - dI_X)\right] + b^2[\tau I_{\sigma\sigma} - dd_\sigma I_\sigma]. \tag{4.232}$$

$$+2\rho b\sigma\phi\left[d_\sigma(1 - dI_X) + \tau I_{X\sigma} + \tau I_\sigma\right] + 2\tau I_\sigma a = 0, \tag{4.233}$$

where $d = X/I - I\tau/2$. Note this is a quasi-linear second order equation. One can then represent I as a formal power series in τ, by

$$I(X, \sigma, \tau, K) = I_0(X, \sigma, K) + I_1(X, \sigma, K)\tau + I_2(X, \sigma, K)\tau^2 + \cdots, \tag{4.234}$$

and then substitute this form into our PDE for I and extract the zeroth order term by letting $\tau \to 0$ to find that

$$\sigma^2\phi^2 d_X^2 + 2\rho b\sigma\phi d_\sigma d_X + b^2 d_\sigma^2 = 1. \tag{4.235}$$

where we note that $d = X/I + O(\tau)$. This is a Hamilton-Jacobi equation which arises in geometry. To understand the geometric context of this equation, we define an inverse metric

$$g^{-1} = \begin{pmatrix} \sigma^2\phi^2 & \rho b\sigma\phi \\ \rho b\sigma\phi & b^2 \end{pmatrix}, \tag{4.236}$$

which has an associated metric given by

$$g = \frac{1}{1-\rho^2}\begin{pmatrix} \sigma^{-2}\phi^{-2} & -\rho b^{-1}\sigma^{-1}\phi^{-1} \\ -\rho b^{-1}\sigma^{-1}\phi^{-1} & b^{-2} \end{pmatrix}, \tag{4.237}$$

where here $\phi = \phi(Ke^S)$. Then with respect to this metric, equation (4.235) can be written as

$$|\nabla d|_g^2 = g^{ij}\nabla_i d\nabla_j d = 1, \tag{4.238}$$

which is a Hamilton Jacobi-equation on the manifold $(\mathbb{R}^2, g)$, where ∇ is the Levi-Civita connection associated with g. The existence theory for this equation is developed in Arnol'd [8]. In geometry, any function d that satisfies equation (4.238) is defined to be a distance function Petersen [135]. Thus computing the zeroth order correction in this form of perturbation theory is equivalent to finding the distance function for the underlying Riemannian manifold. Note that in the case of the SABR model, we have $\phi(S) = S^{\beta-1} = (Ke^X)^{\beta-1}$ and $b(\sigma) = \sigma$.

We now note that achieving a tractable first order correction using the heat kernel method without any simplifying assumptions on the functional form of the implied volatility is quite difficult (although theoretically possible). To see this, we first label the five terms that we have delimited in equation (4.232) Q_i where i indexes the different terms. Thus for instance, $Q_1 = -2\tau I_\tau - I$ and $Q_5 = 2\tau I_\sigma a$. We then multiply each of these terms by I^3, and note that they can be written out explicitly in terms of derivatives of I as

$$Q_1 = -2\tau I^3 I\tau - I^4, \tag{4.239}$$

$$Q_2 = -\frac{\sigma^2\phi^2}{4}\left(-4I^2 + 8XII_X - 4X^2I_X^2 + \tau^2I^4I_X^2 - 4\tau I^3 I_{XX}\right), \tag{4.240}$$

$$Q_3 = \frac{b^2}{4}\left((4X^2 - \tau^2I^4)I_\sigma^2 + 4\tau I^3 I_{\sigma\sigma}\right), \tag{4.241}$$

$$Q_4 = 2b\rho\sigma\phi\left(I_\sigma\left[\frac{\tau}{2}I^3 - XI + X^2I_X - \frac{\tau^2}{4}I^4I_X - \frac{\tau^2}{4}I^4I_X\right] + \tau I_{X\sigma}\right), \tag{4.242}$$

$$Q_5 = 2\tau I^3 I_\sigma a, \tag{4.243}$$

so that equation (4.232) is just given by $\Sigma_{i=1}^5 Q_i = 0$. Now define Q_i' by the following operation: Expand Q_i about $\tau = 0$, differentiate the result, and then let $\tau \to 0$, i.e. take the linear coefficient in τ of Q_i and call the result Q_i'. Specifically, the Q_i' are given by

$$Q_1' = -I_0(2 + I_1), \tag{4.244}$$

$$Q_2' = \sigma^2\phi^2\left(2X^2I_1(I_0)_X^2 + 2I_0^2\left(I_1 - X(I_1)_X\right) + 2XI_0(I_0)_X(X(I_1)_X - 2I_1) + I_0^3(I_XX)^2\right), \tag{4.245}$$

$$Q_3' = b^2\left(2X^2I_1(I_0)_\sigma^2 + 2X^2I_0I_\sigma I1_\sigma + I_0^3(I_0)_{\sigma\sigma}\right), \tag{4.246}$$

$$Q_4' = b\rho\sigma\phi\Big(-2XI_0^2(I_1)_\sigma + 4X^2I_1(I_0)_\sigma(I_0)_X + 2XI_0(-2I_1(I_0)_\sigma + X(I_1)_\sigma(I_0)_X \tag{4.247}$$

$$+X(I_0)_\sigma(I_1)_X) + I_0^3\left((I_0)_\sigma + 2(I_0)_{X\sigma}\right)\Big), \tag{4.248}$$

$$Q_5' = 2aI_0^3(I_0)_\sigma. \tag{4.249}$$

In particular, assuming that we have used the zeroth order equation to solve for I_0, the equation $\sum_{i=1}^5 Q_i = 0$ can be expressed as

$$f_1(X,\sigma)\partial_\tau I_1(X,\sigma) + f_2(X,\sigma)\partial_X I_2(X,\sigma) + f_3(X,\sigma)\partial_\sigma I_3(X,\sigma) = 0, \tag{4.250}$$

where we note that the f_i are explicit functions. One, in principle, can solve this equation John [89, Chp. 2]; however, in practice this is typically quite involved.

If one is willing to introduce additional error into the problem by further expanding the assumed form of the solution about the at the money strike, then he can use an estimate in Proposition 2 of Medvedev [114] to simplify the process of computing an approximate first order correction. This approximation is given by the following: Given the zeroth order term of the implied volatility expansion, a first order approximation is determined by

$$I(X,\sigma,\tau,K) \approx I^{(0)}(X,\sigma,K) + I^{(1)}(0,\sigma,K)\tau, \tag{4.251}$$

with

$$I^{(1)}(0,\sigma,K) = [K\phi\phi' + \phi^2]\frac{\rho\sigma b}{4} + \phi\left[\frac{\rho^2 b^2}{24\sigma} + \frac{b^2}{12\sigma} - \frac{\rho^2 bb'}{6}\right] \tag{4.252}$$

$$+\frac{\sigma^3}{12}\left[K\phi^2\phi' - \frac{K^2\phi(\phi')^2}{2} + K^2\phi^2\phi''\right] + \frac{\phi a}{2}, \tag{4.253}$$

where here ϕ,ϕ', and ϕ'' are all evaluated at K.

We next note that as we have already previously seen, there is a wide class of models that admit an exact distance function. In a similar spirit, in Medvedev [114], the author notes that we can find I_0 in a large class of models. In particular, his first proposition states that if we take the volatility function of the volatility process to be of the form $b(\sigma) = \beta\sigma^\nu$, with other model parameters arbitrary, then if $\nu \neq 1$, we find that

$$I_0(X,\sigma,K) = \sigma(\delta-1)\xi\frac{X}{x}\frac{\sqrt{1+f^2(\zeta)}}{\delta\zeta - f(\zeta)}, \tag{4.254}$$

where we set

$$\xi = \frac{x\beta}{\sigma^\delta}, \quad \delta = 2-\nu, \quad x = \int_0^X \frac{ds}{\phi(Ke^s)}, \quad \zeta_0 = -\frac{\rho}{\delta\sqrt{1-\rho^2}}, \quad \zeta = \frac{\xi}{\sqrt{1-\rho^2}} + \zeta_0, \tag{4.255}$$

and f is given implicitly by solving

$$\frac{\zeta}{(1+f^2)^{\delta/2}} = -\frac{1}{\delta}\rho(1-\rho^2)^{(\delta-1)/2}\int_{-\rho/\sqrt{1-\rho^2}}^{f} \frac{ds}{(1+s^2)^{\delta/2+1}}. \tag{4.256}$$

In the case that $\nu = 1$, we have a simpler formula,

$$I^{(0)}(X,\sigma,K) = \sigma\xi\frac{X}{x}\log\left(\frac{1-\rho}{\zeta - \rho + \sqrt{1-2\rho\xi+\xi^2}}\right). \tag{4.257}$$

The comment we wish to add regarding this work is that this integral can be evaluated. In particular, the anti-derivative is given by

$$\int \frac{ds}{(1+s^2)^{\delta/2+1}} = s\,{}_2F_2\left(\frac{1}{2}, 1+\frac{\delta}{2}, \frac{3}{2}, -s^2\right) \tag{4.258}$$

where here ${}_2F_1$ is the hypergeometric function of the second kind which we previously saw in some of our distance function computations. There are many values of δ where this formula can be expressed as a composition of elementary functions, Daalhuis [41].

5. Symmetries of Finance Equations

We now compute the Lie symmetry groups of several PDE that arise in finance. We first outline our method for computing the symmetries of a given equation and then apply them to both the heat equation and Black-Scholes-Merton equation. We then consider the classical Asian equation whose solution prices an Asian option on an asset which evolves according to lognormal dynamics. This is a second order parabolic equation in two spatial variables. We will use Lie-Point symmetry analysis to reduce the equation to solving an associated equation with only one spatial variable. We then consider several similar Asian type equations which are related to the classical equation by a change of numeraire argument. Thus we build upon the current literature of symmetry methods in finance which includes the following works: Bordag and Mikaelyan [26], Cicogna [34], Craddock and Lennox [37], Craddock and Platen [40], Gazizov and Ibragimov [66], Goard [72], Goard [73], Douden and O' Hara [75], Ivanova et. al. [87], Laurence [103], Naicker et. al. [127], and Sinkala et. al. [147].

5.1. **Introduction to Symmetry Analysis.** We first give an overview of our method of computing the set of Lie point symmetries which form a (semi)group which corresponds to a differential operator. One can use these symmetries to reduce the PDE corresponding to this operator to a simpler equation which involves few independent variables that the initial equation. We will then demonstrate this technique in the case of the heat equation, and then turn to more complicated PDE that arise in finance.

A Lie point transformation of a differential equation is a change of independent and/or dependent variables that maps a given solution of the equation into another solution. Correspondingly, this coordinate transformation leaves the equation form invariant in the new coordinates. Sometimes such a mapping is trivial, i.e. a rescaling of the dependent variable or constant shift of a dependent variable, and other times it allows us to construct the fundamental solution of the PDE, as is the case with the heat equation. A simple example of such a transformation is the change of variables $x \to \lambda x$, $t \to \lambda^2 t$, and $u \to u$ applied to the arguments of a function $u(x,t)$ that solves the one dimensional heat equation $u_t = u_{xx}$. In particular, if $u(x,t)$ solves the heat equation, then so does $u(\lambda x, \lambda^2 t)$ for any $\lambda > 0$. However, if we apply this change of coordinates to the heat equation itself, the chain rule requires that $\partial_t u \to \lambda^{-2}\partial_t u$ and $\partial_x u \to \lambda^{-1}\partial_x u$. Thus $\partial_x^2 u \to \lambda^{-2}\partial_x^2 u$ and in the transformed variables, the heat equation becomes $u_t(\lambda x, \lambda^2 t) = u_{xx}(\lambda x, \lambda^2 t)$ for any $\lambda \neq 0$. However, when expressed in the new coordinates $\tilde{t} = \lambda^2 t$, $\tilde{x} = \lambda x$, and $\tilde{u} = u$, this equation can be written as

$$\lambda^{-2}\tilde{u}_t(\lambda x, \lambda^2 t) = \lambda^{-2}\tilde{u}_{xx}(\lambda x, \lambda^2 t) \to \partial_{\tilde{t}}\tilde{u}(\tilde{x}, \tilde{t}) = \partial_{\tilde{x}}^2\tilde{u}(\tilde{x}, \tilde{t}), \tag{5.1}$$

which shows that the equation is form invariant under this coordinate transformation, i.e., that u in these new coordinates indeed satisfies the heat equation as well.

One of the central usages of point transformations lies in their ability to reduce a given differential equation to a simpler equation with fewer independent variables. In many cases, one can use the simpler equations to construct the fundamental solution of the original partial differential operator. We will show how one can systematically perform such a reduction later, but first recall the prototypical example of the heat equation where such symmetry allows one to reduce a PDE to a simpler solvable ODE. In the case of the heat equation, we previously noted that the scaling transformation $u(x,t) \to u(\lambda x, \lambda^2 t)$ is a symmetry of the equation. If we construct a new minimal independent variable $z = x^2/t$ by seeking the "simplest" variable one can construct from $\tilde{x} = \lambda x$ and $\tilde{t} = \lambda^2 t$ which does not have explicit λ dependence, then we see that when the heat equation is expressed with respect to this variable, it reduces to an ordinary differential equation. In particular, for any suitably differentiable function $f(x,t)$, using the chain rule we can compute

$$\partial_x f = \frac{\partial z}{\partial x}\partial_z f = \frac{2x}{t} f_z, \quad \partial_x^2 f = \frac{4x^2}{t^2} f_{zz} + \frac{2}{t} f_z, \quad \partial_t f = -\frac{x^2}{t^2} f_z. \tag{5.2}$$

Thus if we take $f(x,t) = u(z)$, the heat equation for f requires that

$$\frac{4x^2}{t} u_{zz} + \left(2 + \frac{x^2}{t}\right) u_z = 4zu''(z) + (2+z)u'(z) = 0. \tag{5.3}$$

Note that this is a simple first order ordinary differential equation in u' that can be immediately solved to produce an exact solution of the heat equation,

$$u' = c_1 \frac{e^{-z/4}}{\sqrt{z}}, \quad u(z) = c_1 \int_0^z e^{-\zeta/4} \zeta^{-1/2} d\zeta + c_2, \tag{5.4}$$

where c_1, c_2 are constants of integration. Differentiating the solution u with respect to x and choosing c_1 appropriately yields the fundamental solution of the heat equation, i.e. $u_x = \exp(-x^2/4t)/\sqrt{4\pi t}$. One can then use this fundamental solution to construct the Green's function for the heat operator for suitably prescribed boundary data, which in turn allows one to construct an exact solution to the Cauchy problem for the heat equation (c.f. Evans [54, ch. 2]).

It is possible to observe that the above scaling symmetry of the independent variables of the heat equation is in fact a bona fide symmetry simply by inspection of equation. In addition, note that temporal and spatial translational invariance are also two relatively simple symmetries of the heat equation. A natural question to consider is the following one: do there exist more complicated transformations that cannot reasonably be found by inspection alone?

The answer to this question turns out to in the affirmative, and there is a pseudo-algorithmic method for constructing symmetry transformations for any differential equation which is typically referred to as Lie symmetry analysis. The proper setting for the standard theory of symmetry requires the language of differentiable manifolds, prolongation, and Lie group theory (c.f. Cantwell [28], Lie [109], Olver [131], Olver [132], Stephani [150]). We will refrain from the use of this language and utilize an equivalent method to construct the point symmetries of several PDE that arise in finance.

We first demonstrate this method using the heat equation as an example and note that it can be straightforwardly extended to more complicated partial differential equations.

Let $u(x,t) : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ be a solution of the one dimensional heat equation $u_t = u_{xx}$, and extend u to a function $u(s,x,t) : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}$ and consider the coupled system of the heat equation and the symmetry equation

$$\partial_s u = -T(t,x,u(x,t))\partial_t u - X(t,x,u(x,t))\partial_x u + U(t,x,u(t,x)). \tag{5.5}$$

We call the functions T, X, and U the coefficients of the symmetry equation (or symmetry coefficients for short) and they will be determined by a system of PDE that is specific to the heat equation. More precisely, in the standard theory, these functions correspond to the coefficients of the infinitesimal generator of the local Lie group of symmetries of the heat equation. Our first aim is to determine the functional form of these coefficients by imposing the compatibility condition $[\partial_s, \partial_t]u = 0$ where $[\partial_s, \partial_t] = \partial_s\partial_t - \partial_t\partial_s$ is the commutator (or Lie bracket) of the two operators. Explicitly, this condition requires

$$0 = [\partial_s, \partial_t]u = -2u_{xxx}(u_x T_u + T_x) + u_{xx}(T_t - u_x^2 T_{uu} - 2u_x(X_u + T_{xu}) - 2X_x - T_{xx}) \tag{5.6}$$

$$+u_x(-u_x^2 X_{uu} + u_x(U_{uu} - 2X_{xu}) + 2U_{xu} + X_t - X_{xx}) + U_{xx} - U_t. \tag{5.7}$$

Equation (5.7) is satisfied if

$$0 = T_u = T_x = X_u = U_{uu} = T_t - 2X_x = 2U_{xu} + X_t = U_{xx} - U_t. \tag{5.8}$$

This system of PDEs for T, X, and U was constructed by viewing u_x, u_{xx}, u_{xxx} and all possible products of these functions as independent variables and requiring the associated coefficients to vanish separately, i.e. we interpret the above as a linear superposition of these functions and demand the coefficients of this expression vanish. One could assume that the symmetry coefficients depend on higher derivatives of u and then attempt to find more general solutions of (5.7). We call this process a generalized symmetry analysis, and it is equivalent to the techniques developed in Olver [131, ch. 5]; we do not pursue this further.

