

SERIES EXPANSION OF THE SABR JOINT DENSITY

QI WU

Columbia University

Under the SABR stochastic volatility model, pricing and hedging contracts that are sensitive to forward smile risk (e.g., forward starting options, barrier options) require the joint transition density. In this paper, we address this problem by providing closed-form representations, asymptotically, of the joint transition density. Specifically, we construct an expansion of the joint density through a hierarchy of parabolic equations after applying total volatility-of-volatility scaling and a near-Gaussian coordinate transformation. We then establish an existence result to characterize the truncation error and provide explicit joint density formulas for the first three orders. Our approach inherits the same spirit of a small total volatility-of-volatility assumption as in the original SABR analysis. Our results for the joint transition density serve as a basis for managing forward smile risk. Through numerical experiments, we illustrate the accuracy of our expansion in terms of joint density, marginal density, probability mass, and implied volatilities for European call options.

KEY WORDS: forward smile risk, SABR stochastic volatility model, joint transition density.

1. INTRODUCTION

In the derivative markets for interest rates and currencies, smile risk is a key component that one needs to manage. The SABR stochastic volatility model introduced by Hagan et al. (2002) is a widely used model by practitioners in these institutional markets as the model not only fits the implied volatility smile well, but also generates correct comovements between the underlying level and its smile curve. The implied volatility result for European call options in the original work of Hagan et al. (2002) is adequate for managing single-period smile exposures and has become one of the most widely used formulas in the fixed income market due to its accuracy, simplicity, and clear interpretations of model parameters.

However, as yesterday's exotic products become today's vanilla flow, markets expect simple formulas to manage forward smile risks reflected in the liquidly traded products where the payoff functions involve the underlying states at multiple future temporal points, not just at the maturity. Conditional forward smile exposures, from one future state to another future state, arise from this multiperiodicity feature in the payoff's structure. For instance in the interest rate market, the strike of a forward starting call option on 3 month LIBOR rate with 1 year maturity is set as the future level of the

I am very grateful to a referee for pointing out an error in the calculation in an earlier version of this paper, and I thank both referees for their helpful comments. I would also like to thank Paul Glasserman for introducing me to this problem and for his guidance in carrying out the work. I also thank David Keyes and Matias Courdurier for their comments on an earlier version of this paper. This work is partially supported by NSF grants DMS0410234 and DMS0914529.

Manuscript received May 2009; final revision received February 2010.

Address correspondence to Qi Wu, Department of APAM, Columbia University, New York, NY 10027; e-mail: qw2107@columbia.edu.

DOI: 10.1111/j.1467-9965.2010.00460.x

© 2010 Wiley Periodicals, Inc.

3 month LIBOR rate at a certain future time between now and the option maturity. In the currency market, a barrier option on an exchange rate has payoff functions depending on the underlying states at a few intermediate or a whole continuum of intermediate temporal points throughout the life of the contract. Holding positions of such liquid products with forward starting or barrier features exposes one to forward smile risks, which depends on the transition density from one future state to another future state. Specifically, if a stochastic volatility model is used, the joint transition density of both the underlying and the volatility are necessary. Marginal transition densities are not enough to price and hedge such products.

Unfortunately, the following three properties of the SABR model prevent one from obtaining a strictly closed form solution to the evolution equation that the joint transition density function satisfies. Namely, the nonlinearity from the constant elasticity variance (CEV)-type local volatility function, the coupling between the underlying security process and the lognormal volatility process, and finally the correlation between the two driving Brownian motions. When a diffusion becomes multidimensional and furthermore with the presence of correlation and coupling effect, only a few results are available for the nonaffine class. We therefore turn to asymptotic methods to approximate the joint transition density.

To put our approach in context, we first briefly discuss earlier approaches to the SABR model: the singular perturbation, heat kernel asymptotics, and Malliavin calculus. The singular perturbation method was first applied in Hagan et al. (2002) where their keen observation of the total volatility-of-volatility as a scaling parameter led to a remarkably accurate implied volatility formula for European call options with clear interpretations of how each of the model parameters in the formula affects the shape of a smile curve. Other successful applications of perturbation techniques applied to multiscale stochastic volatility models include Fouque, Papanicolou, and Sircar (2000) and Fouque et al. (2003, 2004). Further, an important existence and uniqueness result for implied volatility function was established in Berestycki, Busca, and Florent (2004) for a general class of two factor level-dependent stochastic volatility models including the SABR and Heston model. A differential geometry approach based on heat kernel asymptotics on Riemannian manifolds was developed in Henry-Labordère (2005) targeting a wider range of stochastic volatility models, including the Heston model and SABR model with mean-reversion in the volatility process, the λ -SABR model. In particular, the λ -SABR model was shown to correspond to the hyperbolic Poincaré half-plane whose geodesic distance is known and a formula for the first-order asymptotic smile was explicitly calculated. Hagan, Lesniewski, and Woodward (2005) used the geometric approach and analyzed specifically the original SABR model and obtained the marginal transition density under various boundary conditions. Bourgade and Croissant (2005) applied the heat kernel asymptotics approach for the same family of generalized SABR models as in Henry-Labordère (2005). In a third approach, Osajima (2007) used the Watanabe–Yoshida theory of the Malliavin calculus and proved the implied volatility formula developed in Hagan et al. (2002) for the dynamic version of the SABR model. In Osajima (2007), a second-order asymptotic expansion in terms of the same total volatility-of-volatility parameter as Hagan et al. (2002) is characterized for both European call prices and implied volatilities.

Given the historical development on the subject, the aim of this paper is to look for a systematic approach of calculating explicitly the joint transition density up to arbitrarily high-order expansions with particular focus on two aspects: the tractability of calculations and the characterization of finite-order truncation error. Our method comprises the following steps.

First, we start with the same assumption as Hagan et al. (2002) in that the total volatility-of-volatility for the entire contract horizon is considered as a small quantity so that it can be treated as a perturbation parameter to scale the problem.

Next we apply a coordinate transformation that greatly simplifies our expansion. After this transformation, in the limit of the total volatility-of-volatility scaling, the SABR generator converges to the generator of a correlated two-dimensional Brownian motion which admits a bivariate Gaussian transition density. Without scaling, it is not possible to completely standardize the SABR model due to presence of coupling effect in the dynamics.

Then with this scaling limit result, we seek a perturbation expansion in the solution of the transformed PDE and obtain a parabolic hierarchy at arbitrary order of scales. By construction, the leading order solution is a bivariate Gaussian distribution and all other higher order expansions are explicitly obtainable through convolutions of the leading order solution against its spatial derivatives. The tractability stems from the convolution structure of Gaussian functions against their moments. And the ultimate form of a finite order expansion at arbitrary order will be of a product of the leading order bivariate Gaussian distribution multiplied by polynomial functions of state variables.

Finally, we show by Duhamel's principle and Young's inequality that the expansion series forms a global L^1 solution with finite-order truncation error of order ε^{n+1} where ε is the total volatility-of-volatility scaling parameter and n is the highest order to which the expansion is carried out.

Relating to the earlier work on the subject, our methodology is closest to the singular perturbation approach in Hagan et al. (2002). In terms of results obtained, Hagan, Lesniewski, and Woodward (2005) derived an explicit marginal density approximation with free boundary condition and discussed in detail its relationship with solutions for Dirichlet (or absorbing), Neumann (or reflecting), and Robin (or mixed) boundary conditions at zero forward, as well as their impact on the implied volatilities at small and large strikes. On the other hand, the heat kernel framework of Henry-Labordère (2005) is set out to be more general for a wider class of stochastic volatility models. Given the historical development in Hagan, Lesniewski, and Woodward (2005) and Henry-Labordère (2005) and our motivation of pricing products with forward starting and barrier features, the scope of this paper is specifically on the joint transition density with free boundary condition. In particular, results developed in this paper are more explicit and easier in terms of calculating higher order terms to refine the approximation. Further, our results are supplemented by an error-bound proof that guarantees the convergence of the proposed expansion. Other work involving the SABR model includes extending interest rate market models to include SABR-consistent smile features (Hagan and Lesniewski 2006; Henry-Labordère 2007; Morini and Mercurio 2007; Rebonato 2007), its moment properties (Andersen and Piterbarg 2007; Lions and Musiela 2005), local time for the SABR model (Benhamou and Croissan 2007), and alternatives to the SABR dynamics (Rogers and Veraart 2008). See Gatheral (2006) for an overview of volatility surface modeling.

The rest of the paper is structured as follows. In Section 2, we state the SABR model and its associated Kolmogorov equations. In Section 3, we introduce the total volatility-of-volatility scaling and the near Gaussian transformation and show that the transformed density function converges to a bivariate Gaussian under the scaling limit. In Section 4, we seek a solution expansion around the limiting distribution and derive the corresponding equation hierarchy. We then show that the expansion converges globally in L^1 and further characterize the truncation error in terms of the order of the scaling parameter. In

Section 5, we provide explicit formulas of the leading, first- and second-order expansion. We conclude the paper in Section 6 with numerical examples to illustrate the accuracy of the results.

2. THE SABR MODEL

We review the definition of the SABR dynamics and the evolution of its joint transition density.

2.1. Transition Probability

The SABR model specifies the joint risk neutral dynamics of a security’s forward price and a volatility process on a measurable space (Ω, \mathcal{F}) equipped with the T -forward martingale measure \mathbb{Q}^T (Musielà and Rutkowski 1997) and a filtration $\{\mathcal{F}_t, t \geq 0\}$ generated by the σ -algebras of the two T -forward Brownian motions $\mathcal{F}_t \triangleq \sigma((W_s^1, W_s^2), 0 \leq s \leq t)$.

On the probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{Q}^T)$, the SABR dynamics reads: $\forall t \geq 0$

$$(2.1) \quad \begin{cases} d\hat{F}_t = \hat{\alpha}_t \hat{F}_t^\beta dW_t^1, & \hat{F}_t > 0, \beta \in [0, 1] \\ d\hat{\alpha}_t = \nu \hat{\alpha}_t dW_t^2, & \hat{\alpha}_t > 0, \nu > 0 \\ \mathbb{E}^{\mathbb{Q}^T}[dW_t^1 dW_t^2] = \rho dt, & \rho \in (-1, 1), \end{cases}$$

where \hat{F}_t is the T -forward price of the considered security with risk-free zero coupon bond as the numéraire. The volatility function is of the form $\hat{\alpha}_t \hat{F}_t^\beta$ with \hat{F}_t^β corresponding to the CEV component and $\hat{\alpha}_t$ following a lognormal process. (W_t^1, W_t^2) are correlated Brownian motions under the measure \mathbb{Q}^T .

To study the SABR dynamics, we first construct its infinitesimal generator and then solve for the joint transition probability density from its associated Kolmogorov equation. To fix notation, we take horizon T as the contract exercise date and call t the current time such that $0 < t < T$. We call f and α backward variables which denote the state values of \hat{F}_t and $\hat{\alpha}_t$. Accordingly, T is the future time and F and A are forward variables denoting the state values of \hat{F}_T and $\hat{\alpha}_T$. The forward Kolmogorov equation (FKE) evolves with respect to f, α and the backward Kolmogorov equation (BKE) with respect to F, A .