We now wish to demonstrate another technique for deriving and solving equations (5.8) which we will utilize for more complicated equations below. First, let $Z = [\partial_s, \partial_t]$, and compute

$$\partial_{u_{xxx}} \partial_{u_x} Z = 2T_v = 0, \tag{5.9}$$

which can be solved to find $T = T_1(\tau, x)$. We then substitute this into Z to simplify equation (5.7). Using the resulting expression, we compute

$$\partial_{u_{xxx}} Z = 2\partial_x T_1, \tag{5.10}$$

which again can be easily solved for $T_1(x, \tau) = T_2(\tau)$. We update Z again with the new form of T, and calculate

$$\partial^2_{u_{xx}} Z = 2X_v, \tag{5.11}$$

which in turn requires that $X = X_0(\tau, x)$. Performing the same procedure again, we find that

$$\partial_{u_{xx}} Z = -\partial_\tau T_2 + 2\partial_x X_0, \tag{5.12}$$

whose solution is given by

$$X_0 = X_1(\tau) + \frac{1}{2} x \partial_\tau T_2. \tag{5.13}$$

We then can calculate

$$\partial^2_{u_x} Z = U_{uu} = 0, \tag{5.14}$$

which requires that U is an affine function

$$U = U_0(\tau, x) + uU_1(\tau, x). \tag{5.15}$$

Using our updated form for the symmetry coefficients, we next compute

$$\partial_{u_x} Z = -\frac{1}{2}\left(2\partial_\tau X_1 + x\partial^2_\tau T_2 + 4\partial_x U_1\right), \tag{5.16}$$

which can be solved to determine

$$U_1 = U_2(\tau) - \frac{x}{2}\partial_\tau X_1 - \frac{x^2}{8}\partial^2_\tau T_2. \tag{5.17}$$

In summary, we have now exhausted taking partial derivatives of Z with respect to different derivatives of u_x and found that the symmetry coefficients must take the restricted forms

$$T = T_2(\tau), \quad U = U_0(\tau, x) + u\left(U_2(\tau) - \frac{x}{2}\partial_\tau X_1 - \frac{x^2}{8}\partial^2_\tau T_2\right), \quad X = X_1(\tau) + \frac{x}{2}\partial_\tau T_2. \tag{5.18}$$

We still require that $Z = 0$, and this condition now reduces to

$$u\partial_\tau U_2 + \frac{u}{4}\partial^2_\tau T_2 - \frac{1}{2}ux\partial^2_\tau X_1 - \frac{1}{8}ux^2\partial^3_\tau T_2 - \partial^2_x U_0 + \partial_\tau U_0 = 0. \tag{5.19}$$

This is an affine function of u that must identically vanish. Thus both the slope and intercept must vanish, which implies

$$\frac{1}{4}\partial_\tau^2 T_2 + \partial_\tau U_2 - \frac{x}{8}\left(4\partial_\tau^2 X_1 + x\partial_\tau^3 T_2\right), \quad \partial_\tau U_0 = \partial_x^2 U_0. \tag{5.20}$$

Thus U_0 is just an arbitrary solution of the heat equation, and we note that our other equation is quadratic in x. Since it must be the zero function as well, each of the coefficients of the x^i must independently vanish. In particular $\partial_\tau^3 T_2 = 0$ so

$$T_2 = \alpha + \beta\tau + \gamma\tau^2, \tag{5.21}$$

where we use Greek letters to denote constants. We also must have

$$\partial_\tau^2 X_1 = 0 \to X_1 = \epsilon + \xi\tau, \tag{5.22}$$

as well as the vanishing of the constant term

$$0 = \frac{u\gamma}{2} + u\partial_\tau U_2 \to U_2 = -\frac{\gamma}{2}\tau + \zeta. \tag{5.23}$$

Substituting these formulas into equation (5.18) and making the regularizing constant redefinitions $\beta \to 2\beta$, $\gamma \to 4\gamma$, $\epsilon \to 2\delta$, $\xi \to \epsilon$, $\zeta \to \eta$, we find that the symmetry coefficients of the heat equation must take the form

$$\begin{pmatrix} T \\ X \\ U \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 2t \\ x \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 4t^2 \\ 4tx \\ -(2t + x^2)u \end{pmatrix} \tag{5.24}$$

$$+\delta \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \epsilon \begin{pmatrix} 0 \\ 2t \\ -xu \end{pmatrix} + \eta \begin{pmatrix} 0 \\ 0 \\ u \end{pmatrix} + U_0 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{5.25}$$

Thus we have found a six dimensional space of symmetries of the heat equation parametrized by $\alpha, \beta, \gamma, \delta, \epsilon$, and η along with an infinite dimensional function space of solutions, parametrized by U_0 of the heat equation. We note that the presence of the U_0 symmetry is due to the fact that the heat equation is a linear PDE. In fact, whenever one computes the symmetry group of a linear equation, he will always have an infinite dimensional symmetry that corresponds to adding an arbitrary solution of the equation in question to a known solution and hence a U_0 type vector will always be present in the symmetry group calculation.

Now that we have constructed the point symmetries of the heat equation, we can use them to construct exact solutions of the equation by choosing values for the Greek constants to restrict to a single symmetry vector and then solve the corresponding symmetry equation. For example, if we focus on the interesting γ-symmetry by fixing $\gamma = 1$ and set all other constants and u_0 to zero, then the resulting symmetry equation (5.127) becomes

$$\partial_s u = -(4tx\partial_x + 4t^2\partial_t + 2t + x^2)u. \tag{5.26}$$

which can be integrated using standard techniques (e.g. the method of characteristics) John [89] yielding a solution of the form

$$u(s,t,x) = \frac{1}{(1+4st)^{1/2}} \exp\left(-\frac{sx^2}{1+4st}\right) u\left(0, \frac{t}{1+4st}, \frac{x}{1+4st}\right). \tag{5.27}$$

In particular, if $u(x,t)$ is any solution of the heat equation, then so is $u(s,t,x)$ for any $s \in \mathbb{R}$. We have thus produced a one parameter family of solutions $u(s,t,x)$ that solve the heat equation given that $u(0,t,x)$ solves the equation. We can think of $u(s,t,x)$ as a solution corresponding to the one parameter family of coordinate transformations

$$t \to \frac{t}{1+4st}, \quad x \to \frac{x}{1+4st}, \quad u \to \frac{1}{\sqrt{1+4st}} \exp\left(-\frac{sx^2}{1+4st}\right) u. \tag{5.28}$$

Clearly, such a symmetry would be difficult to deduce by inspection of the heat equation alone unlike the initial constant scaling symmetry we considered. Note that if we can take $u(0,t,x) \equiv 1$ as our known (trivial) solution of the heat equation and substitute it into equation (5.27), then combining this with the α symmetry that corresponds to time translational invariance allows us to construct the fundamental solution

$$u(s,t,x) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{x^2}{4t}\right), \tag{5.29}$$

of the heat equation.

In addition to generating exact solutions from the point symmetries of the heat equation, we may also use s-independent symmetries to construct associated reduced equations. This is a method that we will employ in the case of Asian type equations to reduce their dimensionality. For example, the s-independent symmetry related to γ is given by setting $\partial_s u = 0$ in the associated symmetry equation which reduces it to

$$0 = (4tx\partial_x + 4t^2\partial_t + 2t + x^2)u, \tag{5.30}$$

with solution

$$u(x,t) = \frac{\exp(-x^2/4t)}{\sqrt{x}} f\left(\frac{t}{x}\right), \tag{5.31}$$

where f is any suitably differentiable function that according to the heat equation satisfies

$$\frac{3x^2}{t^2} f(t/x) + \frac{12x}{t} f'(t/x) + 4f''(t/x) = 0. \tag{5.32}$$

In particular, if we define $z = t/x$, then this may be written as a simple ODE

$$4z^2 f'' + 12zf' + 3f = 0. \tag{5.33}$$

Thus we have reduced the heat equation to solving a simpler first order ODE which can be integrated immediately to find $f = c_1 z^{-1/2} + c_2 z^{-3/2}$. We obtain the fundamental solution of the heat equation by taking $c_2 = 0$.

Note that solving the symmetry equation for each symmetry vector is slightly different from the standard technique. One usually exponentiates the symmetry vectors associated to each Greek constant of integration independently as we will partially demonstrate below. The reason this process is called exponentiation is that it is equivalent to applying the exponential map on a local Lie Group whose tangent space at one point is spanned by the symmetry vectors. Typically, one performs the exponentiation process as follows. Consider the symmetry vector

$$4\tau^2 \frac{\partial}{\partial \tau} + 4\tau x \frac{\partial}{\partial x} - (x^2 + 2t)u\partial_u. \tag{5.34}$$

We can exponentiate this equation by extending the variables x, τ, and u to functions of s and solving the coupled system of ODE

$$\tau_s = 4\tau^2, \quad x_s = 4\tau x, \quad u_s = -(x^2 + 2\tau)u, \tag{5.35}$$

subject to the initial conditions $\tau(0) = \tau$, $x(0) = x$, $u(0) = u$. Doing this yields

$$\tau(s) = \frac{\tau}{1-4s\tau}, \quad x(s) = \frac{x}{1-4s\tau}, \quad u(s) = \exp\left(-\frac{sx^2}{1-4s\tau}\right)\sqrt{1-4s\tau}u, \tag{5.36}$$

from which we constructed a set of coordinate transformations

$$(x,t,u) \to \left(\frac{x}{1-4s\tau}, \frac{\tau}{1-4s\tau}, \sqrt{1-4s\tau}\exp\left(-\frac{sx^2}{1-4s\tau}\right)u\right) \equiv (\tilde{x},\tilde{t},\tilde{u}). \tag{5.37}$$

Inverting this for $\tilde{x},\tilde{\tau}$, and $\tilde{u}$, we find that

$$x = \frac{\tilde{x}}{1+4s\tilde{\tau}}, \quad \tau = \frac{\tilde{\tau}}{1+4s\tilde{\tau}}. \tag{5.38}$$

If $u = f(x,\tau)$ is a solution of the heat equation, then its transformation is given by

$$\tilde{u} = \sqrt{1-4s\tau}\exp\left(-\frac{sx^2}{1-4s\tau}\right)u(x,\tau) = \frac{1}{\sqrt{1+4s\tilde{\tau}}}\exp\left(-\frac{s\tilde{x}^2}{1+4s\tilde{\tau}}\right)f\left(\frac{\tilde{x}}{1+4s\tilde{\tau}}, \frac{\tilde{\tau}}{1+4s\tilde{\tau}}\right), \tag{5.39}$$

so that if $f(x,\tau)$ is any solution of the heat equation, then so is

$$\frac{1}{\sqrt{1+4s\tau}}\exp\left(-\frac{sx^2}{1+4s\tau}\right)f\left(\frac{x}{1+4s\tau}, \frac{\tau}{1+4s\tau}\right). \tag{5.40}$$

Note as $s \to 0$, this solution reduces to $f(x,\tau)$.

5.2. **Correspondence with the Standard Theory.** Before proceeding to our next example, we show how portions of the method we have outlined to construct the fundamental solution of the heat equation correspond to the general theory as outlined in Olver [131]. The reader uninterested in this correspondence can skip this portion of the text.

Performing a symmetry analysis on a system of nonlinear equations is widely regarded as the best way to attempt to construct exact solutions of the system. Although the majority of the symmetry analysis literature is concerned with non-linear PDE, these methods sometimes still prove useful in constructing interesting exact solutions of marginally complicated linear PDE (especially low dimensional linear parabolic PDE that often arise in finance). The goal of the symmetry analysis method is to construct coordinate transformations of independent variables x^i and dependent variables ϕ^j represented by $(x^i,\phi^j) \to (\tilde{x}^i(x^k,\phi^{k'}), \tilde{\phi}^j(x^l,\phi^{l'}))$ that map solutions of a given equation into a new solution. We summarize this method using the notation in Olver [131]. Other less technical references for the symmetry analysis technique are given in Cantwell [28], and Stephani [150].

Consider a system of partial differential equations denoted by

$$\Delta_\nu(x,u^{(n)}) = 0, \quad \nu = 1,\dots,l, \tag{5.41}$$

where $x = (x^1,\dots,x^p)$ is the set of independent variables and $u = (u^1,\dots,u^q)$ is the set of dependent variables; here $1,\dots,q$ are indices for the set of all partial derivatives of u up to order n. For $u = f(x)$, with $f : \mathbb{R}^p \to \mathbb{R}^q$ and component functions f^i, for $i = 1,\dots,q$, we define the n-th prolongation of f to be a mapping

$$\mathrm{pr}^{(n)} f : \mathbb{R}^p \to U^{(n)}, \tag{5.42}$$

given by $\mathrm{pr}^{(n)} f = u^{(n)} = \{u_J = \partial_J f\}$ where J is a multi-index which accounts for all possible derivatives $\partial_J f$ up to order n. We have used $U^{(n)}$ to indicate the domain of the u_J. In other words, the prolongation map lists all derivatives of u up to a given order. For example, if we consider $u = f(x,y)$, we have

$$\mathrm{pr}^{(2)} f(x,y) = (u; u_x, u_y; u_{xx}, u_{xy}, u_{yy}). \tag{5.43}$$

The space $\mathbb{R}^p \times U^{(n)}$ is called the n-th order jet space of $\mathbb{R}^p \times U$, where U denotes the domain of u. The fundamental idea behind the method of symmetry analysis is to view Δ_ν

as a map from the n-th order jet space into $\mathbb{R}^l$, and then to study the subvariety (algebraic submanifold)

$$\mathcal{L}_\Delta = \{(x, u^{(n)}) \in \mathbb{R}^p \times U^{(n)} | \Delta(x, u^{(n)}) = 0\}. \tag{5.44}$$

Now let $M \subset \mathbb{R}^p \times U^{(n)}$ be open. A symmetry group of Δ_ν is a local group (meaning that the group operation is not necessarily always defined) of transformations G acting on an open subset M such that when $(x, u^{(n)}) = (x, \mathrm{pr}^{(n)} f)$ is in the $\mathcal{L}_\Delta$ corresponding to (5.44), then so is $(\tilde{x}, \tilde{u}^{(n)}) = g \cdot (x, \mathrm{pr}^{(n)} f)$ for all $g \in G$ for which $g \cdot (x, \mathrm{pr}^{(n)} f)$ is defined. The difficulty here is deciding how the group element g acts on each component of $\mathrm{pr}^{(n)} f$ given only its action on $\mathrm{pr}^{(0)} f = f$. Specifically, one would like this action to be consistent with the chain rule. In turn, this is facilitated by finding a local, rather than global, notion of g's action on the zeroth order jet space, i.e. on the space of only the dependent and independent variables. This is called the projection of g onto this space. This operation is defined as follows.

Let X be a vector field on only the space of dependent and independent variables $\mathbb{R}^p \times U$. Hence $X : \mathbb{R}^p \times U \to \mathbb{R}^p \times U$. The "infinitesimal" action $(x, u) \mapsto X(x, u)$ induces a local but finite action via the exponential map, as in $(x, u) \mapsto \exp(sX)(x, u)$, the latter defined for s small enough via the Taylor series of the exponential. Here we think of X as an infinitesimal generator of a symmetry $g \in G$, the latter restricted to the action on the zeroth order jet space. We extend X from a vector field on $\mathbb{R}^p \times U$ to one on the associated n-th order jet space $\mathbb{R}^p \times U^{(n)}$, the latter vector field denoted by $\mathrm{pr}^{(n)} X$, as defined in the following:

Theorem 5.1. *(Olver* [131]*) If*

$$X = \sum_i \xi^i(x, u) \frac{\partial}{\partial x^i} + \sum_a \phi_a \frac{\partial}{\partial u^a}, \tag{5.45}$$

then X has prolongation

$$\mathrm{pr}^{(n)} X = X + \sum_{a,J} \phi_a^J(x, u^{(n)}) \frac{\partial}{\partial u_J^a}, \tag{5.46}$$

where

$$\phi_a^J(x, u^{(n)}) = D_J \left(\phi_a - \xi^a u_i^a\right) + \sum_i \xi^i \partial_i u_J^a, \tag{5.47}$$

and the subscripts on u denote partial derivatives and $D_J = D_{j_1} D_{j_2} \cdots D_{j_k}$ is the J-th total derivative (with respect to the independent variables x). In (5.46) the sum is over the multi-index J which indexes all n-th order derivatives.

The vector X we have constructed here is equivalent to the exponential operator in the general symmetry equation we considered in our example of the heat equation, which we recall is

$$L = -T(t, x, u(x, t))\partial_t - X(t, x, u(x, t))\partial_x + U(t, x, u(t, x)). \tag{5.48}$$

Now define the Jacobi matrix of Δ_ν to be

$$J_{\Delta_\nu}(x, u^{(n)}) = \left(\frac{\partial \Delta_\nu}{\partial x^i}, \frac{\partial \Delta_\nu}{\partial u_J^a} \right), \tag{5.49}$$

and say Δ_ν is of maximal rank iff the rank of J_{Δ_ν} is l. The following is the fundamental theorem of the Lie method.

Theorem 5.2. *(Olver* [131]*) Let Δ_ν as in (5.41) be a system of differential equations of maximal rank. If G is a local group of transformations acting on M and*

$$\mathrm{pr}^{(n)} X[\Delta_\nu(x, u^{(n)})] = 0, \quad \text{for} \quad \nu = 1, \ldots, l, \tag{5.50}$$

whenever $\Delta_\nu = 0$*, for every infinitesimal generator* X *of* G*, then* G *is a symmetry group of* Δ_ν*.*

This theorem gives sufficient conditions for a vector field X to be an infinitesimal generator of the symmetry group G for the system Δ_ν. The "infinitesimal criterion" (5.50) is also necessary, so that all such generators are described by (5.50), which, in addition to the equations having maximal rank, they are also "locally solvable" (Olver [131, Chp. 3]). This occurs when, for example, the system is analytic and can be written in Cauchy-Kovalevskaya form, an instance of this form being the evolutionary form of (5.126). Equation (5.50) is equivalent to our commutator equation $[\partial_s, \partial_t]u = 0$ in the heat equation example.

These two theorems enable one to calculate all infinitesimal generators of the symmetry group G of a maximal rank, locally solvable system of equations, such as (5.126). We may then exponentiate the infinitesimal generators to obtain the symmetry group of the system, i.e. an explicit set of coordinate transformations that allow us to map known solutions of the differential equation to new solutions. Finally, these symmetries can be applied to known and usually simple solutions of the system in order to obtain new and hopefully more interesting solutions. For a concrete example of this method, we refer the reader to a calculation of the point symmetry group of the heat equation, as in (Olver [131]), which yields the same results as the heat equation example.

5.3. **Symmetry Analysis of the Black-Scholes-Merton Equation.** There have been multiple papers in the literature that have studied the symmetry group of the Black-Scholes-Merton (BSM) equation (Liu and Wang [111]), (Silderberg [145]), (Singh [146])). We apply our symmetry group construction technique to this equation in order to build intuition for extending it to later more complicated Asian type equations.

The BSM PDE is given by

$$u_\tau + \frac{1}{2}\sigma^2 x^2 u_{xx} + r x u_x - r u = 0, \tag{5.51}$$

where u is the price of a European call on an asset x at time τ, where the risk free interest rate is r. We first note that there is a well known series of coordinate transformations that reduce the BSM equation to the heat equation. Specifically, if we let $x = e^z$, then this equation can be written as

$$u_\tau + \left(r - \frac{1}{2}\sigma^2\right) u_z + \frac{1}{2}\sigma^2 u_{zz} = ru. \tag{5.52}$$

If we then define coordinates $t = T - \tau$, $u = e^{-r\tau}v$, $y = z + (r - \sigma^2/2)t$, then the previous becomes

$$v_t = \frac{\sigma^2}{2} v_{yy}, \tag{5.53}$$

which is just a forward heat equation. Since the BSM and heat equations are linked together by coordinate transformations, we expect that their symmetry groups should be similar. However, note that there is no reason to expect a one to one correspondence of the groups since the spatial translational symmetry $x \to x + \epsilon$ of the heat equation is not present in the BSM equation.

We compute the symmetry group of equation (5.51) using the same methods that we have used for the heat equation. We note that it is natural to choose the appropriate asset numeraire in order to force the drift term of the BSM SDE to vanish, which is tantamount to setting $r = 0$ in the PDE. Thus it really suffices to consider the advection-free equation $u_t + (\sigma^2 x^2/2)u_{xx} = 0$. However, we do not set r to zero, in order to make the problem slightly more interesting.

The BSM symmetry equation is exactly the same as the one for the heat equation since both PDEs have the same dimension. Specifically, it is given by

$$\partial_s u = -T(\tau, x, u(x,\tau))\partial_\tau u - X(\tau, x, u(x,\tau))\partial_x u + U(\tau, x, u(\tau, x)). \tag{5.54}$$

We impose the compatibility condition $[\partial_\tau, \partial_s]u = 0$, and in analogy to the heat equation pick off independent terms to construct a system of PDE for the symmetry coefficients. In particular, the fully expanded commutator is given by

$$Z \equiv [\partial_s, \partial_\tau]u = -rU + ru_x X + u_{xx}x\sigma^2 X - r^2u^2T_u + r^2uu_xxT_u + \frac{1}{2}ruu_{xx}x^2\sigma^2T_u + ru_xu_{xx}x^3\sigma^2T_u \tag{5.55}$$

$$+u_xu_{xx}x^3\sigma^4T_u + \frac{1}{2}u_xu_{xxx}x^4\sigma^4T_u + ruU_u - ruu_xX_u - u_xu_{xx}x^2\sigma^2X_u - \frac{1}{2}ruu_x^2x^2\sigma^2T_{uu} + \frac{r}{2}u_x^3x^3\sigma^2T_{uu} \tag{5.56}$$

$$+\frac{1}{4}u_x^2u_{xx}x^4\sigma^4T_{uu} + \frac{1}{2}u_x^2x^2\sigma^2U_{uu} - \frac{1}{2}u_x^3x^2\sigma^2X_{uu} - r^2uxT_x + r^2u_xx^2T_x + \frac{3}{2}ru_{xx}x^3\sigma^2T_x \tag{5.57}$$

$$+u_{xx}x^3\sigma^4T_x + \frac{1}{2}u_{xxx}x^4\sigma^4T_x + rxU_x - ru_xxX_x \tag{5.58}$$

$$-u_{xx}x^2\sigma^2X_x - ruu_xx^2\sigma^2T_{xu} + ru_x^2x^3\sigma^2T_{xu} + \frac{1}{2}u_xu_{xx}x^4\sigma^4T_{xu} \tag{5.59}$$

$$+u_xx^2\sigma^2U_{xu} - u_x^2x^2\sigma^2X_{xu} - \frac{1}{2}rux^2\sigma^2T_{xx} + \frac{1}{2}ru_xx^3\sigma^2T_{xx} + \frac{1}{4}u_{xx}x^4\sigma^4T_{xx} + \frac{1}{2}x^2\sigma^2U_{xx} \tag{5.60}$$

$$-\frac{1}{2}u_xx^2\sigma^2X_{xx} - ruT_\tau + ru_xxT_\tau + \frac{1}{2}u_{xx}x^2\sigma^2T_\tau + U_\tau - u_xX_\tau, \tag{5.61}$$

which can be factored to one's liking. In order to extract PDE for T, U, and X from Z, we combine terms with like powers of derivatives of u and set them equal in the same manner that we performed the similar computation for the heat equation. We do not simultaneously write down the symmetry coefficient PDEs prior to solving (as was the case with the heat equation), but rather solve them one at a time and use the solutions to continually update the forms of T, U, and X in order to hopefully simplify the equations yet to be considered. We first note that

$$\partial_{u_x}\partial_{u_{xxx}}Z = \frac{1}{2}x^4\sigma^4T_u, \tag{5.62}$$

from with we see $T_u = 0$, so we must have $T = T_1(\tau, x)$. We subsequently find that $\partial_{u_{xxx}}Z = 0$, requiring $T_x = 0$, which further simplifies the form of T to $T = T_2(\tau)$. Now if we substitute T_2 into equation (5.61) and differentiate, we find that

$$\partial_{u_x}\partial_{u_{xx}}Z = -x^2\sigma^2X_u, \tag{5.63}$$

which forces $X_u = 0$ so that $X = X_0(\tau, x)$. Updating our symmetry coefficients again, we now have $T = T_2(\tau)$ and $X = X_0(\tau, x)$ and further calculate

$$\partial_{u_{xx}}Z = x\sigma^2X_0 + \frac{1}{2}x^2\sigma^2\partial_\tau T_2 - x^2\sigma^2\partial_xX_0, \tag{5.64}$$

which can be solved for

$$X_0 = xX_1(\tau) + \frac{x}{2}(\ln x)\partial_\tau T_2. \tag{5.65}$$

Updating the form of the symmetry coefficients and recomputing Z, we see that $\partial^2_{u_x} Z = U_{uu}$, and thus U is affine in u,

$$U(\tau, x, v) = U_0(\tau, x) + uU_1(\tau, x). \tag{5.66}$$

Using this, we next compute

$$\partial_{u_x} Z = \frac{x}{4}\left((2r-\sigma^2)\partial_\tau T_2 - 4\partial_\tau X_1 - 2\ln(x)\partial^2_\tau T_2 + 4x\sigma^2\partial_x U_1\right), \tag{5.67}$$

which can be solved for U_1,

$$U_1 = U_2(\tau) + \frac{1}{4}(\ln x)\partial_\tau T_2 - \frac{r(\ln x)\partial_\tau T_2}{2\sigma^2} + \frac{(\ln x)\partial_\tau X_1}{\sigma^2} + \frac{(\ln x)^2\partial^2_\tau T_2}{4\sigma^2}. \tag{5.68}$$

Using these updated forms of the symmetry coefficients, one can check that $\partial_{u_{xxx}} Z = \partial_{u_{xx}} Z = \partial_{u_x} Z = 0$.

We have now exhausted all derivatives of Z of the form $\partial_{\partial^i_x u} Z$ and have found that the symmetry coefficients must take the form

$$T = T_2(\tau), \quad X = xX_1(\tau) + \frac{x}{2}(\ln x)\partial_\tau T_2(\tau), \tag{5.69}$$

$$U = U_0(\tau, x) + u\left(U_2(\tau) + \frac{1}{4}(\ln x)\partial_\tau T_2 - \frac{r(\ln x)\partial_\tau T_2}{2\sigma^2} + \frac{(\ln x)\partial_\tau X_1}{\sigma^2} + \frac{(\ln x)^2\partial^2_\tau T_2}{4\sigma^2}\right). \tag{5.70}$$

We still need $Z = 0$, and substituting this form of the symmetry coefficients into our expression for Z implies that

$$cZ = 8r\sigma^2 U_0 + u(2r+\sigma^2)^2\partial_\tau T_2 - 2(4u\sigma^2\partial_\tau U_2 + (4ru\partial_\tau X_1 - 2u\sigma^2)\partial_\tau X_1 + u\sigma^2\partial^2_\tau T_2 \tag{5.71}$$

$$+4u(\ln x)\partial^2_\tau X_1 + u(\ln x)^2\partial^3_\tau T_2 + 4rx\sigma^2\partial_x U_0 + 2x^2\sigma^4\partial^2_x U_0 + 4\sigma^2\partial_\tau U_0), \tag{5.72}$$

where we set $c = -8\sigma^2$. Since Z must vanish, so must

$$\partial_v\partial_x Z = -\frac{8}{x}\partial^2_\tau X_1 - \frac{4\ln x}{x}\partial^3_\tau T_2 = 0. \tag{5.73}$$

Now each of these terms must vanish independently, so we need $\partial^2_\tau X_1 = 0$ and $\partial^3_\tau T_2 = 0$ which in turn requires that

$$T_2 = \alpha + \beta\tau + \gamma\tau^2, \quad X_1 = \rho + \xi\tau. \tag{5.74}$$

Using these forms for T_2 and X_1, we can reevaluate Z to find that

$$8r\xi + 4\gamma\sigma^2 - 4\xi\sigma^2 + (2r+\sigma^2)^2(\beta + 2\gamma\tau) + 8\sigma^2\partial_\tau U_2 = 0. \tag{5.75}$$

We can solve this equation to find

$$U_2 = \frac{1}{8\sigma^2}\left(8\zeta\sigma^2 + 4r(r\beta - 2\xi)\tau + \tau(4(r\beta - \gamma + \xi)\sigma^2 + \beta\sigma^4 + \gamma(2r+\sigma^2)^2\tau)\right), \tag{5.76}$$

which concludes the computation. Thus we have found that

$$T = \alpha + \beta\tau + \gamma\tau^2, \quad X = x(\rho + \xi\tau) + \frac{x}{2}(\beta + 2\gamma\tau)\ln x, \tag{5.77}$$

$$U = u\Bigg(\frac{1}{8\sigma^2}\left(8\zeta\sigma^2 + 4r(r\beta - 2\xi)\tau + \tau(4(r\beta - \gamma + \xi)\sigma^2 + \beta\sigma^4 + \gamma(2r+\sigma^2)^2\tau)\right) \tag{5.78}$$

$$\frac{\xi\ln x}{\sigma^2} + \frac{1}{4}(\beta + 2\gamma\tau)\ln x - \frac{r(\beta + 2\gamma\tau)\ln x}{2\sigma^2} + \frac{\gamma\ln x^2}{2\sigma^2}\Bigg) + U_0, \tag{5.79}$$

where here U_0 is an arbitrary solution of the BSM equation. We can extract different symmetry vectors from these equations by setting the Greek constants individually to one while demanding all others are zero. First, we fix $\alpha = 1$ and let the other constants vanish, we find that

$$v_1 = \frac{\partial}{\partial \tau}, \tag{5.80}$$

is a symmetry of the equation. Next, we let $\beta = 1$ with all the other parameters set to zero and see

$$v_2 = \tau\frac{\partial}{\partial \tau} + \frac{x}{2}\ln x\frac{\partial}{\partial x} + \left(\left(\frac{1}{4} - \frac{r}{2\sigma^2}\right)\ln x + \frac{(2r+\sigma^2)^2\tau}{8\sigma^2}\right)u\frac{\partial}{\partial u}. \tag{5.81}$$

Similarly, fixing $\gamma = 1$, one finds that

$$v_3 = \tau^2\frac{\partial}{\partial \tau} + x\tau\ln x\frac{\partial}{\partial x} + \left(\left(\frac{1}{2} - \frac{r}{\sigma^2}\right)\tau\ln x + \frac{\ln^2 x}{2\sigma^2} - \frac{4\sigma^2\tau - (2r+\sigma^2)^2\tau^2}{8\sigma^2}\right)u\frac{\partial}{\partial u}. \tag{5.82}$$

Next, if we let $\rho = 1$ and allow all other constants to vanish, we find that

$$v_4 = x\frac{\partial}{\partial x}. \tag{5.83}$$

Now if $\xi = 1$ and the other Greeks are taken to be trivial, we see that

$$v_5 = x\tau\frac{\partial}{\partial x} + \left(\frac{\ln x}{\sigma^2} - \frac{(8r - 4\sigma^2)\tau}{8\sigma^2}\right)u\frac{\partial}{\partial u}. \tag{5.84}$$

is also a symmetry vector of the BSM equation. Finally, setting $\zeta = 1$, and demanding the other constraints vanish shows the last symmetry vector is

$$v_6 = u\frac{\partial}{\partial u}. \tag{5.85}$$

Thus we see the symmetry group for the BSM equation has the same dimensionality as that of the heat equation.

We now turn to solving the symmetry equations associated with these vectors and finding their associated point symmetries as well as computing their point transformations using classical methods.

In the case of v_1, the symmetry equation and its solution are given by

$$u_s = -u_\tau \to u = f(\tau - s, x), \tag{5.86}$$

for any function f. Similarly, if we solve the ODE system associated with v_1 which is given by,

$$\tau_s = 1, \quad x_s = 0, \quad u_s = 0 \tag{5.87}$$

subject to the initial conditions $u(0) = u$, $x(0) = x$, and $u(0) = u$, we see that the solution is $\tau(s) = s + \tau$, $x(s) = x$, and $u(s) = u$, so the coordinate transformations

$$(x, \tau, u) \to (x, \tau + s, u) \equiv (\tilde{x}, \tilde{\tau}, \tilde{u}) \tag{5.88}$$

are contained in the symmetry group of the BSM equation. Thus if $u = f(x, \tau)$ is a solution of the BSM equation, then so is

$$\tilde{u} = u(x, \tau) = f(\tilde{x}, \tilde{\tau} - s). \tag{5.89}$$

We now proceed to consider the other symmetries of the BSM equation in a similar manner. We next consider the v_2 symmetry equation

$$u_s = -\tau u_\tau - \frac{x}{2}\ln x u_x + \left(\left(\frac{1}{4} - \frac{r}{2\sigma^2}\right)\ln x + \frac{(2r+\sigma^2)^2\tau}{8\sigma^2}\right)u, \tag{5.90}$$

which has a solution given by

$$u = \exp\left(\frac{(2r+\sigma^2)^2\tau}{8\sigma^2}\right) x^{\frac{1}{2}-\frac{2r}{\sigma^2}} f\left(e^{-s}\tau, \ln(\ln(x)) - \frac{s}{2}\right), \tag{5.91}$$

where here f is an arbitrary function. In particular, one can substitute u into the BSM equation in order to reduce it to a simpler equation. Specifically, if we define

$$\bar{\tau} = e^{-s}, \quad \bar{x} = \ln(\ln(x)) - \frac{s}{2}, \tag{5.92}$$

then substitution of u into the BSM equation requires

$$0 = 2\ln^2 x\bar{\tau} f_{\bar{\tau}} - \sigma^2 f_{\bar{x}} - f_{\bar{x}\bar{x}}, \tag{5.93}$$

where here $\ln x = \exp(\bar{x} + s/2)$. If one chooses f to satisfy this equation (e.g. take f to be a constant), then the resulting u solves the BSM equation. This is not particularly interesting in this example, but we will see later in the case of the Asian equation, an analogous computation proves quite useful. We now proceed with the classical way of constructing symmetries from this equation. In order to exponentiate v_2, we need to solve the system of ODE

$$\tau_s = \tau, \quad x_s = \frac{x}{2}\ln x, \quad u_s = \left(\left(\frac{1}{4} - \frac{r}{2\sigma^2}\right)\ln x + \frac{(2r+\sigma^2)^2\tau^2}{8\sigma^2}\right) u, \tag{5.94}$$

subject to the typical initial conditions. The solution of these equations is given by

$$\tau(s) = e^s\tau, \quad x(s) = x^{e^{s/2}}, \quad u(s) = \exp\left(\frac{(e^s-1)(2r+\sigma^2)^2\tau}{8\sigma^2}\right) x^{\frac{(\sigma^2-2r)(e^{s/2}-1)}{2\sigma^2}} u. \tag{5.95}$$

We can simplify the form of this expression by letting $s \to \ln s^2$, so that it can be written as

$$\tau(s) = s^2\tau, \quad x(s) = x^s, \quad u(s) = \exp\left(\frac{(s^2-1)(2r+\sigma^2)^2\tau}{8\sigma^2}\right) x^{\frac{(\sigma^2-2r)(s-1)}{2\sigma^2}} u. \tag{5.96}$$

Thus if we again let $\tilde{\tau} = \tau(s)$, $\tilde{x} = x(s)$, and $\tilde{u} = u(s)$, and let $u = f(x,\tau)$ be any solution of the BSM equation, then

$$\tilde{u} = \exp\left(\frac{(s^2-1)(2r+\sigma^2)^2\tau}{8\sigma^2}\right) x^{\frac{(\sigma^2-2r)(s-1)}{2\sigma^2}} f(x,\tau) \tag{5.97}$$

$$= \exp\left(\frac{(s^2-1)(2r+\sigma^2)^2\tilde{\tau}}{8s^2\sigma^2}\right) \tilde{x}^{\frac{(\sigma^2-2r)(s-1)}{2s\sigma^2}} f(\tilde{x}^{1/s}, s^{-2}\tilde{\tau}), \tag{5.98}$$

is also a solution for any $s > 0$. In particular, if we let $X = x^{1/s}$, $T = \tau/s^2$, and then substitute this last equation into the BSM equation where τ is replaced by $\tilde{\tau}$ (and perform the parallel substitution for x), then $f(X,T)$ must satisfy

$$f_T + \frac{1}{2}\sigma^2 X^2 f_{XX} + rXf_X - rf, \tag{5.99}$$

i.e. f satisfies the BSM equation in these new variables.

We now proceed to consider the slightly more complicated v_3 vector. The symmetry equation for this vector is given by

$$u_s = -\tau^2 u_\tau - x\tau\ln x u_x + \left(\left(\frac{1}{2} - \frac{r}{\sigma^2}\right)\tau\ln x + \frac{\ln^2 x}{2\sigma^2} - \frac{4\sigma^2\tau - (2r+\sigma^2)^2\tau^2}{8\sigma^2}\right) u, \tag{5.100}$$

which can be solved for

$$u(s,\tau,x) = \frac{1}{\sqrt{\tau}} \exp\left(\frac{(2r+\sigma^2)^2\tau^2 + 4\ln^2 x}{8\sigma^2\tau}\right) x^{\frac{1}{2}-\frac{r}{\sigma^2}} f\left(-\frac{1+s\tau}{\tau}, \ln(-\tau^{-1}\ln x)\right), \tag{5.101}$$

for any function f which, if we again define the arguments of f to be T and X, must satisfy

$$\sigma^2\tau^2(f_X - f_{XX}) - 2(\ln^2 x)f_T = 0. \tag{5.102}$$

where one can invert $T = T(\tau)$ and $X = X(x)$ to write this as a PDE only in terms of X and T. Now to exponentiate this vector field, we need to solve

$$\tau_s = \tau^2, \quad x_s = x\tau\ln x, \quad u_s = \left(\left(\frac{1}{2} - \frac{r}{\sigma^2}\right)\tau\ln x + \frac{\ln^2 x}{2\sigma^2} - \frac{4\sigma^2\tau - (2r+\sigma^2)^2\tau^2}{8\sigma^2}\right), \tag{5.103}$$

subject to our usual initial data. The solution to this system of coupled ODEs (after mapping $s \to \sigma^2 s$ for convenience) is given by

$$\tau(s) = \frac{\tau}{1 - \sigma^2 s\tau}, \quad x(s) = x^{\frac{1}{1-\sigma^2 s\tau}}, \tag{5.104}$$

$$u(s) = \sqrt{1 - 2s\sigma^2\tau}\exp\left(\frac{s\left(2r\sigma^2\tau^2 + \left(\ln x - \left(r - \sigma^2/2\right)\tau\right)^2\right)}{1 - 2s\sigma^2\tau}\right)u. \tag{5.105}$$

Defining tilde variables and repeating the above argument, we can conclude that if $f(x,\tau)$ is a solution of the BSM equation, then so is

$$u = \frac{1}{\sqrt{1+\sigma^2 s\tau}}\exp\left(\frac{s}{2}\left(\frac{(\ln x - (r-\sigma^2/2)\tau)^2 + 2r\sigma^2\tau^2}{1+s\sigma^2\tau}\right)\right) f\left(\frac{\tau}{1+\sigma^2 s\tau}, x^{\frac{1}{1+s\sigma^2\tau}}\right), \tag{5.106}$$

which again can be directly verified by substituting u into the BSM equation, and noting that it requires $f(X,T)$ to satisfy the BSM equation in the variables $T = \tau/(1+s\sigma^2\tau)$ and $X = x^{1/(1+s\sigma^2\tau)}$.

This is in some sense the most interesting symmetry of the BSM equation. To understand why, we note that the Green's function for the BSM operator (where we take free space boundary conditions) is Melnikov and Melnikov [115]

$$G(x,t;x_0) = \frac{\exp(-r(T-t))}{x_0\sqrt{2\pi\sigma^2(T-t)}}\exp\left(-\frac{[\ln(x/x_0) + (r-\sigma^2/2)(T-t)]^2}{2\sigma^2(T-t)}\right). \tag{5.107}$$

which bears a close similarity to the v_3 symmetry. In fact, if we take $f(T,X) = e^{rT}$ to be the simplest nontrivial solution of the BSM equation (just as we took a constant here in the case of the heat equation), and choose s appropriately as well as use our time translational symmetry, we can conclude from the symmetries that if f is a solution of the BSM equation, then so is G, i.e. we can produce the fundamental solution of the BSM operator from the symmetry analysis and knowledge of the fact that e^{rT} solves the BSM equation. This is the best possible result that one can expect to achieve from the symmetry analysis methods, i.e. the singly most important piece of information one can potentially obtain from a Lie Symmetry analysis is the fundamental solution of the associated PDE.