Assuming the existence of a transition density function for the conditional probability measure to (2.1), we have them relating to each other in the following sense:

$$\mathbb{P}(F < \hat{F}_T \leq F + dF, A < \hat{\alpha}_T \leq A + dA \mid \hat{F}_t = f, \hat{\alpha}_t = \alpha) = p(t, f, \alpha; T, F, A) dF dA$$

and $p(t, f, \alpha; T, F, A)$ denotes the associated joint transition density. The SABR dynamics are time homogeneous, so the joint transition density depends on T and t only through their difference which we denote by $s := T - t$. We thereafter write the joint transition density as $p(s, f, \alpha; F, A)$.

2.2. Kolmogorov Equation

We rewrite the SABR dynamics from (2.1) in vector form as

$$\begin{pmatrix} d\hat{F}_t \\ d\hat{\alpha}_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} dt + \begin{pmatrix} \hat{\alpha}_t \hat{F}_t^\beta & 0 \\ 0 & \nu \hat{\alpha}_t \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix}$$

with its drift vector, diffusion matrix, and correlation matrix between two Brownian motions denoted by Λ , Σ , and Ω

$$\Lambda(\hat{F}_t, \hat{\alpha}_t) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma(\hat{F}_t, \hat{\alpha}_t) = \begin{pmatrix} \hat{\alpha}_t \hat{F}_t^\beta & 0 \\ 0 & v \hat{\alpha}_t \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

For a general multidimensional diffusion process driven by correlated Brownian motions, the infinitesimal generator \mathcal{A} and its adjoint operator \mathcal{A}^* relate to each other through an inner product such that for any function pair $f, g \in C^2(\mathbb{R}^2)$ that vanishes at $\pm\infty$, we have $\langle \mathcal{A}f, g \rangle = \langle f, \mathcal{A}^*g \rangle$. In our case when \mathcal{A} and \mathcal{A}^* act on the joint transition density $p(t, f, \alpha; T, F, A)$, they take the form

$$\begin{aligned} [\mathcal{A}p(t, f, \alpha; \cdot)](f, \alpha) &= \Lambda(f, \alpha) \cdot \nabla p(t, f, \alpha; \cdot) \\ &\quad + \frac{1}{2} \text{Tr}[\Sigma(f, \alpha) \Omega \Sigma^T(f, \alpha) \times \mathbf{H}p(t, f, \alpha; \cdot)] \\ &= \frac{1}{2} \left[a(f, \alpha) \frac{\partial^2 p}{\partial f^2} + 2b(f, \alpha) \frac{\partial^2 p}{\partial f \partial \alpha} + c(f, \alpha) \frac{\partial^2 p}{\partial \alpha^2} \right] \\ [\mathcal{A}^*p(\cdot; T, F, A)](F, A) &= -\nabla \cdot p(\cdot; T, F, A) \Lambda(F, A) \\ &\quad + \frac{1}{2} \text{Tr}[\mathbf{H} \times p(\cdot; T, F, A) \Sigma(F, A) \Omega \Sigma^T(F, A)] \\ &= \frac{1}{2} \left[\frac{\partial^2 a(F, A)p}{\partial F^2} + 2 \frac{\partial^2 b(F, A)p}{\partial F \partial A} + \frac{\partial^2 c(F, A)p}{\partial A^2} \right] \end{aligned}$$

with $\nabla, \mathbf{H}, \text{Tr}$ denoting the gradient vector, Hessian matrix and trace operator for a symmetric matrix, and \cdot, \times denoting operator products for vectors and matrices. Further the diffusion coefficients $a(f, \alpha), b(f, \alpha), c(f, \alpha)$ and $a(F, A), b(F, A), c(F, A)$ are calculated as

$$\begin{aligned} a(f, \alpha) &= f^{2\beta} \alpha^2, & b(f, \alpha) &= \rho v f^\beta \alpha^2, & c(f, \alpha) &= v^2 \alpha^2 \\ a(F, A) &= F^{2\beta} A^2, & b(F, A) &= \rho v F^\beta A^2, & c(F, A) &= v^2 A^2. \end{aligned}$$

Associated with \mathcal{A}^* and \mathcal{A} , respectively, the FKE and BKE pair for the joint transition density $p(t, f, \alpha; T, F, A)$ then takes the form

$$\begin{aligned} \left[\frac{\partial}{\partial T} - \mathcal{A}^* \right] p(\cdot; T, F, A) &= 0, \quad T > t, \quad \text{starting at } \lim_{T \rightarrow t} p(\cdot; T, F, A) = \delta(F - f) \delta(A - \alpha) \\ \left[\frac{\partial}{\partial t} + \mathcal{A} \right] p(t, f, \alpha; \cdot) &= 0, \quad t < T, \quad \text{terminating at } \lim_{t \rightarrow T} p(t, f, \alpha; \cdot) = \delta(f - F) \delta(\alpha - A) \end{aligned}$$

and the equation we will analyze onwards for $p(t, f, \alpha; T, F, A)$ is with respect to the backward variables f, α where the forward variables F, A are treated as constants. With $s = T - t$, the BKE becomes

$$(2.2) \quad \begin{cases} \left[\frac{\partial}{\partial s} - \frac{1}{2} \left(a(f, \alpha) \frac{\partial^2}{\partial f^2} + 2b(f, \alpha) \frac{\partial^2}{\partial f \partial \alpha} + c(f, \alpha) \frac{\partial^2}{\partial \alpha^2} \right) \right] \\ p(s, f, \alpha; F, A) = 0, s \in (0, T] \\ p(s, f, \alpha; F, A) = \delta(f - F)\delta(\alpha - A), \quad s = 0. \end{cases}$$

When time to maturity $T - t$ shrinks to zero, backward variables (f, α) coincide with forward variables (F, A) . This is the terminal condition at $t = T$ before the change of variable $s = T - t$ and the initial condition at $s = 0$ for (2.2) and it is represented by a two-dimensional delta function, the point measure on a plane.

It is important to point out that although $p(s, f, \alpha; F, A)$ is a function of both f, α and F, A , it is a probability density only in the forward variables F, A with backward variables f, α fixed. The solution to the BKE as a function of f, α with F, A fixed is not in general a probability function. This is particularly true for the BKE associated with the SABR model as one can check that \mathcal{A} is not self-adjoint.

Equation (2.2) is a linear second-order partial differential equation of parabolic type in nondivergence form with coordinate-dependent coefficients. A set of sufficient conditions for the existence of a unique fundamental solution to (2.2) are boundedness, uniform ellipticity, and Hölder continuity in the diffusion coefficient functions $a(f, \alpha), b(f, \alpha), c(f, \alpha)$, see page 3 in Friedman (1983). Unfortunately, the coefficients of (2.2) are neither bounded nor uniformly elliptic, as is very often the case with PDEs arising from diffusion processes. Nor do they satisfy relaxations of these conditions as in Aronson and Besala (1967) and Chen (1986). To the best of our knowledge, equation (2.2) falls outside known regularity conditions for existence and uniqueness of a fundamental solution. We will therefore proceed on the assumption, standard in the mathematical finance literature, that the model admits a transition density satisfying the BKE. In subsequent sections, the equations we derive through scaling and transformation of (2.2) inherit the solvability properties of (2.2) without further assumptions.

3. SCALING AND TRANSFORMATION

We first introduce the total volatility-of-volatility scaling and the near-Gaussian transformation. We then show that due to the correlation and coupling effect, SABR diffusion is not able to be either strictly standardized nor completely decoupled. Instead, the bivariate random vector of the SABR process at any given time will only become a bivariate normal vector in the limit when the scaling parameter vanishes to zero.

3.1. Total Volatility-of-Volatility Scaling

The difficulty of applying typical solution construction methods to equation (2.2) stems from two aspects: the arbitrarily noninteger-valued CEV component β and complex-valued characteristic curves due to the presence of the correlation parameter ρ . The spatial part of equation (2.2) is always elliptic for the range of SABR parameters except at $\rho = \pm 1, [ac - b^2](f, \alpha) = \nu^2 \alpha^4 f^{2\beta} (1 - \rho^2) > 0, \forall f > 0, \alpha > 0, \nu > 0$. Thus the two characteristic ODEs for the parabolic operator in (2.2) have complex-valued solutions. Seeking a closed-form representation of the solution to (2.2) with real value operations faces limited options, and very likely the only option is a series representation.

The asymptotic expansion method is based on the fact that in a wide range of market scenarios, the total volatility-of-volatility v^2T is not very large and can be considered as a small quantity (Hagan et al. 2002). This complies with the spirit of the original SABR methods. It then can be treated as a perturbation parameter $\varepsilon := \sqrt{v^2T}$ assuming $\varepsilon \ll 1$. We take this as the starting point of our power series expansion analysis.

The scaling parameter ε corresponds to the following change of variables:

$$(3.1) \quad \begin{cases} \tau(s) := s/T \\ x(f, \alpha) := f \\ y(f, \alpha) := \alpha/v \end{cases}$$

and we denote $x(F, A), y(F, A)$ by X, Y which is $F, A/v$. Applying (3.1) to equation (2.2) through the chain rule

$$(3.2) \quad \begin{cases} \frac{\partial}{\partial s} = \frac{\partial}{\partial \tau} \frac{d\tau}{ds} = \frac{1}{T} \frac{\partial}{\partial \tau} \\ \frac{\partial^2}{\partial f^2} = \frac{\partial}{\partial x} \left(\frac{\partial}{\partial x} \frac{dx}{df} \right) \frac{dx}{df} = \frac{\partial^2}{\partial x^2} \\ \frac{\partial^2}{\partial f \partial \alpha} = \frac{\partial}{\partial x} \left(\frac{\partial}{\partial y} \frac{dy}{d\alpha} \right) \frac{dx}{df} = \frac{1}{v} \frac{\partial^2}{\partial x \partial y} \\ \frac{\partial^2}{\partial \alpha^2} = \frac{\partial}{\partial y} \left(\frac{\partial}{\partial y} \frac{dy}{d\alpha} \right) \frac{dy}{d\alpha} = \frac{1}{v^2} \frac{\partial^2}{\partial y^2} \end{cases}$$

then equation (2.2) becomes

$$(3.3) \quad \begin{cases} \left[\frac{\partial}{\partial \tau} - \frac{v^2 T}{2} \left(x^{2\beta} y^2 \frac{\partial^2}{\partial x^2} + 2\rho x^\beta y^2 \frac{\partial^2}{\partial x \partial y} + y^2 \frac{\partial^2}{\partial y^2} \right) \right] \\ p(\tau T, x, v y; X, v Y) = 0, \quad \tau \in (0, 1] \\ p(\tau T, x, v y; X, v Y) = \frac{1}{v} \delta(x - X) \delta(y - Y), \quad \tau = 0. \end{cases}$$

Define a new quantity $\tilde{p}_\varepsilon(\tau, x, y; X, Y)$ as

$$(3.4) \quad \tilde{p}_\varepsilon(\tau, x, y; X, Y) \triangleq v p(\tau T, x, v y; X, v Y)$$

and substitute (3.4) into (3.3) together with $\varepsilon^2 = v^2 T$, we have $\tilde{p}_\varepsilon(\tau, x, y; X, Y)$ satisfying

$$(3.5) \quad \begin{cases} \left[\frac{\partial}{\partial \tau} - \frac{\varepsilon^2}{2} \left(x^{2\beta} y^2 \frac{\partial^2}{\partial x^2} + 2\rho x^\beta y^2 \frac{\partial^2}{\partial x \partial y} + y^2 \frac{\partial^2}{\partial y^2} \right) \right] \\ \tilde{p}_\varepsilon(\tau, x, y; X, Y) = 0, \quad \tau \in (0, 1] \\ \tilde{p}_\varepsilon(\tau, x, y; X, Y) = \delta(x - X) \delta(y - Y), \quad \tau = 0. \end{cases}$$