We now look at the final three symmetries. The v_4 and v_6 symmetries are simple and performing the analogues of the above computations one finds that they correspond to constant rescaling invariance of the x and u variables in the heat equation. In particular, if $u(x,\tau)$ solves the heat equation, then so do $u(e^s x,\tau)$ and $e^s u(x,\tau)$, both of which can be seen by inspection.

Thus we finally consider the v_5 symmetry, whose corresponding symmetry equation is

$$u_s = -x\tau + \left(\frac{\ln x}{\sigma^2} - \frac{(8r - 4\sigma^2)\tau}{8\sigma^2}\right)u, \tag{5.108}$$

which has a solution given by

$$u = \exp\left(-\frac{s(\sigma^2 - 2r - s)\tau + 2\ln x}{2\sigma^2}\right) f(e^{-s\tau}x, \tau). \tag{5.109}$$

If we let $X = e^{-s\tau}x$, then substituting the above into the BSM equation, requires that $f(X,\tau)$ satisfy

$$\frac{X^2\sigma^2}{2} f_{XX} + rXf_X + f_X - rf = 0 \tag{5.110}$$

which can be solved for

$$f = c_1\frac{1+rx}{r} + c_2\frac{1+rx}{r}\int^x \exp\left(-\frac{2(r\ln u + \sigma^2\ln(1+ru) - u^{-1})}{\sigma^2}\right) du, \tag{5.111}$$

where the c_i are constants of integration. When this is combined with the previous formula for u it produces a new exact solution to the BSM equation (but it does not satisfy the standard European call option initial data choice for BSM). We finally note that the ODE system for this equation is

$$\tau_s = 0, \quad x_s = x\tau, \quad u_s = \left(\frac{lnx}{\sigma^2} - \frac{(8r - 4\sigma^2)\tau}{8\sigma^2}\right) u, \tag{5.112}$$

which has solution

$$\tau(s) = \tau, \quad x(s) = e^{s\tau}x, \quad u(s) = \exp\left(s\frac{(\sigma^2 - 2r - s)\tau + 2\ln(e^{s\tau}x)}{2\sigma^2}\right) u. \tag{5.113}$$

Thus if $f(x,\tau)$ is a solution of the BSM equation, then so is

$$\exp\left(-\left(\frac{s^2\sigma^2}{2} + s(r - \sigma^2/2)\right)\tau\right) f(xe^{-\sigma^2 s\tau}, \tau), \tag{5.114}$$

which concludes the symmetry analysis for BSM equation. We now turn to the case of Asian options under lognormal dynamics and see that we can also find interesting information from the symmetry group of the corresponding PDE.

5.4. **Symmetries of the Classical Asian Equation.** We now turn attention to analyzing the more complicated two spatial dimensional Asian equation. Typically when one considers PDE that are complex in the sense that their coefficient functions depend on the dependent variables in a nontrivial way, the symmetry group of the equation is no longer interesting. This is intuitive, since one would except equations with simple, e.g. constant, coefficients to have more symmetry in their dependent variables than corresponding equations that have more complicated coefficient function dependencies. The Asian equation turns out to be an exception to this statement which we will show below.

An Asian option is an option whose payoff depends on the time average of an underlying asset's or rate's price instead of the underlying itself. This property makes the option's value more stable in the sense that it is not as prone to devaluation near the maturity time if the underlying experiences large price variations. We will consider the simplest type of European Asian option, namely one with a payoff based on the arithmetic average value whose underlying asset evolves according to lognormal dynamics. Currently there is no known explicit formula defined in terms of elementary functions for the pricing of such an Asian option before expiration for all values of its underlying asset.

Since the payoff depends on the path of the underlying and not the path of the average, the PDE associated to this pricing problem has two spatial variables Ingersoll [85]. There is a small literature on numerical methods for this pricing problem, c.f. Francesco and Pascucci [62], Geman [67], Monti [124], Linetsky [110], Polidoro and Mogavero [136], Thompson [156].

On the other hand, in Vecer [161] and Vecer [162], the author used a change of numeraire amongst other financially insightful arguments to reduce the pricing problem of the Asian option to valuing a certain simpler non time-averaged claim. This approach included the idea of requiring that deterministic, stock-denominated portfolios replicate the Asian forward. The resulting parabolic, single spatial variable PDE can be robustly simulated with simple numerical schemes, i.e. there exist methods that are not prone to numerical instability or other issues that commonly arise in simulation. In Rogers and Shi [140], the authors found a similar reduction to a one dimensional equation. They used an invariance of the Asian option equation under a homogeneous change of scale in the original two dimensional state space. (The resulting PDE may be difficult to numerically simulate, without advecting away certain deterministic processes). We demonstrate how one may arrive at any of these various reductions, including all other possible continuous "point" reductions in a systematic fashion, i.e. by symmetry analysis methods. In addition to finding the symmetry used explicitly in Rogers and Shi [140], and demonstrated implicitly in Vecer [161], we find additional symmetries of the Asian option equation. One such symmetry leads to a new, nontrivial reduction of the equation.

We first review the market model for the type of Asian option that we wish to consider as well as the derivation of the Asian equation. Let $S(t)$ denote the asset process of a stock which evolves by geometric Brownian motion, i.e. let $S(t)$ satisfy

$$dS(t) = rS(t)dt + \sigma S(t)dW. \tag{5.115}$$

Here W, $0 \leq t \leq T$ is a 1-dimensional standard Brownian motion under the risk-neutral measure $\mathbb{P}$, so that r denotes the deterministic rate of return of a risk-free bond in this simple market. Let $K \geq 0$ be the strike of a European Asian call option which is defined by its payoff at time T is given by

$$V(T) = \left(\frac{1}{T}\int_0^T S(t)dt - K\right)^+, \tag{5.116}$$

and let the no-arbitrage and perfectly hedgable fair price $V(t)$ of the option be given by

$$V(t) = u(x(t), y(t), T - t), \tag{5.117}$$

where $x(t) = S(t)$ is the value of the asset underlying the derived security at time t, where $y(t)/t$ is the arithmetic mean of the asset value over $[0, t] \subset [0, T]$, i.e.

$$\frac{y(t)}{t} = \frac{1}{t}\int_0^t x(u)du = \frac{1}{t}\int_0^t S(u)du, \tag{5.118}$$

and where $u = u(x, y, \tau)$ is the solution of the following initial value problem

$$\left(-\partial_\tau - r + rx\partial_x + x\partial_y + \frac{1}{2}\sigma^2 x^2 \partial_x^2\right) u(x, y, \tau) = 0, \tag{5.119}$$

$$u(x, y, 0) = \max\left\{\frac{y}{T} - K, 0\right\}. \tag{5.120}$$

This equation is derived in the same manner as the BSM equation. In particular, one demands the discounted option value $e^{-rt}v(t, S(t), Y(t))$ be drift free, and then applies Ito's lemma to this function and sets the corresponding drift term to be zero. Again, we could use the money market as our numeraire, which has the effect of setting $r = 0$; however, we again keep r general by taking the local currency to be our numeraire to make the problem slightly more interesting.

Here we have set $\tau = \tau(t) = T - t$ to be the time until the option can be exercised. We will refer to equation (5.119) as the Asian equation.

The initial value problem (5.119), (5.120), should not only include the initial data (5.120) (or said terminal data in terms of t) but also the subset of $\mathbb{R}^3$ on which the PDE is valid In our setting this domain is $x > 0$, $y \in \mathbb{R}$, $\tau \in (0, T]$. The asymptotic information $u(x, y, \tau)$ as $x \to 0$ and $y \to -\infty$ is also relevant. See Shreve [144, p.322-323]. The latter boundary conditions encode information that may be deduced directly from the representation of the price as a risk neutral expectation of the payoff (5.116), they are determined by the other information.

For a modern review of Asian options, c.f. Rogers and Shi [140], Shreve [144, p. 320-330]. For a review giving insight into why Asian and other options not written solely on market tradables are, in a certain sense, difficult to price, see Vecer [161, ch. 1,3].

In Vecer [162], the author reduced the Asian initial value problem (5.119), (5.120) to the two variable initial value problem

$$\left(-\partial_\tau + \frac{1}{2}\sigma^2(z - \Delta_X(\tau))^2\partial_z^2\right) g(z, \tau) = 0, \quad z \in \mathbb{R}, \quad \tau \in [0, T], \tag{5.121}$$

$$g(z, 0) = z^+, \quad z \in \mathbb{R}, \quad \text{where} \quad \Delta_X(\tau) = \frac{1 - e^{-r\tau}}{rT}. \tag{5.122}$$

The value process of the reduced equation (5.122) is that of a self-financing portfolio of the underlying and the money market which replicates the associated Asian forward. This can be understood to some extent by noting that, as shown in Vecer [162], the associated fair price of the Asian call is then given in terms of a function $g = g(z, \tau)$ which satisfies (5.122), and takes the form

$$V(t) = v(x(t), y(t), T - t) = x(t) g\left(\frac{X(x(t), y(t), T - t)}{x(t)}, T - t\right), \tag{5.123}$$

where $x(t)$ and $y(t)$ are related to the underlying's evolution as in (5.117), and where

$$X(x, y, \tau) = \Delta_X(\tau)x + e^{-r\tau}(y/T - K) \tag{5.124}$$

In Vecer [162] it is shown that $X(x, y, T - t)$ is the value of a portfolio holding $\Delta_X(\tau)$ shares of the stock at any given time $t \in [0, T]$, capital in the amount of $\Delta_X(T - 0)x(0) + e^{-r(T-0)}(y(0)/T - K) = \Delta_X(T)S(0) - e^{-rT}K$ having been borrowed at time $t = 0$ from the money market to finance the initial stock investment. Since

$$X(x(T), y(T), T - T) = \Delta_X(T - T)x(T) + e^{-r(T-T)}\left(\frac{1}{T}y(T) - K\right) = \frac{1}{T}\int_0^T S(u)du - K, \tag{5.125}$$

the value of the associated forward has been replicated at time T. More importantly, the choice (5.124) for the replicating portfolio (essentially $\Delta_X(\tau)$ and the initial capital) is entirely independent of the market dynamics (5.115).

The reduction in Rogers and Shi [140] is related to (5.122) in a simple way. We now show that this reduction arises naturally as a Lie group invariant solution of the Asian equation. Moreover, we will find an additional class of group invariant solutions which leads to a novel reduction of equation (5.119).

Just as in the case of BSM, we now consider the following coupling of the Asian equation (5.119) to its associated symmetry equation

$$\begin{aligned} \partial_\tau u &= \left(-r + rx\partial_x + x\partial_y + \frac{1}{2}\sigma^2 x^2\partial_x^2\right) u, \\ \partial_s u &= -\left(T(\tau, x, y, u)\partial_\tau + X(\tau, x, y, u)\partial_x + Y(\tau, x, y, u)\partial_y\right) uU(\tau, x, y, u). \end{aligned} \tag{5.126}$$

where we note that the symmetry equation now involves a function Y due to the fact that the PDE has two spatial variables instead of one. Requiring the compatibility condition $Z = [\partial_s, \partial_\tau]v = 0$, we find simple linear PDE's describing T, X, Y, and U's dependence on τ,x,y and u. We do not write out the expression for Z since it is quite involved, but now indicate how we deduce the form of the symmetry coefficients from Z.

We first compute

$$\partial_{u_{xxx}u_x} Z = \frac{1}{2}x^4\sigma^4 T_u = 0, \tag{5.127}$$

which can be solved for $T = T_1(\tau, x, y)$. Assuming this form for T and recomputing Z, we find that

$$\partial_{u_{xxx}} Z = -\frac{1}{2}x^4\sigma^4\partial_x T_1, \tag{5.128}$$

so $T = T_2(\tau, y)$. When we compute $\partial_{u_{xx}u_x} Z$, we find a similar equation for X which requires that $X = X_1(\tau, x, y)$. Next, we again reevaluate Z with the new forms for the symmetry coefficients and compute

$$4\partial_{u_{xx}} Z = 4x\sigma^2 X_1 - 2x^3\sigma^2\partial_y T_2 + 2x^2\sigma^2\partial_\tau T_2 - 4x^2\sigma^2\partial_x X_1 = 0. \tag{5.129}$$

Note that we can interpret this as an equation for T_2 or X_1. We take the latter viewpoint, and solve for

$$X_1 = xX_2(\tau, y) + \frac{x}{2}\left(\ln x\partial_\tau T_2 - x\partial_y T_2\right). \tag{5.130}$$

We next find that $\partial_{u_{xy}u_x} Z = 0$ implies that $Y = Y_1(\tau, x, y)$, and the mixed term $0 = \partial_{u_{xy}} Z$ which implies $Y_1 = Y_2(\tau, y)$ which exhausts all the second derivative terms of Z. We now compute $\partial_{u_x^2} Z = x^2\sigma^2 U_{uu}$ so that $U = U_1(\tau, x, y) + uU_2(\tau, x, y)$. We then compute

$$4\partial_{u_y} Z = 4xX_2(\tau, y) - 6x^2\partial_y T_2(\tau, y) - 4x\partial_y Y_2(\tau, y) + 4x\partial_\tau T_2(\tau, y) + 2x\ln x\partial_\tau T_2(\tau, y) + 4\partial_\tau Y_2(\tau, y). \tag{5.131}$$

Here we see that the various unknown functions in this equation do not depend on x. Hence the equation can only be solved if terms with different coefficient x functions vanish independently. Since only one term has a $\ln x$ in it, we need $\partial_\tau T_3 = 0$, so that $T_3 = \alpha$ where α is a constant of integration. Thus we have completely determined the form of T. To find the other symmetry coefficients, we again reevaluate X and compute

$$\partial_{u_y} Z = xX_2 - x\partial_y Y_2 + \partial_\tau Y_2, \tag{5.132}$$

which implies that $\partial_\tau Y_2(\tau, y) = 0$, so $Y_2 = Y_3(y)$. Substituting this back into the previous equation then forces

$$0 = xX_2 - x\partial_y Y_3 \to X_2 = \partial_y Y_3. \tag{5.133}$$

We now begin to further refine the form of U, by calculating

$$\partial_{u_x} Z = x^2\sigma^2\partial_x U_2 - x^2\partial_y^2 Y_3, \tag{5.134}$$

so that $U_2 = U_3(\tau, y) + x\partial_y^2 Y_3/\sigma^2$. We next compute

$$\partial_u Z = \frac{rx}{\sigma^2}\partial_y^2 Y_3 + \frac{x^2}{\sigma^2}\partial_y^3 Y_3 + x\partial_y U_3 - \partial_\tau U_3. \tag{5.135}$$

The x^2 term must vanish since this is a quadrate equation in x, which implies that

$$Y_3 = \beta + y\gamma + y^2\delta. \tag{5.136}$$

We are nearly finished. We now find that our previous equation has simplified to

$$0 = \frac{2rx}{\sigma^2}\delta + x\partial_y U_3 - \partial_\tau U_3, \tag{5.137}$$

where the term constant in x must vanish, so $U_3 = U_4(y)$. We can then set the linear term to zero and solve to find that

$$U_4(y) = -\frac{2ry\delta}{\sigma^2} + \epsilon, \tag{5.138}$$

which completes the computation. We next make a few aesthetic transformations. First, let $\delta \to \sigma^2\delta$ and then transform $\beta \to \delta$, $\gamma \to \beta$, and $\delta \to \gamma$.

Then the general solution for the symmetry coefficients can be expressed as

$$\begin{pmatrix} \mathcal{T} \\ X \\ Y \\ \mathcal{V} \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ x \\ y \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 2\sigma^2 xy \\ \sigma^2 y^2 \\ 2(x - ry)v \end{pmatrix} + \delta \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \epsilon \begin{pmatrix} 0 \\ 0 \\ 0 \\ v \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ v_0 \end{pmatrix}, \tag{5.139}$$

where $v_0 = v_0(\tau, x, y)$ is any fixed solution of the Asian equation (5.126), and α, β, γ, δ, and ϵ are arbitrary constants. We see these symmetries look less interesting than those which we found in the case of the heat equation; however, they still yield useful information about exact solutions of the Asian equation.

We will see that the symmetry associated with the vector field parameterized by γ in (5.139) leads to a new option with exceptionally simple dynamics and which is C^∞ in all values of the relevant state variables, i.e. in the running average y and stock price x. Before considering this, we first demonstrate how the reduction (5.122) arises from the point symmetry group of the Asian equation (5.126). The reduction is realized by choosing

$$\alpha = \beta - 1 = \gamma = \delta + KT = \epsilon - 1 = u_0 = 0, \tag{5.140}$$

in (5.139). More importantly, in (5.140) we have chosen $\gamma = 0$ in (5.139), and thus have avoided using the new symmetry in order to generate the reduction. We relax this restriction below.

In order to produce the desired reduction (5.122), in addition to choosing the parameters in (5.139) according to (5.140), we seek an associated group invariant solution $u = u(s, \tau, x, y)$ of the Asian equation (5.126), i.e. one in which we demand that the trivial dynamics

$$\partial_s u(s, \tau, x, y) = -\left(T(\tau, x, y, u)\partial_\tau + X(\tau, x, y, u)\partial_x + Y(\tau, x, y, u)\partial_y\right) u + U(\tau, x, y, u) \tag{5.141}$$

associated with (5.127) hold in the space of symmetries, at least for all values of the symmetry coordinate s in a small enough neighborhood of $s = 0$. Thus, instead of constructing a family of solutions parameterized by s that stem fro particular solutions $u = u(0, \tau, x, y)$ of the Asian equation (5.126) we find special solutions of the form $u = u(0, \tau, x, y)$ of the symmetry equation (5.127) for which $u(s, \tau, x, y) = u(0, \tau, x, y)$ holds uniformly in a neighborhood of $s = 0$. The theory described here indicates that there is such a class of "static" solutions of the symmetry equation (5.127), i.e. solutions of the group invariant equation (5.141), which will also be a class of solutions of the Asian equation. With the parameters in (5.139) chosen as in (5.140), one finds that the general solution $v(s, \tau, x, y) = v(0, \tau, x, y) = v(\tau, x, y)$ of the group invariance equation (5.141) is of the form

$$v(x, y, \tau) = xh(x^{-1}(y/T - K), \tau) \equiv xh(z, \tau), \tag{5.142}$$

where h is any sufficiently smooth function of its arguments. One can confirm that there are solutions $u(x, y, \tau)$ of the Asian equation (5.126) of the form (5.142) by noting that

$$\begin{aligned} 0 &= \left(\partial_\tau - \left(-r + rx\partial_x + x\partial_y + \frac{1}{2}\sigma^2 x^2 \partial_x^2\right)\right) v \\ &= \left(\partial_\tau - \left(-r + rx\partial_x + x\partial_y + \frac{1}{2}\sigma^2 x^2 \partial_x^2\right)\right) x h(x^{-1}(y/T - K), \tau) \\ &= x\left(\partial_\tau - \left(L + \frac{1}{2}\sigma^2 z^2 \partial_z^2\right)\right) h(z, \tau), \end{aligned} \tag{5.143}$$

where $L \equiv (T^{-1} - rz)\partial_z$, and where the last step in (5.143) requires a bit of algebra. Thus we have found that there exists solutions of the Asian equation (5.126) of the form indicated in (5.142), and have found that such solutions satisfy the reduced equation

$$0 = \left(\partial_\tau - \left(L + \frac{1}{2}\sigma^2 z^2 \partial_z^2\right)\right) h(z, \tau), \tag{5.144}$$

which is closely related to the reduction (5.122). We recover this reduction precisely by composing away the evolution associated with the deterministic advection term L in (5.144), i.e. by defining $g = g(z, \tau)$ via

$$\begin{aligned} h(z, \tau) &= \exp(\tau L\partial_z) g(z, \tau) = g\left(e^{-r\tau} z + \frac{1 - e^{-r\tau}}{rT}, \tau\right) \\ &= g\left(\frac{e^{-r\tau} z x + \Delta_X(\tau) x}{x}, \tau\right) = g\left(\frac{e^{-r\tau}(y/T - K) + \Delta_X(\tau) x}{x}, \tau\right) = g\left(\frac{X(x, y, \tau)}{x}, \tau\right). \end{aligned} \tag{5.145}$$

Here we used the replicating portfolio X of the Asian forward, as in (5.124). Thus, with the relevant solution to the Asian equation (5.126) given by (5.142), (5.146), we find

$$v(x, y, \tau) = x g\left(\frac{X(x, y, \tau)}{x}, \tau\right), \tag{5.146}$$

which is (5.123). Moreover, conjugating the deterministic evolution $\exp(\tau L\partial_z)$ with the geometric diffusion term in (5.144) finally yields the reduction (5.122): one finds that in (5.144)

$$z^2 \partial_z^2 \mapsto \exp(-\tau L) z^2 \partial_z^2 \exp(\tau L) = (z - \Delta_X(\tau))^2 \partial_z^2, \tag{5.147}$$

because of the definition (5.145), giving the relevant geometric diffusion term in (5.122). That is, one can show

$$\begin{aligned} &\left(\partial_\tau - \left(L + \frac{1}{2}\sigma^2 z^2 \partial_z^2\right)\right) h(z, \tau) = \left(\partial_\tau - \left(L + \frac{1}{2}\sigma^2 z^2 \partial_z^2\right)\right) \exp(\tau L\partial_z) g(z, \tau) \\ &= \exp(\tau L\partial_z) \exp(-\tau L\partial_z) \left(\partial_\tau - \left(L + \frac{1}{2}\sigma^2 z^2 \partial_z^2\right)\right) \exp(\tau L\partial_z) g(z, \tau) \\ &= \exp(\tau L\partial_z) \left(\partial_\tau - \frac{1}{2}\sigma^2 (z - \Delta_X(\tau))^2 \partial_z^2\right) g(z, \tau), \end{aligned} \tag{5.148}$$

which is equivalent to equation (5.122). Finally, from (5.142) and (5.145), one finds the Asian call option boundary data $v(x, y, 0) = (y/T - K)^+$ requires $h(z, 0) = z^+ = g(z, 0)$.

We now turn attention to generating a novel reduction of the Asian equation by considering all time invariant and homogeneous point reductions of the Asian equation (5.126), i.e. by setting $\alpha = u_0 = 0$ in (5.139) but leaving all other parameters arbitrary. More importantly, we consider the new symmetry expressed in (5.139) by allowing $\gamma \neq 0$. We again integrate the associated symmetry invariance equation (5.141) to find a more interesting analog of (5.142), namely, the general solution of (5.141) with these parameters, which can be represented in the form

$$u(\tau,x,y)=\frac{\exp\left[\sigma^{-2}\left(x\frac{Q'(y)}{Q(y)}-\frac{2\Xi}{\Lambda}\operatorname{arctanh}\left(\frac{Q'(y)}{\Lambda}\right)\right)\right]}{(Q(y))^{r\sigma^{-2}}}F\left(\tau,\frac{x}{Q(y)}\right). \tag{5.149}$$

In (5.149), we have introduced

$$\Lambda=\sqrt{\beta^2-4\sigma^2\gamma\delta},\quad \Xi=r\beta+\sigma^2\epsilon,\quad Q(y)=\delta+\beta y+\sigma^2\gamma y^2, \tag{5.150}$$

and $F=F(\tau,z)$ as any sufficiently smooth function in both arguments. One confirms that there are solutions $u(x,y,\tau)$ of the Asian equation (5.126) of the form (5.149) by noting that substitution of (5.149) into (5.126) determines a differential equation for $F=F(\tau,z)$, namely

$$\partial_\tau F=\left(-r+rz\partial_z+\frac{1}{2}\sigma^2z^2\partial_z^2\right)F+\sigma^{-2}\left(\Xi z-\frac{1}{2}\Lambda^2z^2\right)F. \tag{5.151}$$

We emphasize that, up to simple variable changes, (5.151) gives the most general time invariant, homogeneous point reduction of the Asian equation (5.126). In particular, one recovers the reduction (5.143) by choosing the parameters in (5.150) defined in (5.140), while simultaneously using the independent variable

$$z'=\frac{1}{z}=\frac{Q(y)}{x}=\frac{\delta+\beta y+\sigma^2\gamma y^2}{x}=\frac{-KT+y}{x}, \tag{5.152}$$

in (5.149) rather than $z=x/Q(y)$. On the one hand, the choice $z=x/Q(y)$ is evidently more natural from the point of view of making a connection with the BSM equation: We note that (5.151) becomes the BSM equation as Ξ and Λ tend to 0, and in this limit one finds that (5.149) gives

$$u(\tau,x,y)=\frac{\exp\left(-2\sigma^{-2}\frac{x}{\beta'-y}\right)}{(\beta'-y)^{2r\sigma^{-2}}}F\left(\tau,\frac{x}{(\beta'-y)^2}\right). \tag{5.153}$$

On the other hand, we can construct a new, C^∞ put-like option with expiration T and "cut-off" price K by choosing $\beta'=KT$ in (5.153) and using the new independent variable w given by

$$\frac{KT}{\sqrt{K}}w=\frac{1}{\sqrt{z}}=\frac{\beta'-y}{\sqrt{x}}=\frac{KT-y}{\sqrt{x}}. \tag{5.154}$$

Defining $G=G(\tau,w)$ through

$$F(\tau,z)=CG\left(t,\frac{\sqrt{K}}{KT}\frac{1}{\sqrt{z}}\right)=CG\left(\tau,\frac{1-y/KT}{\sqrt{x/K}}\right)\equiv CG(\tau,w), \tag{5.155}$$

where

$$G(0,w)=\theta^+(w)=\theta^+\left(\frac{1-y/KT}{\sqrt{x/K}}\right)=\theta^+(1-y/KT), \tag{5.156}$$

we find the desired put like payoff. (In (5.156), θ^+ indicates the Heaviside or unit step function, which has nontrivial support only for positive values of its argument.) In summary, using (5.153) and (5.155) (with a certain useful choice of C), one finds that an Asian-like option with payoff

$$u(0,x,y) = \frac{K}{2}\frac{T\sigma^2}{1-rT}\frac{\exp\left(-\frac{2}{T\sigma^2}\frac{x/K}{1-y/KT}\right)}{\left(1-y/KT\right)^{2r\sigma^{-2}}}\theta^+\left(1-y/KT\right), \tag{5.157}$$

has values at other times τ remaining until expiration given by

$$u(\tau,x,y) = \frac{K}{2}\frac{T\sigma^2}{1-rT}\frac{\exp\left(-\frac{2}{T\sigma^2}\frac{x/K}{1-y/KT}\right)}{\left(1-y/KT\right)^{2r\sigma^{-2}}}G\left(\tau, \frac{1-y/KT}{\sqrt{x/K}}\right), \tag{5.158}$$

where $G = G(\tau, w)$ satisfies the initial value problem

$$\partial_\tau G = \left(-r + \left(\frac{3}{8}\sigma^2 - \frac{r}{2}\right)w\partial_w + \frac{1}{8}\sigma^2 w^2 \partial_w^2\right)G, \quad G(0,w) = \theta^+(w). \tag{5.159}$$

With the simple initial data indicated, the BSM variant (5.159) has a simple solution $G(\tau, w) = e^{-r\tau}\theta^+(w)$. Thus we find that an Asian-like option with market dependent payoff given by (5.157) has values at other times τ before expiration given by

$$u(\tau,x,y) = e^{-r\tau}\frac{K}{2}\frac{T\sigma^2}{1-rT}\frac{\exp\left(-\frac{2}{T\sigma^2}\frac{x/K}{1-y/KT}\right)}{\left(1-y/KT\right)^{2r\sigma^{-2}}}\theta^+\left(1-y/KT\right). \tag{5.160}$$

Here we evidently restrict to expirations T short enough so that $rT < 1$. The choice of

$$C = \frac{K}{2}\frac{T\sigma^2}{1-rT}, \tag{5.161}$$

in (5.155) requires the restriction $rT < 1$, and the factor in (5.160) was chosen so that

$$\left.\frac{d}{dx}u(\tau,x,Tx)\right|_{x,\tau=0} = -1, \tag{5.162}$$

which then mimics the behavior of a standard European put in the regime of maximum payoff. To see this behavior explicitly, while noting other important differences in this payoff structure from that of a standard European put, see Figure 33. There we plot $u(\tau, x, Tx)$ against the underlyings indicated by x and $y = xT$ with parameter values given by $T\sigma^2 = 0.45$, $rT = 0.40$, and $K = 80$. In this case, one finds that

$$u(0,0,0) = \frac{K}{2}\frac{T\sigma^2}{1-rT} = \frac{80}{2}\frac{0.45}{1-0.4} = 30, \tag{5.163}$$

and the option is similar in value to that of a European put with strike $K = 30$. In order to compare to this contract, in Figure 33 we also plot the payoff of such a European put. There we see that for values of the underlying indicated by x and $y = xT$ for $K \in [0, 80]$ at expiration the value of the new option exceeds that of the standard European put with strike $K = 30$. On the other hand, it should also be clear that, for values of the underlyings x and $y = xT$ below $K = 80$, the value at expiration of the new option is exceeded by that of the standard European put with strike $K = 80$. In Figure 33 we plot $v(\tau, x, Tx)$ versus x for values of time $\tau = \{1T, 0.75T, 0.5T, 0.25T\}$ until expiration, as well as at expiration, $\tau = 0$. Thicker curves correspond to times near expiration. As we have mentioned previously, and as should be clear from either formula (5.160) or Figure 33, this analogue of the European put is C^∞ in all values of its underlyings, and thus at all times up to and including its expiration.

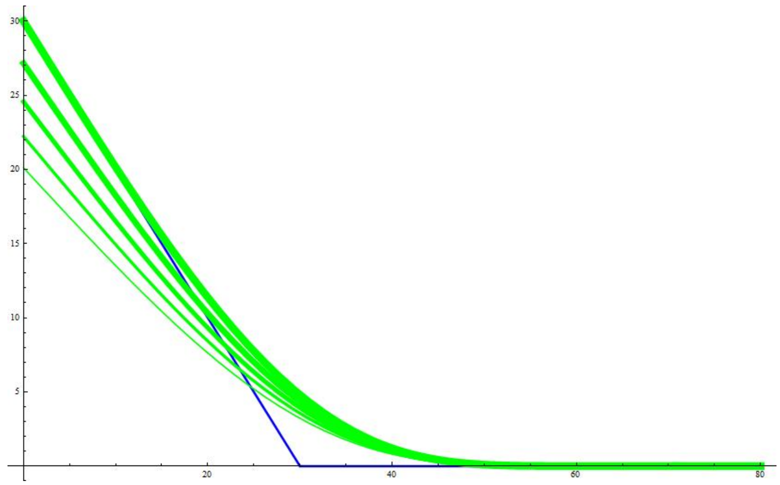

FIGURE 33. $v(\tau, x, xT)$ versus underlying values x and $y = xT$ for times $\tau = 1T, 0.75T, 0.50T, 0.25T$, and $\tau = 0$ until expiration.

Finally, we also note that from (5.160) one finds

$$\Delta \equiv \frac{d}{dx} v(\tau, x, y) < 0, \tag{5.164}$$

$$\Gamma \equiv \frac{d^2}{dx^2} v(\tau, x, y) > 0, \tag{5.165}$$

for all values of x, y. Equation (5.164) shows that to hedge the short position in this put-like option, one must short the underlying stock in the amount indicated, and (5.165) indicates that hedging this short position will always be stable under sudden movements of the main underlying x. See for example Shreve [144, p. 280] to see that, like the European, replicating any Asian-like option is associated with Δ-hedging in the form indicated in (5.164).

In summary, we have constructed the complete maximal 6-dimensional group of point symmetries of the Asian option PDE, thereby extending the known 5-dimensional space of symmetries. The generator of the new symmetry has the unique feature that is depends explicitly on market model parameters. We have demonstrated how group invariant solutions of the Asian PDE (5.126) can be used to construct a novel derivative on a stock and its mean value that, within the indicated simple market model, has $1+1$ dimensional dynamics similar to those arising from the usual Asian option payoff. For this demonstration, we have principally used the new symmetry, which is generated by the infinitesimal symmetry generator associated with the parameter γ in the complete vector space of symmetry generators demonstrated in (5.139). The put-like and market-dependent payoff generated by this new symmetry was shown to have precisely the dynamics of the relevant bond, and also to be C^∞ and have uniform convexity in the financial variable. Thus we present it as an option which should have particularly favorable hedging properties for all values of its contractual underlyings and at all times during the contract.

We emphasize that (5.151) gives the most general time invariant, homogeneous point reduction of the Asian equation (5.126), which includes then all known reductions as special cases. Importantly, equation (5.151) is a simple variation of the BSM equation, and reduces exactly to this for many nontrivial payoff choices, including the one we have described here in some detail. “Nontrivial” in this context means that a claim is written on both the final value of the underlying stock as well as its time average. All such nontrivial claims whose evolutions are described by the BSM equation thus have closed form solutions; computing the time evolving values of any such option (with benign boundary conditions) reduces to integrating the contractual claim against the relevant BSM Green's function.

Before closing, we want to provide the results of a symmetry analysis where we let $r \to 0$ in the Asian equation. Specifically, we find the symmetries of the simpler equation

$$u_\tau = \frac{\sigma^2}{2} x^2 u_{xx} + x u_y. \tag{5.166}$$

It is not possible to compute the symmetry coefficients exactly in this case as we did in the version where $\neq 0$. In particular, we find that the best representation for the coefficients that we can provide is

$$T = \alpha, \quad X = x(\gamma + y\delta), \quad Y = Y_1(\tau, x, y), \quad U = \left(\epsilon - \frac{\gamma}{2} + \frac{x\delta}{\sigma^2}\right) u + U_1(\tau, x, y) \tag{5.167}$$

where here U_1 is an arbitrary solution to the equation, and Y_1 satisfies

$$\partial_\tau Y_1 - \frac{1}{2} x^2 \sigma^2 \partial_x^2 Y_1 - x \partial_y Y_1 + x\gamma + xy\delta. \tag{5.168}$$

Note this is a sourced version of equation (5.166). Thus one could attempt to construct simple solutions to this equation in order generate potentially useful symmetries; however solving this equation in general is tantamount to solving the Asian equation itself. Thus is it is sometimes more interesting to studying pricing equations with non-trivial interest rates.

5.5. **Symmetries of Non Standard Asian Equations.** We now turn attention away from the standard Asian equation and consider a more general class of numeraire invariant analogues. In order to set up the equation, we first need to introduce some notation and follow the set up given in Vecer [160].

Let X_t and Y_t denote two processes that represent the quantity of two assets held at time t, rather than the price of these assets in some reference currency. The price at time t of an asset X in terms of another asset Y is a pairwise relationship of the two assets which is represented by X_Y; this can be regarded as the number of assets Y required to obtain one unit of an asset X. For instance, a natural choice of X and Y is a stock S and dollar \$. In this case, the dollar price of a stock will be denoted by $S_\$$. In the context of foreign exchange, a possible choice for the two assets X and Y is two currencies such as the dollar \$ or the euro €. The price of dollars in terms of Euros is known as an exchange rate, and in this notation will be denoted by $€_\$$. The reciprocal relationship is given by $\$_€$.

Typical financial contracts, such as options, are settled in two or more underlying assets. For example, a European call option has payoff $(X_T - KY_T)^+$, where T is the maturity time of the contract. This contract can be settled at the maturity time T in three natural ways. First, it can be settled in both assets X and Y where the holder has the right to increase his position in the asset X by one unit and decrease his position in the asset Y by K units. It can also be settled only in the asset Y where the holder receives $(X_Y(T) - K)^+$ units of an asset Y. Finally, it can be settled only in the asset X where the holder receives $(1 - KY_X(T))^+$ units of an asset X. Thus there are typically two or more reference assets that could be used when pricing derivative contracts.

The cornerstone of the modern pricing theory of financial contracts is the no-arbitrage principle that assumes on average, one cannot make instantaneous risk free profit from trading in market assets. For no-arbitrage assets X and Y, we will use a slightly weaker result than the Fundamental Theorem of Asset Pricing which when put in the "numeraire invariant" context can be summarized by the following

Theorem 5.3. *Let X and Y be no-arbitrage assets with positive associated price processes. If there is a measure $\mathbb{P}^Y$ such that the price process $X_Y(t)$ is a $\mathbb{P}^Y$ martingale, then there is no arbitrage in the associated pricing model.*

In order to price a contingent claim V whose payoff depends on both X and Y, one can apply this theorem using both X and Y as reference assets. In particular, the price process $V_Y(t)$ is a $\mathbb{P}^Y$ martingale and the price process $V_X(t)$ is a $\mathbb{P}^X$ martingale. Thus we have

$$V_Y(t) = \mathbb{E}_t^Y \left[V_Y(T)\right], \tag{5.169}$$

when we express the price of V in terms of the asset Y. We also have the parallel relationship,

$$V_X(t) = \mathbb{E}_t^X \left[V_X(T)\right], \tag{5.170}$$

when we express the price V in terms of the asset X. The above relationships state that V_t is worth $\mathbb{E}_t^Y \left[V_Y(T)\right]$ units of Y_t or equivalently, $\mathbb{E}_t^X \left[V_X(T)\right]$ units of X_t. This gives us the following relationship which is known as the Change of Numeraire Formula:

Theorem 5.4. *Let X and Y be no-arbitrage assets with positive price process and let V be a no-arbitrage asset. Then*

$$V_t = \mathbb{E}_t^Y \left[V_Y(T)\right] Y_t = \mathbb{E}_t^X \left[V_X(T)\right] X_t. \tag{5.171}$$

We will focus on European Asian options where the underlying follows log normal dynamics. Asian options have a payoff that depends on a contract whose payoff is given by the Asian forward

$$\bar{F}_T = \left[\int_0^T X_Y(t)\mu(dt)\right] Y_T. \tag{5.172}$$

An Asian forward is a contract that pays off a number of units of an asset Y, where this number is the weighted average price of an asset X with respect to the asset Y less a strike price K. Asian options depend on three no-arbitrage assets: X, Y, and $\bar{F}$. However, only X and Y can be used as reference assets since the Asian forward $\bar{F}$ can have a zero or negative price and is thus not an appropriate numeraire.

Asian options are contracts that depend on underlying assets X and Y and upon the average of the price process $X_Y(t)$. The average price process is captured through the Asian Forward.