The quantity $\tilde{p}_\varepsilon(\tau, x, y; X, Y)$ satisfying (3.5) is the transition density function of the following scaled process, denoted by $(\tilde{X}_\tau^\varepsilon, \tilde{Y}_\tau^\varepsilon)$

$$(3.6) \quad \begin{cases} d\hat{X}_t^\varepsilon = \varepsilon \hat{Y}_t^\varepsilon (\hat{X}_t^\varepsilon)^\beta d\tilde{W}_t^1, & \hat{X}_0^\varepsilon = \hat{F}_0 \\ d\hat{Y}_t^\varepsilon = \varepsilon \hat{Y}_t^\varepsilon d\tilde{W}_t^2, & \hat{Y}_0^\varepsilon = \hat{\alpha}_0/\nu \\ \mathbb{E}^{\mathbb{Q}^\varepsilon} [d\tilde{W}_t^1 d\tilde{W}_t^2] = \rho dt, & \rho \in (-1, 1), \quad t \in [0, 1] \end{cases}$$

with generator $\tilde{\mathcal{A}}^\varepsilon$

$$\begin{cases} \tilde{\mathcal{A}}^\varepsilon \triangleq \tilde{\mathcal{L}}_{x,y}^\varepsilon = \frac{1}{2} \left[a_\varepsilon(x, y) \frac{\partial^2}{\partial x^2} + 2b_\varepsilon(x, y) \frac{\partial^2}{\partial x \partial y} + c_\varepsilon(x, y) \frac{\partial^2}{\partial y^2} \right] \\ a_\varepsilon(x, y) = \varepsilon^2 x^{2\beta} y^2, \quad b_\varepsilon(x, y) = \varepsilon^2 \rho x^\beta y^2, \quad c_\varepsilon(x, y) = \varepsilon^2 y^2. \end{cases}$$

The family $\tilde{\mathcal{L}}_{x,y}^\varepsilon$ is of Laplacian type with correlation and coordinate-dependent coefficients. The subscript (x, y) indicates that it acts on the spatial coordinates (x, y) and the subscript ε indicates it is considered as an operator parameterized by ε . Accordingly, $\tilde{p}_\varepsilon(\tau, x, y; X, Y)$ defines a family of scaled transition densities each satisfying a scaled BKE

$$(3.7) \quad \begin{cases} \left[\frac{\partial}{\partial \tau} - \tilde{\mathcal{L}}_{x,y}^\varepsilon \right] \tilde{p}_\varepsilon(\tau, x, y; X, Y) = 0, & 0 < \tau \leq 1 \\ \tilde{p}_\varepsilon(\tau, x, y; X, Y) = \delta(x - X) \delta(y - Y), & \tau = 0. \end{cases}$$

3.2. Near-Gaussian Transformation

Having introduced the scaling parameter ε , we then define a point-wise coordinate transformation which we call the near-Gaussian transformation, and we will show that under this coordinate change, the resulting generator converges to that of a two-dimensional correlated Brownian motion as $\varepsilon \downarrow 0$. The corresponding transition density at this limit is a bivariate Gaussian function around which we will seek a power series expansion assuming the scale parameter ε is small.

The reason that we could not simply find a particular set of change of variables such that \mathcal{A} can be transformed into a standard Laplacian and why instead we need a scaling first is because neither a standardization transform nor a decoupling transform is possible for the original SABR process; the analysis is detailed in the Appendix. The best alternative is then our notion of near-Gaussian transformation.

THEOREM 3.1. *Let x, y, X, Y be the scaled coordinates in (3.1). Define the following point-wise transformation from coordinate (x, y) to a new coordinate (u, v) as:*

$$(3.8) \quad \begin{cases} u := \phi(x, y) \triangleq \int_X^x \frac{dx'}{\sqrt{a_\varepsilon(x', y)}} = \frac{x^{1-\beta} - X^{1-\beta}}{\varepsilon(1-\beta)y} \\ v := \psi(x, y) \triangleq \int_Y^y \frac{dy'}{\sqrt{c_\varepsilon(x, y')}} = \frac{\ln y - \ln Y}{\varepsilon} \end{cases}$$

and denote $(\phi(X, Y), \psi(X, Y))$ by (U, V) which becomes a constant vector $(0, 0)$. Let $\hat{p}_\varepsilon(\tau, u, v; U, V)$ denote the transition density in the new coordinates, shortened as $\hat{p}_\varepsilon(\tau, u, v)$. Then $\hat{p}_\varepsilon(\tau, u, v)$ converges a.e. for τ, u, v and in L_1 with respect to u, v for each τ to a bivariate Gaussian function $\hat{p}_0(\tau, u, v)$:

$$\hat{p}_\varepsilon(\tau, u, v) \rightarrow \hat{p}_0(\tau, u, v) = \frac{1}{2\pi\tau\sqrt{1-\rho^2}} \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2\tau(1-\rho^2)}\right) \quad \text{as } \varepsilon \downarrow 0.$$

Proof. To show the statement in Theorem 3.1 requires proofs of the following three parts. First we calculate explicitly the solution to equation (3.7) after applying (3.8) at the limit $\varepsilon = 0$. Then we show that the sequence of scaled solutions indexed by ε converges to the limit as $\varepsilon \downarrow 0$ using the semigroup property and further show the sequence is bounded above by an integrable function, thus concluding the convergence.

First we verify the inverse map $x(u, v)$, $y(u, v)$ exists, i.e., it is locally invertible. In fact, for $\varepsilon > 0$, $x > 0$, $y > 0$, the inverse map of (3.8) admits a Jacobian with nonzero determinant

$$\forall \varepsilon > 0, \quad x > 0, \quad y > 0$$

$$\left| \frac{\partial(x, y)}{\partial(u, v)} \right| = \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} = \varepsilon^2 x^\beta y^2 \neq 0.$$

Therefore the inverse map of (3.8) is one to one and explicitly given by

$$(3.9) \quad \begin{cases} x(u, v) = [\varepsilon(1 - \beta)u Y e^{\varepsilon v} + X^{1-\beta}]^{\frac{1}{1-\beta}} \\ y(u, v) = Y e^{\varepsilon v}. \end{cases}$$

Let us denote by $\hat{\mathcal{L}}_{u,v}^\varepsilon$ the resulting operator of $\tilde{\mathcal{L}}_{x,y}^\varepsilon$ upon applying (3.8). Through chain rule, $\hat{\mathcal{L}}_{u,v}^\varepsilon$ is calculated as follows:

$$\hat{\mathcal{L}}_{u,v}^\varepsilon \triangleq \frac{1}{2} [a_\varepsilon(x(u, v), y(u, v)) \partial_{xx} + 2b_\varepsilon(x(u, v), y(u, v)) \partial_{xy} + c_\varepsilon(x(u, v), y(u, v)) \partial_{yy}]$$

with

$$\begin{aligned} \partial_{xx} &= \left[\left(\frac{\partial\phi}{\partial x} \right)^2 \right] \frac{\partial^2}{\partial u^2} + \left[2 \frac{\partial\phi}{\partial x} \frac{\partial\psi}{\partial x} \right] \frac{\partial^2}{\partial u \partial v} + \left[\left(\frac{\partial\psi}{\partial x} \right)^2 \right] \frac{\partial^2}{\partial v^2} + \left[\frac{\partial^2\phi}{\partial x^2} \right] \frac{\partial}{\partial u} \\ &\quad + \left[\frac{\partial^2\psi}{\partial x^2} \right] \frac{\partial}{\partial v} \\ \partial_{xy} &= \left[\frac{\partial\phi}{\partial x} \frac{\partial\phi}{\partial y} \right] \frac{\partial^2}{\partial u^2} + \left[\frac{\partial\phi}{\partial y} \frac{\partial\psi}{\partial x} + \frac{\partial\phi}{\partial x} \frac{\partial\psi}{\partial y} \right] \frac{\partial^2}{\partial u \partial v} + \left[\frac{\partial\psi}{\partial x} \frac{\partial\psi}{\partial y} \right] \frac{\partial^2}{\partial v^2} \\ &\quad + \left[\frac{\partial^2\phi}{\partial x \partial y} \right] \frac{\partial}{\partial u} + \left[\frac{\partial^2\psi}{\partial x \partial y} \right] \frac{\partial}{\partial v} \\ \partial_{yy} &= \left[\left(\frac{\partial\phi}{\partial y} \right)^2 \right] \frac{\partial^2}{\partial u^2} + \left[2 \frac{\partial\phi}{\partial y} \frac{\partial\psi}{\partial y} \right] \frac{\partial^2}{\partial u \partial v} + \left[\left(\frac{\partial\psi}{\partial y} \right)^2 \right] \frac{\partial^2}{\partial v^2} \\ &\quad + \left[\frac{\partial^2\phi}{\partial y^2} \right] \frac{\partial}{\partial u} + \left[\frac{\partial^2\psi}{\partial y^2} \right] \frac{\partial}{\partial v} \end{aligned}$$

and

$$\begin{cases} \frac{\partial \phi}{\partial x} = \frac{1}{\varepsilon x^\beta y} \\ \frac{\partial \phi}{\partial y} = \frac{x^{1-\beta} - X^{1-\beta}}{\varepsilon(1-\beta)} \left(\frac{-1}{y^2} \right) = \frac{-u}{y} \\ \frac{\partial^2 \phi}{\partial x^2} = \frac{-\beta}{\varepsilon x^{\beta+1} y} \\ \frac{\partial^2 \phi}{\partial xy} = \frac{-1}{\varepsilon x^\beta y^2} \\ \frac{\partial^2 \phi}{\partial y^2} = \frac{x^{1-\beta} - X^{1-\beta}}{\varepsilon(1-\beta)} \left(\frac{2}{y^3} \right) = \frac{2u}{y^2} \end{cases} \quad \begin{cases} \frac{\partial \psi}{\partial x} = 0 \\ \frac{\partial \psi}{\partial y} = \frac{1}{\varepsilon y} \\ \frac{\partial^2 \psi}{\partial x^2} = 0 \\ \frac{\partial^2 \psi}{\partial xy} = 0 \\ \frac{\partial^2 \psi}{\partial y^2} = \frac{-1}{\varepsilon y^2}. \end{cases}$$