The weight of the average at different times is given by the measure μ in $\bar{F}_T$. Averaging is chosen to be continuous with uniform weights when

$$\mu(dt) = \frac{1}{T}dt, \tag{5.173}$$

or discrete when

$$\mu(dt) = \frac{1}{n}\sum_{k=1}^{n} \delta_{kT/n}(t)dt. \tag{5.174}$$

We can use change of numeraire formula as long as the assets X and Y are no-arbitrage assets. This is not the case when X is a stock S and Y is dollars \$, when the Asian Forward contract becomes

$$\bar{F}_T = \left[\int_0^T S_{\$}(t)\mu(dt)\right] \$(T). \tag{5.175}$$

However, we can still rewrite this contract in terms of no-arbitrage assets when the bond price follows a deterministic term structure $B_t^T = e^{-r(T-t)}\$(t)$ as

$$\bar{F}_T = \left[\int_0^T S_{B^T}(t)e^{-r(T-t)}\mu(dt)\right] B_T^T, \tag{5.176}$$

which is of the form (5.172), where the underlying two no arbitrage assets are S and B^T. This is helpful since hedging must be done in no-arbitrage assets as opposed to arbitrage assets such as currencies. Thus it also makes sense to consider continuous averaging in the form

$$\mu(dt) = \frac{1}{T}e^{-r(T-t)}dt, \tag{5.177}$$

or discrete averaging in the form

$$\mu(dt) = \frac{1}{n}\sum_{k=1}^{n} \delta_{kT/n}(t)e^{-r(T-t)}dt. \tag{5.178}$$

Since the payoff of the Asian option can be expressed using both assets Y or X, we again have two natural definitions for Asian type options. An Asian option can be considered either as a contract that pays off $f\left(X_Y(T), \bar{F}_Y(T)\right)$ units of Y_T, or as a contact that pays off $g\left(Y_X(T), (F_X(T)\right)$ units of X_T.

The two payoffs need to represent the same contract, so we must have

$$f\left(X_Y(T), \bar{F}_Y(T)\right) Y_T = g\left(Y_X(T), \bar{F}_X(T)\right) X_T, \tag{5.179}$$

which leads to the following symmetric relationship between f and g

$$f(x,y) = g\left(\frac{1}{x}, \frac{y}{x}\right) x, \quad \text{or} \quad g(x,y) = f\left(\frac{1}{x}, \frac{y}{x}\right) x. \tag{5.180}$$

For instance, the Asian Call option with a fixed strike pays off

$$\left(\int_0^T X_Y(t)\mu(dt) - K\right)^+ Y_T. \tag{5.181}$$

This corresponds to the choice of $f(x,y) = (y-K)^+$ or $g(x,y) = (y-Kx)^+$ in the above definition of the Asian option. An Asian call option with a floating strike pays off

$$\left(\int_0^T X_Y(t)\mu(dt)Y_T - KX_T\right)^+, \tag{5.182}$$

which corresponds to the choice of $g(x,y) = (y-K)^+$ or $f(x,y) = (y-Kx)^+$. Asian options with fixed or floating strike are the two most typical payoffs.

Since the price of an Asian option depends on the price of the Asian forward, finding the replicating portfolio for the Asian Forward contract will prove crucial for pricing the option contract. Moreover, it turns out that the hedge of the forward is model independent. Let write the hedge for the Asian Forward in the following form

$$\bar{F}_t = \bar{\Delta}_t^X X_t + \bar{\Delta}_t^Y Y_t, \tag{5.183}$$

where $\bar{\Delta}_t^X$ represents the number of units of an asset X at time t, and $\bar{\Delta}_t^Y$ represents the number of units of an asset Y at time t. The hedging portfolio should be self-financing, so in particular, it must satisfy

$$d\bar{F}_Y(t) = \bar{\Delta}_t^X dX_Y(t). \tag{5.184}$$

Theorem 5.5. *Let X and Y be two no-arbitrage assets. Then the replicating portfolio for the Asian Forward contract that pays off*

$$\bar{F}_T = \left[\int_0^T X_Y(t)\mu(dt)\right] Y_T, \tag{5.185}$$

is given by

$$\bar{F}_t = \left[\int_t^T \mu(ds)\right] X_t + \left[\int_0^t X_Y(s)\mu(ds)\right] Y_t. \tag{5.186}$$

This result does not depend on the dynamics of the price $X_Y(t)$.

We now consider how to price an Asian option where the underlying is assumed to follow log normal dynamics. In the lognormal model, we assume the following price process dynamics:

$$dX_Y(t) = \sigma X_Y(t) dW_t^Y, \tag{5.187}$$

with a corresponding expression for the inverse price given by

$$dY_X(t) = \sigma Y_X(t) dW_t^X. \tag{5.188}$$

The relationship between the Brownian motions W_t^Y and W_t^X is given by

$$dW_t^X = -dW_t^Y + \sigma dt, \tag{5.189}$$

and the Radon-Nikodým derivative $Z_t = d\mathbb{P}^X / d\mathbb{P}^Y$ is a log normal process given by

$$Z_t = \frac{d\mathbb{P}^X}{d\mathbb{P}^Y} = \exp\left(\sigma W_t^Y - \frac{1}{2}\sigma^2 t\right). \tag{5.190}$$

Asian options pay off either $f\left(X_Y(T), \bar{F}_Y(T)\right)$ units of Y_T, or $g\left(Y_X(T), \bar{F}_X(T)\right)$ units of X_T. When $f(x,y) = g(1/x, y/x)x$, the two payoffs represent the same contract. Let V be the Asian option contract. We have

$$V_T = f\left(X_Y(T), \bar{F}_Y(T)\right) Y_T = g\left(Y_X(T), \bar{F}_X(T)\right) X_T, \tag{5.191}$$

and thus we can compute the price of the contract with respect to both reference assets X and Y from the martingale property of $V_Y(t)$ and $V_X(t)$. In particular,

$$V_Y(t) = \mathbb{E}_t^Y \left[V_Y(T)\right], \tag{5.192}$$

when we express the price of the claim V in terms of the asset Y and

$$V_X(t) = \mathbb{E}_t^X \left[V_X(T)\right], \tag{5.193}$$

when we express the price of the claim V in terms of the asset X. In the geometric Brownian motion model, the price process is Markovian and thus we can define the following functions

$$u(t,x,y) = \mathbb{E}^Y \left[f\left(X_Y(T), \bar{F}_Y(T)\right) \middle| X_Y(t) = x, \bar{F}_Y(t) = y \right], \tag{5.194}$$

which represents the price of the Asian option in terms of the reference asset Y, and

$$v(t,x,y) = \mathbb{E}^X \left[g\left(Y_X(T), \bar{F}_X(T)\right) \middle| Y_X(t) = x, \bar{F}_X(t) = y \right], \tag{5.195}$$

which represents the price of the Asian option in terms of the reference asset X. The change of numeraire formula now reads as

$$V_t = u\left(t, X_Y(t), (\bar{F}/Y)_t\right) Y_t = v\left(t, Y_X(t), \bar{F}_X(t)\right) X_t. \tag{5.196}$$

This gives us the following relationship of the functions u and v:

$$u(t,x,y) = v\left(t, \frac{1}{x}, \frac{y}{x}\right)x, \quad \text{or} \quad v(t,x,y) = u\left(t, \frac{1}{x}, \frac{y}{x}\right)x. \tag{5.197}$$

We now consider a type of Asian option whose payoff is given by $f\left(X_Y(T), \bar{F}_Y(T)\right)Y_T$. A typical example is an Asian call option with the payoff function $f(x,y) = y^+$, or an Asian put option with payoff function $f(x,y) = y^-$. The price of the Asian forward $\bar{F}$ with respect to the reference asset Y has the following evolution:

$$d\bar{F}_Y(t) = \bar{\Delta}_t^X dX_Y(t) = \bar{\Delta}_t^X \sigma X_Y(t) dW_t^Y. \tag{5.198}$$

The price of the Asian Option contract with payoff $V_T = f\left(X_Y(T), \bar{F}_Y(T)\right)Y_T$ satisfies

$$V_Y(t) = \mathbb{E}_t^Y\left[\frac{f\left(X_Y(T), \bar{F}_Y(T)\right)Y_T}{Y_T}\right] = \mathbb{E}^Y\left[f\left(X_Y(T), \bar{F}_Y(T)\right)\middle| X_Y(t), \bar{F}_Y(t)\right] \tag{5.199}$$

when computed with respect to the reference asset Y.

Let us denote the price of the Asian option with respect to the reference asset Y by

$$u(t,x,y) = \mathbb{E}^Y\left[f\left(X_Y(T), \bar{F}_Y(T)\right)\middle| X_Y(t) = x, \bar{F}_Y(t) = y\right]. \tag{5.200}$$

This price $u\left(t, X_Y(t), \bar{F}_Y(t)\right)$ is a $\mathbb{P}^Y$ martingale, and thus du has a zero dt term by the Martingale representation theorem. Using Ito's formula, we get

$$du = u_t dt + u_x dX_Y(t) + u_y d\bar{F}_Y(t) \tag{5.201}$$

$$+\frac{1}{2}\left[u_{xx} dX_Y(t)^2 + 2u_{xy} dX_Y(t) d\bar{F}_Y(t) + u_{yy} d\bar{F}_Y(t)^2\right] \tag{5.202}$$

$$= [u_t + \frac{1}{2}\sigma^2 x^2 (u_{xx} + 2\bar{\Delta}_t^X u_{xy} + (\bar{\Delta}_t^X)^2 u_{yy})]dt + u_x dX_Y(t) + u_y d\bar{F}_Y(t). \tag{5.203}$$

Since the dt term is zero, we obtain the following partial differential equation:

$$u_t(t,x,y) + \frac{1}{2}\sigma^2 x^2\left[u_{xx}(t,x,y) + 2\bar{\Delta}_t^X u_{xy}(t,x,y) + (\bar{\Delta}_t^X)^2 u_{yy}(t,x,y)\right] = 0. \tag{5.204}$$

The terminal condition is given by

$$u(T,x,y) = f(x,y). \tag{5.205}$$

When the payoff of the option depends only on the price of the Asian forward $\bar{F}_Y(T)$, the terminal condition simplifies to

$$u(T,x,y) = f(y), \tag{5.206}$$

but the pricing partial differential equation remains the same since the pricing problem still depends on $X_Y(t)$. The price of the Asian forward $\bar{F}_Y(t)$ is Markovian in the pair $\left(\bar{F}_Y(t), X_Y(t)\right)$. When the payoff of the option depends only on the price $X_Y(t)$, there is no dependence on the variable y, and the pricing partial differential equation reduces to

$$u_t(t,x) + \frac{1}{2}\sigma^2 x^2 u_{xx}(t,x) = 0, \tag{5.207}$$

with the terminal condition

$$u(T,x) = f(x).$$

This is not surprising, in this case we obtain just the pricing partial differential equation for a European option with a payoff of $f(X_Y(T))$ units of Y_T.

Now consider the Asian option contract that pays off $g\left(Y_X(T), \bar{F}_X(T)\right)$ units of X_T. The evolution of the Asian Forward price under the reference asset X can be rewritten as

$$d\bar{F}_X(t) = \bar{\Delta}_t^Y Y_X(t) = \left[\bar{F}_Y(t) - \bar{\Delta}_t^X X_Y(t)\right] dY_X(t) \tag{5.208}$$

$$= \left[\bar{F}_Y(t) - \bar{\Delta}_t^X X_Y(t)\right] \sigma Y_X(t) dW_t^X = \sigma\left[\bar{F}_X(t) - \bar{\Delta}_t^X\right] dW_t^X. \tag{5.209}$$

The second equality $\bar{\Delta}_t^Y = \bar{F}_Y(t) - \bar{\Delta}_t^X \bar{X}_Y(t)$ follows from the relation $\bar{F}_t = \bar{\Delta}_t^X X_t + \bar{\Delta}_t^Y Y_t$. Note that unlike the price evolution of $\bar{F}_Y(t)$, the price evolution of $\bar{F}_X(t)$ is Markov in just one variable, and thus contracts whose payoff depends only on $\bar{F}_X(T)$ admit a simpler, partial differential equation with one spatial variable.

The price of the Asian option with respect to the reference asset X is given by

$$V_X(t) = \mathbb{E}_t^X\left[\frac{g\left(Y_X(T), \bar{F}_X(T)\right) X_T}{X_T}\right] = \mathbb{E}^X\left[g\left((Y_X(T), \bar{F}_X(T)\right) \middle| Y_X(t), \bar{F}_X(t)\right]. \tag{5.210}$$

If we let

$$v(t, x, y) = \mathbb{E}^X\left[g\left(Y_X(T), \bar{F}_X(T)\right) \middle| Y_X(t) = x, \bar{F}_X(t) = y\right], \tag{5.211}$$

then the price of the Asian option $v(t, Y_X(t), \bar{F}_X(t))$ is a $\mathbb{P}^X$ martingale, and thus the dt term of dv is zero. Using the evolution of the Asian Forward under the reference asset X, we get

$$dv = v_t dt + v_x dY_X(t) + v_y d\bar{F}_X(t) + \frac{1}{2}\left[v_{xx} dY_X(t)^2 + 2v_{xy} dY_X(t) d\bar{F}_X(t) + v_{yy} d\bar{F}_X(t)^2\right] \tag{5.212}$$

$$= \left[v_t + \frac{1}{2}\sigma^2(x^2 v_{xx} + 2x(y - \bar{\Delta}_t^X) v_{xy} + (y - \bar{\Delta}_t^X)^2 v_{yy})\right] dt + v_x dY_X(t) + v_y d\bar{F}_X(t). \tag{5.213}$$

Since the dt term is zero, the following partial differential equation must be satisfied:

$$v_t(t, x, y) + \frac{1}{2}\sigma^2\left[x^2 v_{xx}(t, x, y) + 2x(y - \bar{\Delta}_t^X) v_{xy}(t, x, y) + (y - \bar{\Delta}_t^X)^2 v_{yy}(t, x, y)\right] = 0, \tag{5.214}$$

with the terminal condition

$$v(T, x, y) = g(x, y). \tag{5.215}$$

When the payoff of the Asian option contract does not depend on $Y_X(T)$, but only on $\bar{F}_X(T)$, the pricing equation does not depend on the variable x, and thus it is equivalent to the following reduced partial differential equation:

$$v_t(t, y) + \frac{1}{2}\sigma^2(y - \bar{\Delta}_t^X)^2 v_{yy}(t, y) = 0, \tag{5.216}$$

with the terminal condition

$$v(T, y) = g(y). \tag{5.217}$$

This reduced partial differential equation was first found in Vecer [161], and later studied in Vecer [162].

On the other hand, when the payoff does not depend on $\bar{F}_X(T)$, but only on $Y_X(T)$, the pricing equation does not depend on the variable y, and thus it is equivalent to the partial differential equation that corresponds to the BSM equation with zero interest rates

$$v_t(t, x) + \frac{1}{2}\sigma^2 x^2 v_{xx}(t, x) = 0, \tag{5.218}$$

with the usual terminal condition

$$v(T, x) = g(x). \tag{5.219}$$

5.6. **Symmetries of New Asian Equations.** We now wish to study the symmetries of these Asian type equations. We first apply this method to construct an exact solution of the new Asian type equations. Consider equation (5.204), which can be written as

$$\mathcal{A}u \equiv \left(\partial_\tau - \frac{\sigma^2}{2}x^2\left(\partial_{xx} + 2\Delta(\tau)\partial_x\partial_y + \Delta^2(\tau)\partial_{yy}\right)\right)u(\tau,x,y) = 0, \tag{5.220}$$

where either

$$\Delta(\tau) = \frac{1-e^{-r\tau}}{rT} \text{ for } r>0, \quad \text{or} \quad \Delta(\tau) = \lim_{r\to 0^+}\frac{1-e^{-r\tau}}{rT} = \frac{\tau}{T} \text{ for } r=0. \tag{5.221}$$

We denote the price of the Asian option with respect to the reference asset Y by v and set $\tau = T - t$ to be the time to expiration. Now couple a symmetry equation of the form

$$\partial_s u(\tau,s,x,y) = -Tu_\tau - Xu_x - Yu_y + U, \tag{5.222}$$

where $T = T(\tau,x,y,u(\tau,s,x,y))$ with similar expressions for X, Y, and U, to equation (5.220). Next, in a manner similar to our previous considerations, we consider the compatibility condition $Z \equiv [\partial_s,\partial_\tau]u = 0$ in order to construct the functional forms for the symmetry coefficients. We do not write this condition out explicitly, since it is fairly complicated, but as in the case of the classical Asian equation, we indicate how we extract and solve the PDEs for the symmetry coefficients from Z. We first note that

$$0 = \partial_{u_{xxx}u_x}Z = \frac{1}{2}x^4 T_v, \tag{5.223}$$

so that $T = T_1(\tau,x,y)$. We next find that

$$\partial_{u_{xxx}}Z = \frac{1}{2}x^4\sigma^4\left(\Delta(\tau)\partial_y T_1 + \partial_x T_1\right), \tag{5.224}$$

which can be solved for $T_1 = T_2(\tau, y - x\Delta)$. Next, the equation $\partial_{u_{xx}u_x}Z = 0$ forces $X = X_1(\tau,x,y)$. After reevaluating Z with these new forms of the symmetry coefficients, we find that $\partial_{u_{xx}}Z = 0$ implies

$$X_1 = -\frac{x}{2}\left(x\Delta'\partial_2 T_2 - 2X_2 - \ln x T_\tau\right), \tag{5.225}$$

where here we use the notation ∂_2 to denote partial differentiation of a function with respect to its second argument; thus in the above we are differentiating T_2 with respect to $y - x\Delta$. We next find that $\partial_{u_{yy}u_y}Z = 0$ implies that $Y_1 = Y_1(\tau,x,y)$. We now note that $\partial_{u_{yy}}Z = 0$ can be solved for

$$2Y_1 = 2xX_2\Delta + 2Y_2 + 2xT_2\Delta' - x^2\Delta\Delta'\partial_2 T_2 + x\ln x\partial_\tau T_2. \tag{5.226}$$

Next, $\partial_{u_x^2}Z = 0$ requires that U is affine in v,

$$U = U_1(\tau,x,y) + uU_2(\tau,x,y). \tag{5.227}$$

We next update Z, and then compute

$$4\partial_{u_x}Z = 2x^2\Delta''\partial_2 T_2 - 2x^2\sigma^2\Delta'\partial_2 T_2 + 4x^2\Delta'\partial_2 X_2 - 2x^3(\Delta')^2\partial_2^2 T_2 + x\sigma^2\partial_2 T_2 - 4x\partial_\tau X_2 \tag{5.228}$$

$$+2x^2\Delta'\partial_\tau\partial_2 T_2 + 2x^2\ln x\Delta'\partial_\tau\partial_2 T_2 - 2x\ln x\partial_\tau^2 T_2 - 4x^2\sigma^2\Delta\partial_y U_2 - 4x^2\sigma^2\partial_x U_2. \tag{5.229}$$

Equating different powers of x, we see that $T_2 = \alpha$ where α is a constant of integration. This greatly simplifies the current forms of our other symmetry coefficients. We now can solve $\partial_{u_x}Z = 0$ for

$$V_2 = V_3(\tau, y - x\Delta) + \frac{x}{\sigma^2}\Delta'\partial_2 X_2 - \frac{\ln x}{\sigma^2}\partial_\tau X_2. \tag{5.230}$$

We now can consider the next equation

$$\partial_{v_y} Z = x\Delta' \partial_2 Y_2 - \partial_\tau Y_2 - xX_2\Delta' - x\alpha\Delta'', \tag{5.231}$$

which equating powers in x again requires that

$$Y_2 = Y_3(y - x\Delta), \quad X_2 = -\frac{\alpha\Delta''}{\Delta'} + \partial_2 Y_2(\tau, Y). \tag{5.232}$$

Now, by demanding the constant term vanishes, we can further refined this expression to $Y_2 = Y_3(y - x\Delta)$. Next, we can compute $\partial_v Z = 0$ and again equation powers of x to see that $Y_3^{(3)} = 0$, which has solution

$$Y_3(y - x\Delta) = \beta + \gamma(y - x\Delta) + \delta(y - x\Delta)^2. \tag{5.233}$$

A second result of the equating powers is that

$$V_3 = V_4(\tau) + \frac{2Y\delta\Delta''}{\sigma^2\Delta'}. \tag{5.234}$$

Now to further simplify the form of $\partial_v Z$, we need to assume that

$$\Delta'''\Delta' - (\Delta'')^2 = 0, \tag{5.235}$$

must hold, or else the symmetry group becomes nearly trivial, and as a consequence, it cannot be used to construct reductions of (5.220). Thus either Δ is constant or it must be of the form $\Delta(\tau) = e^{\tau c_1}c_2 + c_3$. We have already chosen Δ to be of this form for specific choices of the c_i for financial reasons, but it is interesting to note that the symmetry analysis in some sense forces this upon us.

After substituting this into $\partial_v Z = 0$, we find that the equation reduces to enforcing $\partial_\tau V_4 = 0$, so we let $V_4 = \epsilon$. We finally note that equating powers of x in $Z = 0$ requires that $2\delta - Y_2'' = 0$, from which we conclude that

$$Y_2 = \rho + \xi Y + \delta Y^2, \tag{5.236}$$

which yields the full form of the symmetry coefficients. Now $Z = 0$ reduces to U_1, being a solution of our initial PDE, which again is an expected symmetry since the equation is linear. Renaming and rescaling constants allows us to write the complete form for the symmetry coefficients as

$$\begin{pmatrix} T \\ X \\ Y \\ U \end{pmatrix} = \alpha \begin{pmatrix} T \\ rTx \\ x \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ x \\ y \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 2\sigma^2 Tx(y - x\Delta) \\ \sigma^2 T(y^2 - x^2\Delta^2) \\ 2(x - rTy)u \end{pmatrix} + \delta \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \epsilon \begin{pmatrix} 0 \\ 0 \\ 0 \\ u \end{pmatrix} + u_0 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{5.237}$$

where $u_0 = u_0(\tau, x, y)$ is any solution of (5.220).

There are a few important points that need to be noted about the calculation that results in the solution (5.237). First, the equation analogous to (5.7) is more complicated; however, we used essentially the same techniques we used in the heat equation in order to obtain a system of PDE for the symmetry coefficients of this equation. This system was solved by solving simple subsystems of PDE and substituting solutions back into the symmetry coefficient system which results in simpler PDEs. Repeating this process several times, one eventually arrives as (5.237). Secondly, we note that for a significant portion of the calculation, we are able to leave $\Delta(\tau)$ as a general function.

Now we turn attention to the construction and interpretation of the Lie point symmetries associated to the five constants of equation (5.237). We first consider the α symmetry by letting $\alpha = 1$ and setting all other constants and u_0 to zero. Then the symmetry equation becomes

$$u_s = -Tu_\tau - rTxu_x - xu_y. \tag{5.238}$$

To solve this equation, we use the standard technique for first order linear PDE and consider the ODE associated ODE system

$$\dot{\tau}(s) = -T, \quad \dot{x}(s) = -rTx(s), \quad \dot{y}(s) = -x, \quad \dot{u}(s) = 0, \tag{5.239}$$

subject to the initial data $(\tau, x, y, v)(0) = (\tau, x, y, u)$. Integrating the system, we find

$$\tau(s) = \tau - Ts, \quad x(s) = xe^{-rTs} = T\Delta'(Ts)x, \quad y(s) = y - \Delta(sT)x, \quad u(s) = u. \tag{5.240}$$

Thus if $u(\tau, x, y)$ solves equation (5.220), then so does

$$u(\tau - Ts, T\Delta'(Ts)x, y - \Delta(sT)x), \quad s \in \mathbb{R}. \tag{5.241}$$

The domain of s should be suitably restricted to make the arguments of v consistent with the domains of the independent variables.

Next, we consider the simpler β-symmetry. The nontrivial part of the associated ODE system is

$$\dot{x} = -x, \quad \dot{y} = -y, \tag{5.242}$$

which has solution $x(s) = xe^{-s}$, $y(s) = ye^{-s}$. Thus if $u(\tau, x, y)$ solves the equation, then so does $u(\tau, xe^{-s}, ye^{-s})$. In other words, equation (5.237) is invariant under simultaneous scaling of spatial variables, which is clear from inspection of the equation.

The γ symmetry is more interesting and will lead to an exact solution. The ODE system for this symmetry is more complicated,

$$\dot{\tau} = 0, \quad \dot{x} = 2\sigma^2 Tx(y - x\Delta), \quad \dot{y} = \sigma^2 T(y^2 - x^2\Delta^2), \quad \dot{u} = 2(x - rTy)u. \tag{5.243}$$

The $\dot{x}$ and $\dot{y}$ equations decouple, so we solve them first to find

$$x(s) = \frac{4}{(2T\sigma^2 sc_1 + c_2)^2}, \quad y(s) = \frac{4\Delta + 2c_1(2T\sigma^2 sc_1 + c_2)}{(2sT\sigma^2 c_1 + c_2)^2}. \tag{5.244}$$

The initial data requires $x = 4/c_2^2$, $y = (4\Delta + 2c_1c_2)/c_2^2$, i.e. $c_2 = \pm\sqrt{x}/2$ or $c_1 = (\pm y \mp x\Delta)x^{-1/2}$, $c_2 = \pm 2x^{-1/2}$. In any case, we find the same solution

$$x(s) = \frac{x}{(1 + sT\sigma^2(y - x\Delta))^2}, \quad y(s) = \frac{y + s\sigma^2 T(y - x\Delta)^2}{(1 + sT\sigma^2(y - x\Delta))^2}. \tag{5.245}$$

We can now solve the u equation to find

$$u(s) = \exp\left(\frac{2sx(rT\Delta - 1)}{sT(-y + x\Delta)\sigma^2 - 1}\right)(1 + sT(y - x\Delta)\sigma^2)^{-2r/\sigma^2}u. \tag{5.246}$$

Thus if $u(\tau, x, y)$ solves equation (5.237), then so does

$$\frac{\exp\left(\frac{2sx(1 - rT\Delta)}{1 + sT(y - x\Delta)\sigma^2}\right)}{(1 + sT(y - x\Delta)\sigma^2)^{2r/\sigma^2}} u\left(\tau, \frac{x}{(1 + sT\sigma^2(y - x\Delta))^2}, \frac{y + s\sigma^2 T(y - x\Delta)^2}{(1 + sT\sigma^2(y - x\Delta))^2}\right). \tag{5.247}$$

The final two symmetries are simple and, like the β symmetry, easily verified by inspection. The δ symmetry corresponds to a y translational symmetry, i.e. if $u(\cdot, y)$ is a solution, then so is $u(\cdot, y + s)$. Finally, the ϵ symmetry is another rescaling symmetry i.e. if u is a solution, then so is $e^{-s}u$.

We also can exponentiate these vector fields just as in the classical analysis; however, we find that they do not give any new information from that which we have previously mentioned. We note that we are not able to exponentiate the γ vector due to the fact that we are not able to solve the nonlinear coupled ODE system associated to the X and Y components of the symmetry vector.

We now consider the second canonical equation (5.214), which we can write in terms of $\tau = T - t$ as

$$u_\tau - \frac{\sigma^2}{2}\left[x^2 u_{xx} + 2x(y - \Delta(\tau))u_{xy} + (y - \Delta(\tau))^2 u_{yy}\right] = 0, \tag{5.248}$$

where we take $\Delta(\tau)$ as before. We again extend u to $u(s, \tau, x, y)$ and couple (5.214) to the symmetry equation (5.127) that we previously used. In this case, the symmetry coefficients are given by

(5.249)

$$\begin{pmatrix} T \\ X \\ Y \\ U \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ -rx \\ T^{-1} - ry \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 0 \\ x \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ x \\ 0 \\ u(1 + r/\sigma^2) \end{pmatrix} + \delta \begin{pmatrix} 0 \\ 0 \\ 0 \\ u \end{pmatrix} + u_0 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Note that the symmetry group is simpler than in the previous case. This should be expected since the new equation has coefficients that are non-autonomous in y which has the effect of restricting the symmetry group. It is interesting to note that the β and γ symmetries have no analogues in the previous computation. This is due to the fact that the explicit x dependence of this equation differs from the previous.

The computation required to conclude the symmetry coefficients take the form (5.249) is again very long. It is similar in form to the previous computation in that one can extract simple partial differential equations from the compatibility condition, solve them to update the forms of the symmetry coefficients, and substitute the resulting expressions back into the compatibility condition which in turn simplifies the equation. However, there are several places where parts of the calculation differs significantly from the previous case, i.e. the computation is not parallel to the previous. It in fact appears more complicated at intermediate stages, but simplifies drastically near the end.

We now compute the point symmetries associated with (5.249). If we take $\alpha = 1$ and set the rest of the parameters to zero, then the symmetry equation becomes

$$v_s = -v_\tau + rxv_x + (ry - T^{-1})v_y = 0, \tag{5.250}$$

which has solution

$$v\left(s - \tau, e^{r\tau}x, \frac{e^{r\tau}(rTy - 1)}{rT}\right). \tag{5.251}$$

For the β symmetry, we find that the symmetry equation $u_s = -xu_y$ has solution $u(\tau, x, y - sx)$. The γ symmetry equation has solution

$$\exp\left(\frac{s(r + \sigma^2)}{\sigma^2}\right) v(\tau, e^{-s}x, y). \tag{5.252}$$

and the δ symmetry just corresponds to a constant rescaling of u.

We next turn attention to the third equation which is given by

$$u_\tau = \frac{\sigma^2}{2}x^2\left((x\Delta - 1)^2 u_{xx} + 2y\Delta(x\Delta - 1)u_{xy} + y^2\Delta^2 u_{yy}\right) \tag{5.253}$$

The symmetry group for this equation is nearly trivial. An analysis similar to the above produces three symmetries. The first is just constant rescaling of the dependent variable, i.e. $w \to \pm e^\lambda w$ where λ is any real number. The second is time translation invariance, and the third is $w \to w + w_0$ where w_0 is any solution of this equation; note like in the other cases that this just represents the linearity of the equation.

References

[1] L. V. Ahlfors. Conformal Invariants: Topics in Geometric Function Theory. *American Mathematical Society Chelsea Publishing* (2010).

[2] L. B. G. Andersen, V. V. Piterbarg. Interest Rate Modeling Volumes 1,2, and 3. *Atlantic Financial Press* (2010).

[3] L. B. G. Andersen. Option Pricing with Quadratic Volatility: a Revisit. Working Paper. Available at SSRN: `http://ssrn.com/abstract=1118399` (Feb. 2010).

[4] L. B. G. Andersen, R. B. Ratcliffe. Extended Libor Market Models with Stochastic Volatlity. *Journal of Computational Finance*, **9** No. 1, p. 1-40. Available at `http://ssrn.com/abstract=844190` (2005).

[5] L. B. G. Andersen, V. V. Piterbarg. Moment Explosions in Stochastic Volatility Models. *Finance and Stochastics*, **11** No. 1, p. 29-50. Available at `http://ssrn.com/abstract=559481` (2007).

[6] J. Andreasen, B. Jensen, R. Poulsen. Eight Valuation Methods in Financial Mathematics: The Black-Scholes Formula as an Example. *Mathematical Scientist* **23** No. 1, p. 18-40, (1998).

[7] A. Antonov, T. Misirpashaev. Projection on a Quadratic Model by Asymptotic Expansion with an Application to LMM Swaption. Working Paper. Available at `http://ssrn.com/abstract=1421139` (Jun. 2009).

[8] V. I. Arnol'd, Mathematical Methods of Classical Mechanics, *Springer-Verlag* (1989).

[9] G. B. Arfken, H. J. Weber, Mathematical Methods for Physicists, Sixth Edition: A Comprehensive Guide. *Academic Press* (2005).

[10] G. B. Arous, E. P. Hsu, P. Laurence, T. H. Ouyang, Asymptotics of Implied Volatility in Stochastic Volatility Models. Preprint (2010).

[11] M. Avellaneda, From SABR to Geodesics. Talk. `math.nyu.edu/faculty/avellane/AvellanedaQCongress2005.pdf` (2005).

[12] I. G. Avramidi, Analytic and Geometric Methods for Heat Kernel Applications in Finance. Available at `http://infohost.nmt.edu/~iavramid/notes/hkt/hktutorial13.pdf` (2007).

[13] I. G. Avramidi. Heat Kernel on Homogeneous Bundles over Symmetric Spaces. *Comm. Math. Phys.* **288** No. 3, p. 963-1006. Available at `http://arxiv.org/abs/math/0701489` (2008).

[14] I. G. Avramidi, R. Schimming. Algorithms for the Calculation of the Heat Kernel Coefficients. *Quantum Theory under the Influence of External Conditions*, Ed. M. Bordag, Teubner-Texte zur Physik, **30**, p. 150-162. Available at `http://arxiv.org/abs/hep-th/9510206` (1995).

[15] E. Benhamou, O. Croissant. Local Time for the SABR Model: Connection with the 'Complex' Black Scholes and Applications to CMS and Spread Options. Working Paper. Available at `http://ssrn.com/abstract=1064461` (Oct. 2007).

[16] H. Berestycki, J. Busca, and I. Florent. Computing the Implied Volatility in Stochastic Volatility Models. *Communications on Pure and Applied Mathematics*, **57** Iss. 10, p. 1352-1373 (2004).

[17] H. Berestycki, J. Busca, and I. Florent. Asymptotics and Calibration of Local Volatilty Models. *Quantitative Finance*, **2** p. 61-69 (2002).

[18] F. Black, M. Scholes. The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy* **81** No. 3. p. 637-654 (1973).

[19] G. Bluman, S. Anco. Symmetry and Integration Methods for Differential Equations. Second Ed. *Springer* (2010).

[20] A. N. Borodin, P. Salminen. Handbook of Brownian Motions – Facts and Formulae. *Probability and Its Applications: Birkhauer* (1996).

[21] A. Brace, D. Gatarek, and M. Musiela. The Market Model of Interest Rate Dynamics. *Mathematical Finance* **7** No. 2, p. 127-147 (1997).

[22] D. R. Brecher, A.E. Lindsay. Results on the CEV Process, Past and Present. Working Paper. Available at `http://ssrn.com/abstract=1567864` (Mar. 2010).

[23] D. Brigo, F. Mercurio. Interest Rate Models – Theory and Practice: With Smile, Inflation and Credit. Second Edition. *Springer Finance* (2006).

[24] D. Brigo, F. Mercurio. A Deterministic-Shift Extension of Analytically-Tractable and Time-Homogeneous Short-Rate Models. *Finance and Stochastics* **5** No. 3, p. 369-387 (2001).

[25] A. M. Bronstein, M. M. Bronstein, R. Kimmel, Numerical Geometry of Non-Rigid Shapes. *Springer* (2008).

[26] L. A. Bordag, A. Mikaelyan, Models of Self-Financing Hedging Strategies in Illiquid Markets – Symmetry Reductions and Exact Solutions. Preprint. Available at `http://arxiv.org/abs/1008.2663` (Aug. 2010).

[27] P. Bourgade, O. Croissant. Heat Kernel Expansion for a Family of Stochastic Volatility Models: Delta-Geometry. Working Paper. Available at `http://ssrn.com/abstract=844229` (Nov. 2005).

[28] B. J. Cantwell, *Introduction to Symmetry Analysis*, First Edition. *Cambridge University Press* (2002).

[29] P. Carr, D. B. Madan. Option Valuation Using the Fast Fourier Transform. *Journal of Computational Finance* **2**, p. 61-73 (1999).

[30] J. Cheeger, D.G. Ebin. Comparison Theorems in Riemannian Geometry, *AMS Chelsea Publishing* (2008).

[31] W. Chen, N. Costanzino, J. Liechty, A. Mazzucato, and V. Nistor. Closed-Form Asymptotics and Numerical Approximations of 1D Parabolic Equations with Applications to Option Pricing. Working Paper (2010).

[32] A. S. Cherny, H. J. Engelbert. Singular Stochastic Differential Equations. *Springer*. (2005).

[33] B. Chow, P. Lu, and L. Ni. Hamilton's Ricci Flow. *AMS Graduate Studies in Mathematics 77* (2006).

[34] G. Cicogna, D. Vitali. Generalised Symmetries of Fokker-Plank-type Equaiton. J. Phys. A: Math. and Gen. **22** No. 11, p. L453-L456 (1989).

[35] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A Theory of the Term Structure of Interest Rates.*Econometrica* **53**, p.385-407 (1985).

[36] M. Craddock, K. A. Lennox. Lie Symmetry Methods for Multidimensional Linear, Parabolic PDEs and Diffusions. *Quantitative Finance Research Centre Research Paper Series*, Paper No. **274** (2010).

[37] M. Craddock, K. A. Lennox. Lie Group Symmetries as Integral Transforms of Fundamental Solutions. *Journal of Differential Equations* **232** p. 652-674 (2007).

[38] Y. Z. Chen, S. H. Lo, and J. R. Xie. Numerical Procedure for Evaluating Geodesics on a Surface. *Computer Aided Civil and Infrastructure Engineering* **4** Iss. 4, p. 317-320 (1989).

[39] F. Corielli, P. Foschi, and A. Paucucci. Parametrix Approximation of Diffusion Transition Densities. *SIAM J. Financial Math* **1** p. 833-867 (2010).

[40] M. Craddock, E. Platen. Symmetry Group Methods for Fundamental Solutions. *J. of Differential Equations* **207** p.285-302 (2004).

[41] A. O. Daalhuis. "Hypergeometric Function" in F. W. J. Olver, et.al. NIST Handbook of Mathematical Functions. *Cambridge University Press* (2010).

[42] F. Delbaen, W. Schachermayer. The Mathematics of Arbitrage. *Springer Finance* textbf13 (2008).

[43] C.G. Ding. Algorithm AS275: Computing the Non-Central Chi-Squared Distribution Function. *Applied Statistics* **41** p. 478-482 (1992).

[44] M. P. do Carmo. Differential Geometry of Surfaces. *Prentice-Hall* (1976).

[45] M. P. do Carmo. Riemannian Geometry. *Birkhauser* (1992).

[46] P. Doust. No Arbitrage SABR. Working Paper (2010).

[47] D. J. Duffy. Finite Difference Methods in Financial Engineering: A Partial Differential Equation Approach. *Wiley* (2006).

[48] B. Dupire. Pricing with a Smile. *Risk* **7** No. 1, p. 18-20 (1994).

[49] B. Dupire. Pricing and Hedging with Smiles. *Proceedings of AFFI Conference* (June 1993).

[50] B. Dupire. A Unified Theory of Volatility. Paribas Working Paper (1996).

[51] B. Dupire. Arbitrage Pricing with Stochastic Volatility. SORT Research Paper, Paribas Capital Markets (1993).

[52] E. Ekstrom, J. Tysk. Boundary Behaviour of Densities For Non-Negative Diffusions. Working Paper (2011).