Organizing terms, we then have

$$\hat{\mathcal{L}}_{u,v}^\varepsilon = \frac{1}{2} [l_\varepsilon(u, v) \partial_{uu} + 2m_\varepsilon(u, v) \partial_{uv} + n_\varepsilon(u, v) \partial_{vv} + j_\varepsilon(u, v) \partial_u + k_\varepsilon(u, v) \partial_v],$$

where $l_\varepsilon(u, v)$, $m_\varepsilon(u, v)$, $n_\varepsilon(u, v)$, $j_\varepsilon(u, v)$, $k_\varepsilon(u, v)$ are

$$\begin{aligned} l_\varepsilon(u, v) &= \left[a_\varepsilon(x, y) \left(\frac{\partial \phi}{\partial x} \right)^2 + 2b_\varepsilon(x, y) \frac{\partial \phi}{\partial x} \frac{\partial \phi}{\partial y} + c_\varepsilon(x, y) \left(\frac{\partial \phi}{\partial y} \right)^2 \right] = 1 - 2\rho u \varepsilon + u^2 \varepsilon^2 \\ m_\varepsilon(u, v) &= \left[a_\varepsilon(x, y) \frac{\partial \phi}{\partial x} \frac{\partial \psi}{\partial x} + b_\varepsilon(x, y) \left(\frac{\partial \phi}{\partial x} \frac{\partial \psi}{\partial y} + \frac{\partial \phi}{\partial y} \frac{\partial \psi}{\partial x} \right) + c_\varepsilon(x, y) \frac{\partial \phi}{\partial y} \frac{\partial \psi}{\partial y} \right] = \rho - \varepsilon u \\ n_\varepsilon(u, v) &= \left[a_\varepsilon(x, y) \left(\frac{\partial \psi}{\partial x} \right)^2 + 2b_\varepsilon(x, y) \frac{\partial \psi}{\partial x} \frac{\partial \psi}{\partial y} + c_\varepsilon(x, y) \left(\frac{\partial \psi}{\partial y} \right)^2 \right] = 1 \\ j_\varepsilon(u, v) &= \left[a_\varepsilon(x, y) \frac{\partial^2 \phi}{\partial x^2} + 2b_\varepsilon(x, y) \frac{\partial^2 \phi}{\partial x \partial y} + c_\varepsilon(x, y) \frac{\partial^2 \phi}{\partial y^2} \right] = -(2\rho + \beta x^{\beta-1} y) \varepsilon + 2u \varepsilon^2 \\ k_\varepsilon(u, v) &= \left[a_\varepsilon(x, y) \frac{\partial^2 \psi}{\partial x^2} + 2b_\varepsilon(x, y) \frac{\partial^2 \psi}{\partial x \partial y} + c_\varepsilon(x, y) \frac{\partial^2 \psi}{\partial y^2} \right] = -\varepsilon. \end{aligned}$$

Then (3.7) becomes an equation in coordinates (u, v)

$$(3.10) \quad \begin{cases} \left[\frac{\partial}{\partial \tau} - \hat{\mathcal{L}}_{u,v}^\varepsilon \right] \tilde{p}_\varepsilon(\tau, x(u, v), y(u, v); X, Y) = 0, & 0 < \tau \leq 1 \\ \tilde{p}_\varepsilon(\tau, x(u, v), y(u, v); X, Y) = \frac{1}{\varepsilon^2 X^\beta Y^2} \delta(u) \delta(v), & \tau = 0 \end{cases}$$

with $(x(u, v), y(u, v))$ given by (3.9).

Next define a new quantity $\hat{p}_\varepsilon(\tau, u, v; U, V)$ as

$$(3.11) \quad \hat{p}_\varepsilon(\tau, u, v; U, V) \triangleq \varepsilon^2 X^\beta Y^2 \tilde{p}_\varepsilon(\tau, x(u, v), y(u, v); X, Y)$$

and plug (3.11) into equation (3.10), we have $\hat{p}_\varepsilon(\tau, u, v; U, V)$ satisfying

$$(3.12) \quad \begin{cases} \left[\frac{\partial}{\partial \tau} - \hat{\mathcal{L}}_{u,v}^\varepsilon \right] \hat{p}_\varepsilon(\tau, u, v; U, V) = 0, & 0 < \tau \leq 1 \\ \hat{p}_\varepsilon(\tau, u, v; U, V) = \delta(u)\delta(v), & \tau = 0 \end{cases}$$

with $\hat{\mathcal{L}}_{u,v}^\varepsilon$ and $x^{\beta-1}y$ in terms of u, v given by

$$(3.13) \quad \hat{\mathcal{L}}_{u,v}^\varepsilon = \frac{1}{2} \left[\begin{aligned} &(1 - 2\rho\varepsilon + \varepsilon^2u^2) \frac{\partial^2}{\partial u^2} + 2(\rho - \varepsilon u) \frac{\partial^2}{\partial uv} + \frac{\partial^2}{\partial v^2} \\ &+ [-(2\rho + \beta x^{\beta-1}y)\varepsilon + 2u\varepsilon^2] \frac{\partial}{\partial u} + (-\varepsilon) \frac{\partial}{\partial v} \end{aligned} \right]$$

$$x^{\beta-1}y = \frac{Ye^{\varepsilon v}}{\varepsilon(1 - \beta)uYe^{\varepsilon v} + X^{1-\beta}}.$$

We will shorten $\hat{p}_\varepsilon(\tau, u, v; U, V)$ as $\hat{p}_\varepsilon(\tau, u, v)$ from now on.

Setting $\varepsilon = 0$ in (3.13), we obtain the limiting generator as

$$\hat{\mathcal{A}}^0 = \frac{1}{2} \left[\frac{\partial^2}{\partial u^2} + 2\rho \frac{\partial^2}{\partial uv} + \frac{\partial^2}{\partial v^2} \right]$$

with $\hat{p}_0(\tau, u, v)$ denoting the solution to the corresponding limiting equation of (3.12).

Then $\hat{p}_0(\tau, u, v)$ solves

$$\begin{cases} \frac{\partial \hat{p}_0}{\partial \tau} - \left[\frac{\partial^2 \hat{p}_0}{\partial u^2} + 2\rho \frac{\partial^2 \hat{p}_0}{\partial uv} + \frac{\partial^2 \hat{p}_0}{\partial v^2} \right] (\tau, u, v) = 0, & \tau \in (0, 1] \\ \hat{p}_0(\tau, u, v) = \delta(u)\delta(v), & \tau = 0 \end{cases}$$

and is explicitly given by

$$\hat{p}_0(\tau, u, v) = \frac{1}{2\pi\tau\sqrt{1 - \rho^2}} \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2\tau(1 - \rho^2)}\right).$$

The associated semigroup sequence T_τ^ε defined via $\tilde{p}_\varepsilon(\tau, u, v)$ as

$$(T_\tau^\varepsilon f)(u, v) \triangleq \int f(u, v) \tilde{p}_\varepsilon(\tau, u, v; u', v') du' dv' \in C_b, \quad \forall f \in C_b$$

is a strong continuous contraction semigroup. Then $\forall f \in C_b^2(\mathbb{R}_+^2)$,

$$(T_\tau^\varepsilon f) \rightarrow (T_\tau f) \quad \text{on } C_b^2(\mathbb{R}_+^2) \quad \text{as } \varepsilon \rightarrow 0$$

which implies the convergence of the kernel Ethier and Kurtz (1986)

$$\hat{p}_\varepsilon(\tau, u, v) \rightarrow \hat{p}_0(\tau, u, v) \quad \text{a.e. as } \varepsilon \rightarrow 0.$$

Further at each fixed ε , there exists a constant $K > 0$ such that for all $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$ and all $(u, v) \in \mathbb{R}^2$, we have

$$\begin{aligned} a_\varepsilon(u, v)\xi_1^2 + 2b_\varepsilon(u, v)\xi_1\xi_2 + c_\varepsilon(u, v)\xi_2^2 &\leq (1 - \varepsilon\rho u + \varepsilon^2u^2)\xi_1^2 + (\rho - \varepsilon u)(\xi_1^2 + \xi_2^2) + \xi_2^2 \\ &\leq K(1 + u^2 + v^2)(\xi_1^2 + \xi_2^2), \end{aligned}$$

then each $\hat{p}_\varepsilon(\tau, u, v)$ is bounded above by Gaussian kernel (Fabes and Stroock 1984, 1986; Fabes 1992) which is clearly integrable

$$|\hat{p}_\varepsilon(\tau, u, v)| \leq \frac{C_K}{2\pi\tau} \exp\left(-\frac{u^2 + v^2}{2\tau C_K}\right)$$

for some constant C_K depending only on K . Thus not only $\hat{p}_\varepsilon \rightarrow \hat{p}_0$ as $\varepsilon \downarrow 0$ but also the sequence \hat{p}_ε is bounded by a integrable function for all ε . Taking (u, v) as a continuity point, Lebesgue Dominated Convergence Theorem then concludes the convergence. \square

4. SERIES EXPANSION

4.1. Convergence of Expansion

With the total volatility-of-volatility scaling and near-Gaussian transformation established in Section 3, we have successfully transformed the original problem (2) into (3.12) which we will rewrite here as

$$(4.1) \quad \left\{ \begin{array}{l} \left[\frac{\partial}{\partial \tau} - \hat{\mathcal{L}}_{u,v}^\varepsilon \right] \hat{p}_\varepsilon(\tau, u, v) = 0, \quad 0 < \tau \leq 1 \\ \hat{p}_\varepsilon(\tau, u, v) = \delta(u)\delta(v), \quad \tau = 0 \\ \text{where } \hat{\mathcal{L}}_{u,v}^\varepsilon = \frac{1}{2} \left[(1 - 2\rho u\varepsilon + \varepsilon^2 u^2) \frac{\partial^2}{\partial u^2} + 2(\rho - \varepsilon u) \frac{\partial^2}{\partial u \partial v} + \frac{\partial^2}{\partial v^2} \right. \\ \quad \left. + [-(2\rho + \beta x^{\beta-1} y)\varepsilon + 2u\varepsilon^2] \frac{\partial}{\partial u} + (-\varepsilon) \frac{\partial}{\partial v} \right] \\ x^{\beta-1} y = \frac{Y e^{\varepsilon v}}{\varepsilon(1 - \beta)u Y e^{\varepsilon v} + X^{1-\beta}} \end{array} \right.$$

and so far everything is exact and no approximation is made. To prepare for a power series expansion, we further expand $x^{\beta-1} y$ to the first order of ε

$$x^{\beta-1} y = [X^{\beta-1} Y] + [[(\beta - 1)X^{2(\beta-1)} Y^2]u + [X^{\beta-1} Y]v]\varepsilon + O(\varepsilon^2).$$

With the limiting solution $\hat{p}_0(\tau, u, v)$ established in Theorem 3.1, we will show that the following proposed power series representation of $\hat{p}_\varepsilon(\tau, u, v)$ converges in the sense that the finite partial sum of this expansion is a global L^1 solution to (4.1) with $x^{\beta-1} y$ expanded the second order of ε . By a global L^1 solution, we mean that the infinite power series expansion has bounded L^1 norm for every $\tau \in (0, 1)$.