[53] L. C. Evans. An Introduction to Stochastic Differential Equations Version 1.2. Available at `http://math.berkeley.edu/~evans/SDE.course.pdf` (2010).

[54] L. C. Evans. Partial Differential Equations Second Edition. *AMS Graduate Studies in Mathematics* **19** (2010).

[55] T. Frankel. The Geometry of Physics: An Introduction. *Cambridge University Press* Second Edition (2003).

[56] W. Feller. Two Singular Diffusion Problems. *Annals of Math, Second Series* **54**, No. 1 p.173-182 (1951).

[57] M. Forde, A. Jacquier. Small-Time Asymptotics for Implied Volatility Under the Heston Model. *International Journal of Theoretical and Applied Finance* **12** Iss. 6, p. 861-876 (2009).

[58] M. Forde. Exact Pricing and Large-Time Asymptotics for the Modified SABR Model and the Brownian Exponential Functional. Forthcoming in *International Journal of Theoretical and Applied Finance* (2010).

[59] M. Forde. The large maturity smile for the SABR model with non-zero correlation. Preprint. (2010).

[60] M. Forde. Small-Time Asymptotics for a General Local-Stochastic Volatility Model Using the Heat Kernel Expansion. Preprint (2010).

[61] J. P. Fouque, G. Papanicolaou, K. R. Sircar. Derivatives in Financial Markets with Stochastic Volatility. *Cambridge University Press* (2000).

[62] M. Francesco, A. Pascucci. On the Complete Model with Stochastic Volatility by Hobson and Rogers. *Proc. of the Royal. Soc.* **460** No. 2051 (2004).

[63] J. Gatheral. The Volatility Surface: A Practitioner's Guide. *Wiley* (2006).

[64] J. Gatheral, E.P. Hsu, P.M. Laurence, C. Ouyang, T. H. Wang. Asymptotics of Implied Volatility in Local Volatiltiy Models. To appear in *Mathematical Finance* (2010).

[65] P. Gauthier. P.H.Y. Rivaille, Fitting the Smile, Smart Parameters for SABR and Heston. Working Paper. Available at `http://ssrn.com/abstract=1496982` (2009)

[66] R.K. Gazizov, N.H. Ibragimov. Lie Symmetry Analysis of Differential Equations in Finance. *Nonlinear Dynamics* **17**, p. 387-407 (1998).

[67] H. Geman, M. Yor. Bessel processes, Asian option, and Perpetuitites. *Mathematical Finance* **3** p. 349-375 (1993).

[68] H. Geman, A. Eydeland. Domino Effect. *Risk* **8** p. 65-67 (1995).

[69] P. B. Gilkey, K. Kirsten, J. H. Park, and D. Vassilevich. Asymptotics of the Heat Equation with 'Exotic' Boundary Conditions or with Time Dependent Coefficients. *Nucl. Phys. Proc. Suppl.* **104** p.63-70. Preprint Available at `http://arxiv.org/abs/math-ph/0105009` (2002).

[70] P. B. Gilkey, K. Kirsten, and J. H. Park. Heat Content Asymptotics for Spectral Boundary Conditoins. *Trends Math.* **5** p.49-58. Preprint Available at `http://arxiv.org/abs/math-ph/0209012` (2002).

[71] P. Glasserman. Monte Carlo Methods in Financial Engineering. *Springer* (2004).

[72] J. Goard. Lie Symmetry Methods in Finance – An Example of the Bond Procing Equation. *Proceedings of the World Congress on Engineering 2008 Vol II* WCE (2008).

[73] J. Goard. Fundamental Solutions to Kolmogorov Equations via Reduction to Canonical Form. *Jour. App. Math. and Decision Sci.*, p. 1-24 (2006).

[74] W. M. Goldman. Complex Hyperbolic Geometry. *Clarendon Press Oxford Mathematical Monographs* (1999).

[75] S. Gounden, J.G. O'Hara. An Analytic Formula for the Price of an American-style Asian Option of Floating Strike Type. *Applied Mathematics and Computation* **217** Iss. 7, p.2923-2936. Preprint Available at `http://ssrn.com/abstract=1403746`(2010).

[76] A. V. Haastrecht, A. Pelsser. Efficient, Almost Exact Simulation of the Heston Stochastic Volatility Model. Accepted in *International Journal of Theoretical and Applied Finance.* Preprint Available at `http://ssrn.com/abstract=1621544` (2010).

[77] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward. Managing smile risk. *Wilmott Magazine* Sept. p. 84-108 (2002).

[78] P.Hagan, A. Lesniewski. LIBOR Market Model with SABR Style Stochastic Volatility. Working Paper available at `http://lesniewski.us/papers/working/SABRLMM.pdf` (2008).

[79] P. Hagan, A. Lesniewski, D. Woodward, Probability Distribution in the SABR Model of Stochastic Volatility. Working Paper available at `http://lesniewski.us/papers/working/ProbDistrForSABR.pdf` (2005).

[80] P. S. Hagan, D. E. Woodward. Equivalent Black Volatilitites. *Applied Mathematical Finance* **6** p. 147-157 (1999).

[81] R. Hamilton. Ricci Flow for Surfaces in Collected Papers on Ricci Flow. *International Press* (2003).

[82] S. Hassani. Mathematical Physics. *Springer* (1999).

[83] S. L. Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies* **6** No. 2 p. 327-343 (1993).

[84] E. P. Hsu. Stochastic Analysis on Manifolds. *AMS Graduate Studies in Mathematics* **38** (2002).

[85] J. E. Ingersoll Jr. Theory of Financial Decision Making. First Edition. *Rowman and Littlefield Publishers, Inc.* (1987).

[86] O. Islah. Solving SABR in Exact Form and Unifying it with LIBOR Market Model. Working Paper available at `http://ssrn.com/abstract=1489428` (Oct. 2009).

[87] N. M. Ivanova, C. Sophocleous, P. G. L. Leach. Group Classification of a Class of Equations Arising in Financial Mathematics. *Journal of Mathematical Analysis and Applications* **372** Iss. 1, p. 273-286 (2010).

[88] J. D. Jackson. Classical Electrodynamics. *Wiley* (1999).

[89] F. John. Partial Differential Equations. Forth Edition. *Springer* (1982).

[90] F. Jones. Lebesgue Integration on Euclidean Space. *Bartlet Publishers Inc.* (2001).

[91] M. Joshi. The Concepts and Practice of Mathematical Finance. *Cambridge University Press* (2007).

[92] J. Jost. Riemannian Geometry and Geometric Analysis. Fifth Edition. *Springer* (2008).

[93] L. Josefsson, P. Persson, Conformal Array Antenna Theory and Design. *Wilet* (2006).

[94] M. Kac. On Distributions of Certain Wiener Functionals. *Transactions of the American Mathematical Society* **65** p. 113 (1949).

[95] C. Kahl, P. Jackel. Not-so-Complex Logarithms in the Heston Model. *Wilmott Magazine*, p. 74-103.

[96] P. E. Kloeden, E. Platen. Numerical Solution of Stochastic Differential Equations. *Springer* (1999).

[97] D. Kristensen. A. Mele. Adding and Subtracting Black-Scholes: New Approach to Approximating Derivative Prices in Continuous – Time Models. Working Paper (2011).

[98] P. H. Labordere. Analysis, Geometry, and Modeling in Finance: Advanced Methods in Option Pricing. *Chapman and Hall/CRC Financial Mathematics Series* (2008).

[99] P. H. Labordere, A General Asymptotic Implied Volatility for Stochastic Volatility Models. Working Paper available at `http://arxiv:cond-mat/0504317` (2005).

[100] P. H. Labordere. Unifying the BGM and SABR Models: a Short Ride in Hyperbolic Geometry. Working Paper available at `http://ssrn.com/abstract=877762` (Jan 2007).

[101] P.H. Labordere. Calibration of Local Stochastic Volatility Models to Market Smiles. *Risk* **9** p. 112-117 (2009).

[102] P. Laurence, T. H. Wang, Closed Form Solutions for Quadratic and Inverse Quadratic Term Structure Models. *International Journal of Theoretical and Applied Finance* **8** No. 8, p. 1059-1083 (2005).

[103] P. Laurence. Implied Volatility, Fundamental Solutions, Asymptotic Analysis, and Symmetry Methods. Talk at Cal Tech April (2008).

[104] R. W. Lee. The Moment Formula for Implied Volatility at Extreme Strikes. *Mathematical Finance* **14** No. 3, p. 469-480. Preprint available at `http://ssrn.com/abstract=558885`. (2004).

[105] R. Lee. Implied and Local Volatilities Under Stochastic Volatility. *International Journal of Theoretical and Applied Finance*, **4** No. 1, p. 45-89 (2001).

[106] A. Lesniewski. WKB Method for Swaption Smile `http://lesniewski.us/papers/presentations/Courant020702.pdf` (2002).

[107] A. Lewis. Option Valuation under Stochastic Volatility with Mathematica Code. *Finance Press* (2000).

[108] A. Lewis. Geometries and Smile Asymptotics for a Class of Stochastic Volatility Models. Talk available at `www.optioncity.net/pubs/UCSB2007Talk.pdf` (2007).

[109] Lie, S. On the integration of a Class of Linear Partial Differential Equations by Definite Integrals. *Archiv for Mathematik og Naturvidenskab VI* No. 3, 1881, 328368. Reprinted in *Gesammelte Abhandlundgen* **3**, paper XXXV, CRC Handbook of Lie Group Analysis of Differential Equations **2**. *CRC Press* (1995).

[110] V. Linetsky. Spectral Expansions for Asian (Average Price) Options. *Operations Research* **52** No. 6, p. 856 - 867 (2004).

[111] Y. Liu, D. S. Wang. Symmetry Analysis of the Option Pricing Model with Dividend Yield from Financial Markets. Accepted in *Applied Mathematics Letters* (2010).

[112] M. Mateljevic, I. Anic, and S. Taylor. Asymptotic Curvature Bounds for Conformally Flat Metrics on the Plane. *Filomat* **24** No. 2 p. 93-100 (2010).

[113] D. M. McAvity, H. Osborn. A DeWitt Expansion of the Heat Kernel for Manifolds with a Boundary, *Class. and Quantum Grav.* **8** p. 603-638 (1991).

[114] A. N. Medvedev. Asymptotic Methods for Computing Implied Volatilities Under Stochastic Volatility. Working Paper available at `http://ssrn.com/abstract=667281` (2004).

[115] M. Y. Melnikov, Y. A. Melnikov. Construction of Green's Functions for the Black-Scholes Equation. *Electronic Journal of Differential Equations*. No. 153, p. 1-14 (2007).

[116] F. Memoli, G. Sapiro. Fast computation of weighted distance functions and geodesics on implicit hyper-surfaces. *J. Comput. Phys.* **173** No. 2, p. 730-764 (2001).

[117] F. Mercurio, M. Morini. Joining the SABR and Libor Market Models Together. *Risk* p.80-85 (Mar. 2009).

[118] F. Mercurio, A. Pallavicini. Smiling at Convexity: Bridging Swaption Skews and CMS Adjustments. **Risk** p. 64-69 (Aug. 2006).

[119] R. Merton. Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science (The RAND Corporation)* **4** No. 1 p. 141-183 (1973).

[120] R. C. Merton. Continuous Time Finance. *Blackwell Publishing* (1992).

[121] J. W. Milnor, J. Stasheff. *Characteristic Classes. Princeton University Press* (1974).

[122] J. W. Milnor. Morse Theory. *Annals of Mathematics Studies* **51** (1963).

[123] S. Molchanov. Diffusion processes and Riemmanian Geometry. *Russian Math. Surveys* **30** No. 1, p. 1-63 (1975).

[124] L. Monti, A. Pascucci. Obstacle Problem for Arithmetic Asian Options. *C. R. Acad. Sci. Paris, Ser. I 347* p.1443-1446 (2009).

[125] M. Morini, F. Mercurio. No-Arbitrage Dynamics for a Tractable SABR term Structure Libor Market Model. *Bloomberg Portfolio Research Paper No. 2010-03-FRONTIERS* (2007).

[126] M. Nakahara. Geometry, Topology, and Physics. *Taylor and Francis Group* (2003).

[127] V. Naicker, K. Andriopoulos, PGL Leach. Symmetry Reductions of a Hamilton-Jacobi-Bellman Equation Arising in Financial Mathematics. *Journal of Nonlinear Mathematical Physics* **12** No. 2, p. 268-283 (2005).

[128] V. Naicker, J. G. O' Hara, and P. G. L. Leach, Symmetries and the Merton Portfolio Selection Model. *CCFEA Technical Report* WP041-10. Working paper available at `http://www.essex.ac.uk/ccfea/research/WorkingPapers/2010/CCFEA\%20WP041-10.pdf` (2010).

[129] J. Obloj. Fine-Tune Your Smile: Correction to Hagan et al. Working Paper available at `http://www.arxiv.org/abs/0708.0998` (2008).

[130] B. Øksendal. Stochastic Differential Equations: An Introduction with Applications. Sixth Edition. *Springer* (2008).

[131] P. Olver. *Applications of Lie Groups to Differential Equations*. Second Edition. *Springer*, New York (2000).

[132] F. W. Olver. Asymptotics and Special Functions. Second Edition. *CRC Press*, (1997).

[133] Y. Osajima. The Asymptotic Expansion Formula of Implied Volatility for Dynamic SABR Model and FX Hybrid Model. Working paper available at `http://ssrn.com/abstract=965265` (2007).

[134] L. Paulot. Asymptotic Implied Volatility at the Second Order with Application to the SABR Model. Working paper available at `http://ssrn.com/abstract=1413649` (2007).

[135] P. Petersen. Riemannian Geometry. Second Edition. *Springer* (2006).

[136] S. Polidoro, C. Mogavero. A Finite Difference Method For a Boundary Value Problem Related to the Kolmogorov Equation. Working Paper (2011).

[137] P. E. Protter. Stochastic Integration and Differential Equations. Second Edition. *Springer* (2005).

[138] R. Rebonato, K. McKay, R. White. The SABR/LIBOR Market Model: Pricing, Calibration and Hedging for Complex Interest-Rate Derivatives. *Wiley* (2009).

[139] R. Rebonato, J.P. Barjaktarevic. Approximate Solutions for the SABR Model: Improving on the Hagan Expansion. ICBI presentation (2010).

[140] L. Rogers, Z. Shi. The Value of an Asian Option. *Journal of Applied Probability* **32** No. 4 p. 1077-1088 (1995).

[141] Rogers, L.C.G., Zane, O. Saddlepoint Approximations to Options Prices. *Annals of Applied Probability* **9**, 493-503 (1999).

[142] W. Rudin. Real and Complex Analysis. *McGraw-Hill* (1987).

[143] N. A. Sharp. Comparison of Numerial Methods for the Integration of the Black Hole Geodesic Equations. *J. of Computational Physics.* **41** p. 295-308 (1981).

[144] S. E. Shreve. Stochastic Calculus for Finance II: Continuous-Time Models (Vol 2). *Springer Finance* (2004).

[145] G. Silberberg. Discrete Symmetries of the Black-Scholes Equation. *Proceedings of 10th International Conference in MOdern GRoup ANalysis* p. 190-197 (2005).

[146] J.P. Singh, S. Prabakaran. Group Properties of the Black Scholes Equation and its Solutions. *Electronic Journal of Theoretical Physics* **5**, No. 18 p. 51-60 (2008).

[147] W. Sinkala, P.G. Leach, and J.G. O'Hara. Zero-Coupon Bond Prices in the Vasicek and CIR Models: Their Computation as Group-Invariant Solutions. *Mathematical Methods in the Applied Sciences* **31** p. 665-678 (2008).

[148] S. Spichak, V. Stognii. Symmetry Classification and Exact Solutions of the One-Dimensional Fokker-Plank Equation with Arbitrary Coefficients of Drift and Diffusion. *J. Phys. A: Math. Gen* **32** p. 8341-8353 (1999).

[149] N. E. Steenrod. The Topology of Fibre Bundles. *Princeton University Press* (1999).

[150] H. Stephani, M. MacCallum. Differential Equations: Their Solution Using Symmetries. *Cambridge University Press* (1990).

[151] N. Sukhomlin, J. M. Ortiz. Equivalence and New Exact Solutions to the Black Scholes and Diffusion Equations. *Applied Mathematics E-Notes* **7** p.206-213 (2007).

[152] A. Takahashi, K. Takehara, M. Toda. Computation in an Asymptotic Expansion Method. *CARF Working Paper Series CARF-F-149*. Working paper available at `http://ssrn.com/abstract=1413924` (May 2009)

[153] S. Taylor, S. Glasgow. A Novel Reduction of the Simple Asian Option and Lie-Group Invariant Solutions. *International Journal of Theoretical and Applied Finance* **12** Iss. 8 p. 1197-1212 (2009).

[154] S. Glasgow, S. Taylor, and J. Vecer. Symmetries of Generalized Asian Equations. In preparation (2010).

[155] Tehranchi, M. Asymptotics of Implied Volatility Far From Maturity. *Journal of Applied Probability*, **46** (3), 629-650 (2009).

[156] G. W. P. Thompson. Fast Narrow Bounds on the Value of Asian Options. Working paper available at `http://www.jbs.cam.ac.uk/research/working\_papers/2002/wp0209.pdf` (2002).

[157] S. R. S. Varadhan. On the behavior of the Fundamental Solution of the Heat Equation with Variable Coefficients. *Comm. Pure. Appl. Math.* **20** Iss. 2, p. 431-455 (1967).

[158] S. R. S. Varadhan. Diffusion Processes in a Small Time Interval. *Comm. Pure. Appl. Math.* **20** p. 659-685 (1967).

[159] D. V. Vassilevich. Heat Kernel Expansion: User's Manual. *Phys. Rep.* **388** p. 279-360. Preprint available at `http://arxiv.org/abs/hep-th/0306138` (2003).

[160] J. Vecer. Stochastic Finance: A Numeraire Approach. *CRC Press* (2010).

[161] J. Večeř. A new PDE approach for pricing arithmetic average Asian Options. *Journal of Computational Finance* **4** No. 4, p. 105-113 (2001).

[162] J. Večeř. Unified Asian Pricing. *Risk* **15** No. 6, p. 113-116 (2002).

[163] J. Večeř, M. Xu. Pricing Asian Options in a Semimartingale Model. *Quantitative Finance* **4** No. 2, p. 170-175 (2001).

[164] B. H. Voorhees. Solving Geodesic Equations in Space of Locally Constant Connections in Mathematical Essays on Growth and the Emergence of Form edited by P. L. Antonelli. *University of Alberta Press* (1985).

[165] R. M. Wald. General Relativity. *The University of Chicago Press* (1984).

[166] Sheng-Li Wang. Symmetry Group Classification of a Stochastic Volatility Model in Mathematical Finance. Master's Thesis (2006).

[167] F. W. Warner. Foundations of Differentiable Manifolds and Lie Groups. *Springer* (1983).

[168] G. West. Calibration of the SABR Model in Illiquid Markets. Working Paper available at `http://www.riskworx.com/insights/sabr/sabrilliquid.pdf` (2005).

[169] Q. Wu. Series Expansion of the SABR Joint Density. Working Paper available at `http://ssrn.com/abstract=1367966` (Jan 2010).

[170] L. Ying, E. Candes. Fast Geodesics Computation with the Phase Flow Method. *J. Comput. Phys.* **220** p. 6-18 (2006).