Plugging the following expansion ansatz:

$$(4.2) \quad \hat{p}_\varepsilon = \hat{p}_0 + \varepsilon \hat{p}_1 + \varepsilon^2 \hat{p}_2 + \dots + \varepsilon^n \hat{p}_n + \dots$$

into (4.1) for positive integer $n \geq 1$, and equating like orders of ε for both the equation and the initial condition, we obtain the following hierarchy:

$$\begin{aligned} \left[\frac{\partial}{\partial \tau} - \frac{1}{2} \mathcal{L}_\varepsilon^0 \right] \hat{p}_0(\tau, u, v) &= 0; & \hat{p}_0(0, u, v) &= \delta(u)\delta(v) \\ \left[\frac{\partial}{\partial \tau} - \frac{1}{2} \mathcal{L}_\varepsilon^0 \right] \hat{p}_1(\tau, u, v) &= \frac{1}{2} [\mathcal{L}_\varepsilon^1 \hat{p}_0](\tau, u, v); & \hat{p}_1(0, u, v) &= 0 \end{aligned}$$

$$\begin{aligned}
 \left[\frac{\partial}{\partial \tau} - \frac{1}{2} \mathcal{L}_\varepsilon^0 \right] \hat{p}_2(\tau, u, v) &= \frac{1}{2} [\mathcal{L}_\varepsilon^1 \hat{p}_1 + \mathcal{L}_\varepsilon^2 \hat{p}_0](\tau, u, v); & \hat{p}_2(0, u, v) &= 0 \\
 & \vdots \\
 \left[\frac{\partial}{\partial \tau} - \frac{1}{2} \mathcal{L}_\varepsilon^0 \right] \hat{p}_n(\tau, u, v) &= \frac{1}{2} [\mathcal{L}_\varepsilon^1 \hat{p}_{n-1} + \mathcal{L}_\varepsilon^2 \hat{p}_{n-2}](\tau, u, v); & \hat{p}_n(0, u, v) &= 0 \\
 & \vdots
 \end{aligned}
 \tag{4.3}$$

with $\mathcal{L}_\varepsilon^0$, $\mathcal{L}_\varepsilon^1$ and $\mathcal{L}_\varepsilon^2$ denoting the source operators at each order of ε

$$\begin{cases}
 \mathcal{L}_\varepsilon^0 = \frac{\partial^2}{\partial u^2} + 2\rho \frac{\partial^2}{\partial u \partial v} + \frac{\partial^2}{\partial v^2} \\
 \mathcal{L}_\varepsilon^1 = -2\rho u \frac{\partial^2}{\partial u^2} - 2u \frac{\partial^2}{\partial u \partial v} - (2\rho + \beta X^{\beta-1} Y) \frac{\partial}{\partial u} - \frac{\partial}{\partial v} \\
 \mathcal{L}_\varepsilon^2 = u^2 \frac{\partial^2}{\partial u^2} + [(2 + \beta(1 - \beta) X^{2(\beta-1)} Y^2) u - (\beta X^{\beta-1} Y) v] \frac{\partial}{\partial u}.
 \end{cases}
 \tag{4.4}$$

We call $\hat{p}_0(\tau, u, v)$ the leading order expansion to \hat{p}_ε and it is the solution to the first equation in the hierarchy (4.3). $\hat{p}_1(\tau, u, v)$ is called the first-order expansion which solves the second equation in (4.3) and is explicitly obtainable by applying Duhamel’s principle once $\hat{p}_0(\tau, u, v)$ is obtained. Similarly $\hat{p}_2(\tau, u, v)$ is the second-order expansion which solves the third equation in (4.3) once \hat{p}_1 and \hat{p}_0 are obtained. Progressively, the expansion hierarchy in (4.3) up to finite order n can be obtained explicitly.

Let us denote the partial sum truncated at the n th order expansion by

$$\hat{p}_\varepsilon^n = \hat{p}_0 + \varepsilon \hat{p}_1 + \dots + \varepsilon^n \hat{p}_n, \quad \text{for } n \geq 1.
 \tag{4.5}$$

We then introduce the mixed norm $L^2_\tau((0, 1)) \times L^1_{u,v}(\mathbb{R}^2)$ to make precise the expansion and convergence

$$\|f(\tau, u, v)\|_{L^2_\tau((0,1)) \times L^1_{u,v}(\mathbb{R}^2)} := \left[\int_0^1 \left| \int_{\mathbb{R}^2} |f(\tau, u, v)| \, du \, dv \right|^2 \, d\tau \right]^{\frac{1}{2}}.$$

A function $f(\tau, u, v)$ is further said to be in the space of $L^2_\tau((0, 1)) \times L^1_{u,v}(\mathbb{R}^2) \cap C^{1,2}((0, 1) \times \mathbb{R}^2)$

$$\left\{ f : C^{1,2}((0, 1) \times \mathbb{R}^2) \rightarrow \mathbb{R} \mid \left[\int_0^1 \left| \int_{\mathbb{R}^2} |f(\tau, u, v)| \, du \, dv \right|^2 \, d\tau \right]^{\frac{1}{2}} < \infty \right\}$$

if it is differentiable once in τ and twice in u, v and further it has finite L^1 norm in the spatial variables u, v on \mathbb{R}^2 and finite L^2 norm in the temporal variable τ on interval $(0, 1)$. We then have the following result.

THEOREM 4.1. *Let $\hat{p}_\varepsilon : (0, 1) \times \mathbb{R}^2 \mapsto \mathbb{R}$ be a $C^{1,2}$ function that solves (4.1). Let $\hat{p}_i : (0, 1) \times \mathbb{R}^2 \mapsto \mathbb{R}, i = 1, \dots, n$ be a hierarchy of $C^{1,2}$ functions that solves the equation hierarchy in (4.3) and let \hat{p}_ε^n be the partial sum of \hat{p}_i defined in (4.5), then there exists a constant $C > 0$, independent of ε , such that $\forall \varepsilon > 0$*

$$(4.6) \quad \|\hat{p}_\varepsilon(\tau, u, v) - \hat{p}_\varepsilon^n(\tau, u, v)\|_{L^2_{\tau((0,1)) \times L^1_{u,v}(R^2)}} \leq C\varepsilon^{n+1}.$$

Proof. The proof consists of the following three steps. The first step is to find an integral representation for the truncation error after n th-order expansion. We denote the truncation error as

$$\xi_\varepsilon^n := \hat{p}_\varepsilon - \hat{p}_\varepsilon^n.$$

This requires an evolution equation that ξ_ε^n satisfies and it is obtained by summing up the expansion hierarchy. Then by Duhamel’s principle, we will show that we can represent the solution to the error equation as a convolution of a correlated bivariate Gaussian kernel with source terms that involves only the n th and $(n - 1)$ th order expansion.

In the next step, we observe that for expansion at any given order n other than the leading order, the corresponding equation in hierarchy (4.3) involves the source operators $\mathcal{L}_\varepsilon^1, \mathcal{L}_\varepsilon^2$ (4.4) which are differentiations of expansions at order $n - 1$ and $n - 2$. Specifically, the highest order effect of the source operator acting on the fundamental solution of hierarchy (4.3) will be multiplication by spatial polynomials $u^m v^n$ and temporal polynomials τ^{-l} . All (m, n, l) are positive integers.

Finally, through Young’s inequality, we are able to obtain L^1 control for the spatial convolution as both of the kernel as functions in the spatial domain is of Gaussian and Gaussian moments type and hence L^p for $p \geq 1$. For the temporal integration, the kernel as functions in the temporal variable does not have enough decay within the subinterval of $(0,1)$ and in fact one of them is only L^2 above, hence we obtain a L^2 control on the temporal integration. □

Step I: Equation for ξ_ε^n

Recalling that we have an expansion ansatz as in (4.2), let us formally complete the expansion as

$$(4.7) \quad \hat{p}_\varepsilon = \hat{p}_0 + \varepsilon \hat{p}_1 + \varepsilon^2 \hat{p}_2 + \dots + \varepsilon^n \hat{p}_n + \xi_\varepsilon^n = \hat{p}_\varepsilon^n + \xi_\varepsilon^n$$

with ξ_ε^n denoting the truncation error after the expansion to the n th order. Plugging (4.7) into (4.3) and with some algebra, we find all the “lower-order” terms in the subtraction cancel and what remains is the equation for the error term ξ_ε^n

$$(4.8) \quad \left[\frac{\partial}{\partial \tau} - \frac{1}{2} \mathcal{L}_\varepsilon^0 \right] \xi_\varepsilon^n(\tau, u, v) = \frac{1}{2} \varepsilon^{n+1} [\mathcal{L}_\varepsilon^1 \hat{p}_n + \mathcal{L}_\varepsilon^2 \hat{p}_{n-1} + \varepsilon \mathcal{L}_\varepsilon^2 \hat{p}_n](\tau, u, v), \quad \xi_\varepsilon^n(0, u, v) = 0.$$

The solution to the error equation will then be established in the following proposition as a direct consequence of applying Duhamel’s principle to (4.8) for an inhomogeneous evolution equation with $[\partial_\tau - \frac{1}{2} \mathcal{L}_\varepsilon^0]$ as the evolution operator. And indeed, the full equation hierarchy shares the same left-hand side as $[\partial_\tau - \frac{1}{2} \mathcal{L}_\varepsilon^0]$ and the solution structure of the truncation error will be similar to that of the expansions.

PROPOSITION 4.2. *Consider the inhomogeneous Cauchy problem*

$$(4.9) \quad \begin{aligned} \left[\frac{\partial}{\partial \tau} - \frac{1}{2} \left(\frac{\partial}{\partial u^2} + 2\rho \frac{\partial}{\partial uv} + \frac{\partial}{\partial v^2} \right) \right] \hat{p}_\varepsilon(\tau, u, v) &= f(\tau, u, v), & \tau > 0 \\ \hat{p}_\varepsilon(\tau, u, v) &= g(u, v), & \tau = 0 \end{aligned}$$

the solution to (4.9) is represented via that of the homogeneous problem through the initial condition $g(\tau, u, v)$ and the source terms $f(\tau, u, v)$ as

$$(4.10) \quad \hat{p}_\varepsilon(\tau, u, v) = \int_{\mathbb{R}^2} G(\tau, u - u', v - v')g(u', v') du' dv' + \int_0^\tau \int_{\mathbb{R}^2} G(\tau - s, u - u', v - v')f(s, u', v') du' dv' ds,$$

where the kernel $G(\cdot)$ in (4.10) is the fundamental solution of (4.9) solving the following the homogeneous problem

$$\left[\frac{\partial}{\partial \tau} - \frac{1}{2} \left(\frac{\partial}{\partial u^2} + 2\rho \frac{\partial}{\partial uv} + \frac{\partial}{\partial v^2} \right) \right] G(\tau, u, v) = 0, \quad \tau > 0$$

$$G(\tau, u, v) = \delta(u)\delta(v), \quad \tau = 0$$

and is given by

$$(4.11) \quad G(\tau, u, v) = \frac{1}{2\pi\tau\sqrt{1-\rho^2}} \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2\tau(1-\rho^2)}\right).$$

Proof. Equation (4.10) is simply Duhamel’s principle and (4.11) follows from recognizing the equation as the generator for two-dimensional Brownian motion with correlation ρ .

Equation (4.11) is the bivariate Gaussian limit that we saw in Theorem 3.1 as the limiting density function to a correlated bivariate Brownian motion. With Proposition 4.2, the solution to the error equation (4.8) is given by

$$(4.12) \quad \xi_\varepsilon^n(\tau, u, v) = \frac{1}{2}\varepsilon^{n+1} \int_0^\tau \int_{\mathbb{R}^2} G(\tau - s, u - x, v - y) \times [\mathcal{L}_\varepsilon^1 p^n + \mathcal{L}_\varepsilon^2 p^{n-1} + \varepsilon \mathcal{L}_\varepsilon^2 p^n](s, x, y) dx dy ds.$$

For instance, the truncation error after the first-order expansion will be

$$\xi_\varepsilon^1 = \frac{1}{2}\varepsilon^2 \int_0^\tau \int_{\mathbb{R}^2} G(\tau - s, u - x, v - y) [\mathcal{L}_\varepsilon^1 p^1 + \mathcal{L}_\varepsilon^2 p^0 + \varepsilon \mathcal{L}_\varepsilon^2 p^1](s, x, y) dx dy ds.$$

Thus we have successfully obtained an equation for the truncation error and observe it is of order ε^{n+1} from the coefficients in (4.12). What remains is to show that the integral in (4.12), aside from the coefficient ε^{n+1} , is bounded. \square

Step II: Effect of Source Operator on the Green’s Function.

To control the size of the integral in (4.12), we need an analysis of the effect of the source operator $\mathcal{L}_\varepsilon^1$ and $\mathcal{L}_\varepsilon^2$ on the expansions, precisely at order n and $n - 1$. In fact $\mathcal{L}_\varepsilon^1, \mathcal{L}_\varepsilon^2$ are differentiations in the spatial variables whose effect can be characterized as multiplication by a polynomial function in the state variables τ, u, v . The complication in our case is the fact that $G(\tau, u, v)$ is of Gaussian type only in u, v , but not in τ ; further it is bivariate with correlation.

For now let us first write down explicitly the solution representation to the expansion hierarchy up to order n in (4.3). Invoking (4.10) in Proposition 4.2 and identifying individually the initial condition and the source term at different expansion order, the solutions to the hierarchy up to order n in (18) can be represented by

$$\begin{aligned}
 \hat{p}_0(\tau, u, v) &= G(\tau, u, v); && \text{for } n = 0 \\
 \hat{p}_1(\tau, u, v) &= \int_0^\tau \int_{\mathbb{R}^2} G(\tau - s, u - x, v - y) \\
 &\quad \times \left[\frac{1}{2} \mathcal{L}_\varepsilon^1 G \right] (s, x, y) dx dy ds; && \text{for } n = 1 \\
 \hat{p}_2(\tau, u, v) &= \int_0^\tau \int_{\mathbb{R}^2} G(\tau - s, u - x, v - y) \\
 &\quad \times \left[\frac{1}{2} \mathcal{L}_\varepsilon^1 \hat{p}_1 + \frac{1}{2} \mathcal{L}_\varepsilon^2 \hat{p}_0 \right] (s, x, y) dx dy ds; && \text{for } n = 1 \\
 &\quad \vdots \\
 \hat{p}_n(\tau, u, v) &= \int_0^\tau \int_{\mathbb{R}^2} G(\tau - s, u - x, v - y) \frac{1}{2} \\
 &\quad \times \left[\mathcal{L}_\varepsilon^1 \hat{p}_{n-1} + \mathcal{L}_\varepsilon^2 \hat{p}_{n-2} \right] (s, x, y) dx dy ds; && \text{for } n \geq 2.
 \end{aligned}
 \tag{4.13}$$

To shorten notation, we use “ \otimes ” to represent this specific form of integral where convolution in the spatial variables u, v is defined for the full domain while in the temporal variable τ it is a partial convolution on an interval

$$[f \otimes g](\tau, u, v) := \int_0^\tau \int_{\mathbb{R}^2} f(\tau - s, u - x, v - y) G(s, x, y) dx dy ds.$$

Then (4.13) in this simplified notation becomes

$$\begin{aligned}
 \hat{p}_0(\tau, u, v) &= G; && \text{for } n = 0 \\
 \hat{p}_1(\tau, u, v) &= G \otimes \left[\frac{1}{2} \mathcal{L}_\varepsilon^1 G \right]; && \text{for } n = 1 \\
 \hat{p}_2(\tau, u, v) &= G \otimes \frac{1}{2} \mathcal{L}_\varepsilon^1 \left[G \otimes \left[\frac{1}{2} \mathcal{L}_\varepsilon^1 G \right] \right] + G \otimes \left[\frac{1}{2} \mathcal{L}_\varepsilon^2 G \right] \\
 &\quad \vdots \\
 p^n(\tau, u, v) &= G \otimes \frac{1}{2} \left[\mathcal{L}_\varepsilon^1 p^{n-1} + \mathcal{L}_\varepsilon^2 p^{n-2} \right] && \text{for } n \geq 2
 \end{aligned}$$

with the solution to the hierarchy written down, we can now examine the effect of the two source operators $\mathcal{L}_\varepsilon^1, \mathcal{L}_\varepsilon^2$ in (19) on $p^n(\tau, u, v)$ where G is given by (4.11).

On explicit calculation, we see the exact form of the first-order derivative $\partial_u G, \partial_v G$ and the second-order derivative $\partial_{uu} G, \partial_{uv} G$. They are the four core components of the effect of the source operator $\mathcal{L}_\varepsilon^1$ and $\mathcal{L}_\varepsilon^2$ on G .

$$\begin{aligned}
 \frac{\partial G(\tau, u, v)}{\partial u} &= \left[\frac{u - \rho v}{(1 - \rho^2)\tau} \right] G(\tau, u, v) \\
 &= \left[\frac{1}{2\pi(1 - \rho^2)^{3/2}} \right] \left[\frac{u - \rho v}{\tau^2} \right] \exp \left(-\frac{u^2 - 2\rho uv + v^2}{2\tau(1 - \rho^2)} \right)
 \end{aligned}$$

$$\begin{aligned}
\frac{\partial G(\tau, u, v)}{\partial v} &= \left[\frac{v - \rho u}{(1 - \rho^2)\tau} \right] G(\tau, u, v) \\
&= \left[\frac{1}{2\pi(1 - \rho^2)^{3/2}} \right] \left[\frac{v - \rho u}{\tau^2} \right] \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2\tau(1 - \rho^2)}\right) \\
\frac{\partial^2 G(\tau, u, v)}{\partial u^2} &= \left[\frac{(u - \rho v)^2 + (-1 + \rho^2)\tau}{(1 - \rho^2)^2\tau^2} \right] G(\tau, u, v) \\
&= \left[\frac{1}{2\pi(1 - \rho^2)^{5/2}} \right] \left[\frac{(u - \rho v)^2 + (-1 + \rho^2)\tau}{\tau^3} \right] \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2\tau(1 - \rho^2)}\right) \\
\frac{\partial^2 G(\tau, u, v)}{\partial u \partial v} &= \left[\frac{(-\rho(u^2 + v^2) + (1 + \rho^2)uv) + \rho(1 - \rho^2)\tau}{(1 - \rho^2)^2\tau^2} \right] G(\tau, u, v) \\
&= \left[\frac{1}{2\pi(1 - \rho^2)^{5/2}} \right] \left[\frac{(-\rho(u^2 + v^2) + (1 + \rho^2)uv) + \rho(1 - \rho^2)\tau}{\tau^3} \right] \\
&\quad \times \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2\tau(1 - \rho^2)}\right).
\end{aligned}$$

As we see, $\partial_u G$ is of the form $u \exp(-u^2)$ in its highest order of u , the same for v up to a constant with a different sign. In terms of τ it is of the form $\frac{1}{\tau} \exp(-\frac{1}{\tau})$ in the highest order of τ . Further $\partial_{uu} G$ and $\partial_{uv} G$ are of the form $u^2 \exp(-\frac{1}{\tau})$ in the highest order of u and $\frac{1}{\tau^2} \exp(-\frac{1}{\tau})$ in the highest order of τ . Here we single out the highest order term in τ for analysis of the size due to the fact that $\forall \tau \in (0, 1)$, $\frac{1}{\tau}$ is an increasing function in n . More generally, when $\frac{\partial^{m+n}}{\partial^m u \partial^n v}$ acts on $G(\tau, u, v)$, the result is a polynomial coefficient with u^{m+n} and v^{m+n} as the highest order terms in the numerator and τ^{m+n} as the highest order term in the denominator. Without losing generality, we consider the highest order effect in terms of u and τ

$$\frac{\partial^{m+n}}{\partial^m u \partial^n v} G(\tau, u, v) \sim u^{m+n} \frac{1}{\tau^{m+n}} G(\tau, u, v),$$

where (m, n) are positive integers and by \sim we mean the highest order effect of the source operator.

If we can control the size of the mixed convolution type quantity $\frac{\partial^{m+n}}{\partial^m u \partial^n v} \otimes G$, then the size of expansion $p_\varepsilon^n(\tau, u, v)$ can be bounded as it consists of a finite number of items which are equal or smaller than the above quantity. The same argument applies to $p_\varepsilon^{n-1}(\tau, u, v)$.

Step III: Convolution Control.

Pertaining the analysis in step II, we next seek control for the following quantity:

$$(4.14) \quad I(\tau, u, v) \triangleq \int_0^\tau \int_{\mathbb{R}^2} G(\tau - s, u - x, v - y) \left[x^{m+n} \frac{1}{s^{l+1}} G(s, x, y) \right] (s, x, y) dx dy ds$$

and claim the following.

PROPOSITION 4.3. $\|I\|_{L_\tau^2((0,1))L_{uv}^1(\mathbb{R}^2)} < +\infty$.

Proof. Recall the classical Young’s inequality states the following. Let $f \in L^p$ and $g \in L^q$, then the convolution

$$f * g \triangleq \int_{\mathbb{R}^n} f(x - \xi)g(\xi) d\xi$$

satisfies inequality

$$\| f * g \|_{L^r} \leq \| f \|_{L^p} \| g \|_{L^q}$$

for

$$1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q} \quad \text{and} \quad 0 \leq p, q, r \leq +\infty.$$

The first observation is that (4.14) is an integral of convolution type. More specifically with G of two-dimensional Gaussian form in mind, the integration in the spatial variables x and y , omitting nonvariable dependent coefficients, is a full two-dimensional convolution on \mathbb{R}^2

$$(4.15) \quad [\exp(-(x^2 - 2\rho xy + y^2))] * [x^{m+n} \exp(-(x^2 - 2\rho xy + y^2))].$$

In contrast, the integration in the temporal variable s is of the form of a partial convolution as the range is a subset of finite interval $(0, \tau)$ on $(0, 1)$

$$(4.16) \quad \left[\frac{1}{t} \exp\left(-\frac{1}{t}\right) \right] * \left[\frac{1}{t^{l+1}} \exp\left(-\frac{1}{t}\right) \right].$$

Here “*” stands for the standard convolution on the full domain.

The second observation is that (4.15) and (4.16) as functions of space and as functions of time converge in different spaces. In fact $\exp(-(x^2 - 2\rho xy + y^2))$ and $x^{m+n} \exp(-(x^2 - 2\rho xy + y^2))$ have finite $L^p(\mathbb{R}^2)$ norm for $p \geq 1$ since they are Gaussian and Gaussian moments while for $t^{-1} \exp(-t^{-1})$ and $t^{-(l+1)} \exp(-t^{-1})$ we do not have such a uniform result on the whole temporal domain \mathbb{R}_+ . Indeed $t^{-1} \exp(-t^{-1})$ is not $L^1(\mathbb{R}_+)$ as the integration doesn’t converge on the half open interval \mathbb{R}_+ , its $L^p(\mathbb{R}^2)$ norm is finite only for $p \geq 2$. Here we will stay with an L^2 argument as our purpose in the section is to show boundedness rather than obtain an explicit solution. The case for $t^{-(l+1)} \exp(-t^{-1})$ is nicer, it is always a L^p function for $p \geq 1$ and even for the half open real line \mathbb{R}_+ since $l + 1$ is always larger than 1 due to the source effect and the L^p integration converges uniformly for $p \geq 1$.

Therefore both G and $(u, v)^{m+n} G$ are L^1 in \mathbb{R}^2 . By Young’s inequality with $p = 1, q = 1$, we have $r = 1$ for

$$\begin{aligned} \| G(\tau - s, u, v) * (u)^{m+n} G(s, u, v) \|_{L^1(\mathbb{R}^2)} &\leq \| G(\tau - s, u, v) \|_{L^1(\mathbb{R}^2)} \cdot \| (u)^{m+n} G(s, u, v) \|_{L^1(\mathbb{R}^2)} \\ &\leq \frac{C_1}{\tau - s} \exp\left(\frac{-C_2}{\tau - s}\right) \cdot \frac{C_3}{s} \exp\left(\frac{-C_4}{s}\right) \\ &\leq C_5 \frac{1}{\tau - s} \exp\left(\frac{-1}{\tau - s}\right) \cdot \frac{1}{s} \exp\left(\frac{-1}{s}\right), \end{aligned}$$

where C_1, C_2, C_3, C_4, C_5 are constants chosen to ensure the inequalities follows and they are all bounded below from zero.

Then as $t^{-(l+1)}\exp(-t^{-1})$ is uniformly bounded in L^p for $p \geq 1$ and $t^{-1}\exp(-t^{-1})$ is only L^2 and above, the best one will obtain is the case for $p = 1, q = 2$, hence $r = 2$ as

$$\begin{aligned} & \left\| C_5 \int_0^\tau \frac{1}{\tau-s} \exp\left(\frac{-1}{\tau-s}\right) \cdot \frac{1}{s^{l+1}} \exp\left(\frac{-1}{s}\right) ds \right\|_{L^2([0,1])} \\ & \leq \left\| C_5 \int_0^\infty \frac{1}{\tau-s} \exp\left(\frac{-1}{\tau-s}\right) \cdot \frac{1}{s^{l+1}} \exp\left(\frac{-1}{s}\right) ds \right\|_{L^2(\mathbb{R}_+)} \\ & = \left\| C_5 \frac{1}{\tau} \exp\left(\frac{-1}{\tau}\right) * \frac{1}{s^{l+1}} \exp\left(\frac{-1}{s}\right) \right\|_{L^2(\mathbb{R}_+)} \\ & \leq C_6 \left\| \frac{1}{\tau} \exp\left(\frac{-1}{\tau}\right) \right\|_{L^2(\mathbb{R}_+)} \cdot \left\| \frac{1}{s^{l+1}} \exp\left(\frac{-1}{s}\right) \right\|_{L^1(\mathbb{R}_+)} \\ & \leq C_6 \cdot C_7 \cdot C_8. \end{aligned}$$

Again C_6, C_7, C_8 are constants bounded away from zero.

Finally,

$$\|I\|_{L^2_t([0,1]) \times L^1_{uv}(\mathbb{R}^2)} \leq C_6 \cdot C_7 \cdot C_8 < +\infty.$$

□

5. JOINT DENSITY FORMULAS

Having proved the convergence of the expansion, we will calculate explicit expansion formulas up to the second order and illustrate the accuracy by numerical examples.

5.1. Leading Order

Recall the solution to the leading order problem is given by a bivariate normal distribution

$$(5.1) \quad \hat{p}_0(\tau, u, v) = G(\tau, u, v) = \frac{1}{2\pi\tau\sqrt{1-\rho^2}} \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2\tau(1-\rho^2)}\right).$$

5.2. First Order

Then the first-order system solves the following inhomogeneous problem with source term involving leading order solution. Together with zero initial condition, it reads:

$$\begin{cases} \left[\frac{\partial}{\partial \tau} - \frac{1}{2} \mathcal{L}_\varepsilon^0 \right] \hat{p}_1(\tau, u, v) = \frac{1}{2} [\mathcal{L}_\varepsilon^1 \hat{p}_0](\tau, u, v), & \forall 0 < \tau \leq 1 \\ \mathcal{L}_\varepsilon^1 = -2\rho u \frac{\partial^2}{\partial u^2} - 2u \frac{\partial^2}{\partial u \partial v} - (2\rho + \beta X^{\beta-1} Y) \frac{\partial}{\partial u} - \frac{\partial}{\partial v} \\ \hat{p}_1(0, u, v) = 0. \end{cases}$$

By Duhamel representation, we have $\hat{p}_1(\tau, u, v)$ as

$$(5.2) \quad \hat{p}_1(\tau, u, v) = \int_0^\tau \int_{\mathbb{R}^2} \frac{1}{2} [\mathcal{L}_\varepsilon^1 \hat{p}_0](s, x, y) G(\tau-s, u-x, v-y) dx dy ds.$$

As $\hat{p}_0(\tau, u, v) = G(\tau, u, v)$ from (5.2), we first calculate $\frac{1}{2}\mathcal{L}_\varepsilon^1 G$ as

$$\begin{aligned} \left[\frac{1}{2}\mathcal{L}_\varepsilon^1 G\right](\tau, u, v) &= f_{10}(\tau, u, v)G(\tau, u, v), \quad \text{with} \\ f_{10}(\tau, u, v) &= \left[-\frac{(2\rho + \beta X^{\beta-1} Y)(u - \rho v) + (v - \rho u)}{2(-1 + \rho^2)}\right] \frac{1}{\tau} + \left[\frac{uv(u - \rho v)}{(-1 + \rho^2)}\right] \frac{1}{\tau^2}. \end{aligned}$$

Then $\hat{p}_1(\tau, u, v)$ is obtained as

$$(5.3) \quad \begin{cases} \hat{p}_1(\tau, u, v) = \int_0^\tau \int_{\mathbb{R}^2} [f_{10}G](s, x, y)G(\tau - s, u - x, v - y) dx dy ds \\ \quad = C_1 \left[a_{11}(u, v) + a_{10}(u, v) \frac{1}{\tau} \right] \hat{p}_0(\tau, u, v), \end{cases}$$

where the coefficients $C_1, a_{11}(u, v), a_{10}(u, v)$ are

$$(5.4) \quad \begin{cases} C_1 = \frac{1}{2(-1 + \rho^2)} \\ a_{11}(u, v) = -\beta X^{\beta-1} Y(u - \rho v) \\ a_{10}(u, v) = uv(u - \rho v). \end{cases}$$

Finally,

$$\hat{p}_1(\tau, u, v) = \frac{-1}{4\pi\tau^2(1 - \rho^2)^{3/2}} [(u - \rho v)(uv - \beta X^{\beta-1} Y\tau)] \exp\left(-\frac{u^2 - 2\rho uv + v^2}{2\tau(1 - \rho^2)}\right).$$

5.3. Second Order

In the second-order expansion, the inhomogeneous problem will have two source terms involving solutions at both the first order and the leading order.

$$\begin{cases} \left[\frac{\partial}{\partial \tau} - \frac{1}{2}\mathcal{L}_\varepsilon^0\right] \hat{p}_2(\tau, u, v) = \frac{1}{2}[\mathcal{L}_\varepsilon^1 \hat{p}_1 + \mathcal{L}_\varepsilon^2 \hat{p}_0](\tau, u, v), \quad \forall 0 < \tau \leq 1 \\ \mathcal{L}_\varepsilon^1 = -2\rho u \frac{\partial^2}{\partial u^2} - 2u \frac{\partial^2}{\partial u \partial v} - (2\rho + \beta X^{\beta-1} Y) \frac{\partial}{\partial u} - \frac{\partial}{\partial v} \\ \mathcal{L}_\varepsilon^2 = u^2 \frac{\partial}{\partial u^2} + [(2 + \beta(1 - \beta) X^{2(\beta-1)} Y^2)u - (\beta X^{\beta-1} Y)v] \frac{\partial}{\partial u} \\ \hat{p}_2(0, u, v) = 0. \end{cases}$$

With $[\frac{1}{2}\mathcal{L}_\varepsilon^1 \hat{p}_1](\tau, u, v)$ and $[\frac{1}{2}\mathcal{L}_\varepsilon^2 \hat{p}_0](\tau, u, v)$ calculated as

$$\left[\frac{1}{2}\mathcal{L}_\varepsilon^1 \hat{p}_1\right](\tau, u, v) = f_{11}(\tau, u, v)G(\tau, u, v),$$

where

$$f_{11}(\tau, u, v) = \frac{(a - 2\rho)(a - \rho)}{4(-1 + \rho^2)} + \frac{1}{\tau^3} \frac{2u^4v^2 - 4u^3v^3\rho + 2u^2v^4\rho^2}{4(-1 + \rho^2)^2} \\ + \frac{1}{\tau^2} \frac{-2u^4 - 3au^3v + 7u^3v\rho - 3uv^3\rho(-1 + a\rho) + u^2v^2(-5 + 6a\rho - 3\rho^2)}{4(-1 + \rho^2)^2} \\ + \frac{1}{\tau} \frac{v^2(a - \rho)\rho(-2 + a\rho) + u^2(5 + a^2 - 3a\rho - 3\rho^2) + uv(5a - 2(4 + a^2)\rho + a\rho^2 + 4\rho^3)}{4(-1 + \rho^2)^2}$$

and

$$\left[\frac{1}{2} \mathcal{L}_\varepsilon^2 \hat{p}_0 \right] (\tau, u, v) = f_{20}(\tau, u, v) G(\tau, u, v),$$

where

$$f_{20}(\tau, u, v) = \frac{1}{\tau^2} \frac{u^2(u - v\rho)^2}{2(-1 + \rho^2)^2} + \frac{1}{\tau} \frac{(1 + b)u^2 + cv^2\rho - uv(c + b\rho)}{2(-1 + \rho^2)}.$$

Then $\hat{p}_2(\tau, u, v)$ is obtained as

$$(5.5) \quad \left\{ \begin{aligned} \hat{p}_2(\tau, u, v) &= \int_0^\tau \int_{\mathbb{R}^2} [f_{11}G + f_{20}G](s, x, y) G(\tau - s, u - x, v - y) dx dy ds \\ &= C_2 \left[a_{23}(u, v)\tau + a_{22}(u, v) + a_{21}(u, v)\frac{1}{\tau} + a_{20}(u, v)\frac{1}{\tau^2} \right] \hat{p}_0(\tau, u, v), \end{aligned} \right.$$

where the coefficients $C_2, a_{23}(u, v), a_{22}(u, v), a_{21}(u, v), a_{20}(u, v)$ are

$$\left\{ \begin{aligned} C_2 &= \frac{1}{24(1 - \rho^2)^2} \\ a_{23}(u, v) &= (8 - 3a^2 - 6b) + (12a)\rho + (-28 + 3a^2 + 12b)\rho^2 + (-12a)\rho^3 + (20 - 6b)\rho^4 \\ a_{22}(u, v) &= u^2[3a^2 - 12a\rho + 2(5 + \rho^2 + 3b(-1 + \rho^2))] \\ &\quad - 2uv[-3(a + c) + (10 + 3a^2 - 3b)\rho + 3(-3a + c)\rho^2 + (2 + 3b)\rho^3] \\ &\quad + v^2[-2 + (2 + 3(a - 2\rho)^2)\rho^2 + 6c\rho(-1 + \rho^2)] \\ a_{21}(u, v) &= u^4 + \rho^2v^4 + 2[-3a + 4\rho]u^3v + 2[\rho^2(-3a + 4\rho)]uv^3 \\ &\quad + 2[-2 + 6a\rho - 7\rho^2]u^2v^2 \\ a_{20}(u, v) &= 3u^4v^2 - 6u^3v^3\rho + 3u^2v^4\rho^2 \end{aligned} \right.$$

with a, b, c given by

$$a = 2\rho + \beta X^{\beta-1} Y, \quad b = 2 + \beta(1 - \beta) X^{2(\beta-1)} Y^2, \quad c = \beta X^{\beta-1} Y.$$

5.4. Explicit Formulas

Summarizing the result obtained so far, we start from the total volatility-of-volatility scaling in (3.1) and introduce the scale parameter $\varepsilon = v\sqrt{T}$ so $p(T - t, f, \alpha; T, F, A)$ becomes $\tilde{p}_\varepsilon(\tau, x, y; X, Y)$ as in (3.4). We then applied the near Gaussian transformation $(x, y) \rightarrow (u, v)$ in (3.8) to partially standardize the scaled SABR density from $\tilde{p}_\varepsilon(\tau, x, y; X, Y)$ to $\hat{p}_\varepsilon(\tau, u, v; U, V)$ as in (3.11). Finally, $\hat{p}_\varepsilon(\tau, u, v; U, V)$ is expanded in terms of ε (4.5) around the limiting solution $\hat{p}_0(\tau, u, v)$ (5.1).

Based on Theorem 4.1, the joint transition density $p(T - t, f, \alpha; F, A)$ at time to maturity $T - t$ conditional on the current state values f, α has the following series expansion:

$$(5.6) \quad p_n(T - t, f, \alpha; F, A) = \frac{1}{vTF^\beta A^2} \sum_{k=0}^n (v\sqrt{T})^k \hat{p}_k \left(\frac{T-t}{T}, \frac{f^{1-\beta} - F^{1-\beta}}{\alpha(1-\beta)\sqrt{T}}, \frac{\ln(\alpha/A)}{v\sqrt{T}} \right).$$

As we solve the expansion hierarchy (4.3) to the first three orders with $\hat{p}_0, \hat{p}_1, \hat{p}_2$ obtained in (5.2), (5.3), and (5.5), we express explicitly the density approximations in the following in terms of the original forward and backward variables $(t, f, \alpha; T, F, A)$ and model parameters (β, ρ, ν) .

(i) Approximation up to zero, first, and second order

$$(5.7) \quad p_0(T - t, f, \alpha; F, A) = \frac{1}{vTF^\beta A^2} \hat{p}_0(\tau, u, v)$$

$$(5.8) \quad p_1(T - t, f, \alpha; F, A) = \frac{1}{vTF^\beta A^2} [\hat{p}_0(\tau, u, v) + v\sqrt{T}\hat{p}_1(\tau, u, v)] \\ = \frac{1}{vTF^\beta A^2} \left[1 + \frac{v\sqrt{T}}{2(-1 + \rho^2)} (a_{11} + a_{10}/\tau) \right] \hat{p}_0(\tau, u, v)$$

$$(5.9) \quad p_2(T - t, f, \alpha; F, A) \\ = \frac{1}{vTF^\beta A^2} [\hat{p}_0(\tau, u, v) + v\sqrt{T}\hat{p}_1(\tau, u, v) + v^2T\hat{p}_2(\tau, u, v)] \\ = \frac{1}{vTF^\beta A^2} \left[1 + \frac{v\sqrt{T}}{2(-1 + \rho^2)} (a_{11} + a_{10}/\tau) \right. \\ \left. + \frac{v^2T}{24(1 - \rho^2)^2} (a_{23}\tau + a_{22} + a_{21}/\tau + a_{20}/\tau^2) \right] \hat{p}_0(\tau, u, v).$$

In (5.7), (5.8), and (5.9), transformed variables (τ, u, v) , zero-order expansion $\hat{p}_0(\tau, u, v)$, spatial coefficients at first order a_{11}, a_{10} and second order $a_{23}, a_{22}, a_{21}, a_{20}$ are explicitly given in terms of original variables (t, f, α, T, F, A) and model parameters β, ρ, ν as follows:

(ii) Transformed variables (τ, u, v) , $\hat{p}_0(\tau, u, v)$ and zero-order expansion $\hat{p}_0(\tau, u, v)$

$$(5.10) \quad \left\{ \begin{aligned} &\tau = \frac{T-t}{T}, \quad u = \frac{f^{1-\beta} - F^{1-\beta}}{\alpha(1-\beta)\sqrt{T}}, \quad v = \frac{\ln(\alpha/A)}{v\sqrt{T}} \\ &\hat{p}_0(\tau, u, v) = \frac{1}{2\pi\tau\sqrt{1-\rho^2}} \exp \left[-\frac{u^2 - 2\rho uv + v^2}{2\tau(1-\rho^2)} \right] \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 - \frac{t}{T} \right) \\ &\quad \times \exp \left[-\frac{\left(\frac{f^{1-\beta} - F^{1-\beta}}{\alpha(1-\beta)\sqrt{T}} \right)^2 - 2\rho \left(\frac{f^{1-\beta} - F^{1-\beta}}{\alpha(1-\beta)\sqrt{T}} \right) \left(\frac{\ln(\alpha/A)}{v\sqrt{T}} \right) + \left(\frac{\ln(\alpha/A)}{v\sqrt{T}} \right)^2}{2(1-\rho^2) \left(1 - \frac{t}{T} \right)} \right]. \end{aligned} \right.$$

(iii) Spatial coefficients at first and second order

$$(5.11) \quad \begin{cases} a_{11} = -\beta F^{\beta-1} A v^{-1} (u - \rho v), & a_{10} = u^2 v - \rho u v^2 \\ a_{23} = \rho^4 (20 - 6b) + \rho^3 (-12a) + \rho^2 (-28 + 3a^2 + 12b) + \rho (12a) + (8 - 3a^2 - 6b) \\ a_{22} = u^2 [3a^2 - 12a\rho + 6b(-1 + \rho^2) + 2\rho^2 + 10] \\ \quad - 2uv[\rho^3(2 + 3b) + \rho^2(-9a + 3c) + \rho(10 + 3a^2 - 3b) - (3a + 3c)] \\ \quad + v^2[(2 + 3(a - 2\rho)^2)\rho^2 + 6c\rho(-1 + \rho^2) - 2] \\ a_{21} = u^4 + v^4 \rho^2 + u^3 v(8\rho - 6a) + uv^3(8\rho^3 - 6a\rho^2) + u^2 v^2(-14\rho^2 + 12a\rho - 4) \\ a_{20} = 3u^4 v^2 - 6u^3 v^3 \rho + 3u^2 v^4 \rho^2 \\ a = 2\rho + \beta F^{\beta-1} A v^{-1}, \quad b = 2 + \beta(1 - \beta) F^{2(\beta-1)} A^2 v^{-2}, \quad c = \beta F^{\beta-1} A v^{-1}. \end{cases}$$

Through (5.7)–(5.10), expressions are explicit up to algebraic, logarithmic, and exponentiation operations. For $p(T - t, f, \alpha; F, A)$ to be a conditional probability function, the variables will be the time to maturity $T - t$, the future price F , and the future volatility A . The model parameters (β, ρ, v) and current level of underlying and volatility (f, α) are constants.

6. NUMERICAL EXAMPLES

To evaluate the accuracy of our result, we report numerical comparisons between ours, that of finite difference solver, and those from earlier work in Hagan et al. (2002) and Hagan, Lesniewski, and Woodward (2005). In particular, the comparisons are made in terms of four aspects: joint density, marginal density, conservation of probability mass, and implied volatility for European call options.

6.1. Joint Transition Density

Both point-wise l^∞ error and global l^2 error are compared for system (4.1) between expansions obtained at second-order expansion (5.5) and finite difference solver where ADI scheme is used to discretize (4.1), see page 64 in Morton and Mayers (2005) for details. The scheme is unconditionally stable and the discretization error is second order in both temporal and spatial variables with Dirichlet boundaries.

Equation (4.1) has Dirac initial mass centered at $(u, v) = (0, 0)$ and the solution to (4.1) decays very rapidly as the domain tends from the center $(0, 0)$ to \mathbb{R}^2 . We choose the domain under numerical evaluation to be $\mathcal{T}(\tau) \times \Omega(u, v) = [0, 1] \times [-6, 6] \times [-6, 6]$ in absolute values with a partition of 100 time steps in τ and 101×101 spatial grids in (u, v) . Dirichlet boundary conditions are imposed at $\hat{p}_\varepsilon(\tau, u, v) |_{\partial\Omega(u, v)} = 0, \forall \tau \in [0, 1]$ together with the discrete Dirac initial condition being $\hat{p}_\varepsilon(0, u, v) |_{(0,0)} = \frac{1}{\Delta u \Delta v}$ at the center and zero elsewhere $\hat{p}_\varepsilon(0, u, v) |_{\Omega(u, v)/(0,0)} = 0$. The domain chosen as such is consistent with discretizations in the sense that the numerical solution outside $\Omega(u, v)$ is observed to be below the level of discretization error. At each step of temporal marching, we have a 10201×10201 sparse matrix and we use a Bi-conjugate Gradient solver to solve the resulting linear system to precision 10^{-12} .

TABLE 6.1
 Point-wise l^∞ Error and Global l^2 Error between Joint Densities from Second-Order Expansion in Equation (5.9) and Joint Densities from Finite Difference Solver at T Equal to 1, 3, 6, 12 Months and ν Equal to 10%, 40%, 80%, 100%

$\ \hat{p}_\varepsilon - (\hat{p}_0 + \varepsilon \hat{p}_1 + \varepsilon^2 \hat{p}_2) \ $		$\beta = 0.0001$		$\beta = 0.5000$		$\beta = 0.9999$	
ν	T (months)	$\ \cdot\ _{l^\infty}$	$\log_{10}(\ \cdot\ _{l^2})$	$\ \cdot\ _{l^\infty}$	$\log_{10}(\ \cdot\ _{l^2})$	$\ \cdot\ _{l^\infty}$	$\log_{10}(\ \cdot\ _{l^2})$
0.1	1	0.2204%	-3.4093	0.2204%	-3.4093	0.2204%	-3.4089
0.4	1	0.2206%	-3.4066	0.2206%	-3.4066	0.2216%	-3.4023
0.8	1	0.2219%	-3.3352	0.2219%	-3.3352	0.2219%	-3.3209
1.0	1	0.2234%	-3.2009	0.2234%	-3.2008	0.2234%	-3.1845
0.1	3	0.2204%	-3.4092	0.2204%	-3.4092	0.2204%	-3.4065
0.4	3	0.2214%	-3.3727	0.2214%	-3.3726	0.2214%	-3.3385
0.8	3	0.2292%	-2.7376	0.2292%	-2.7375	0.2292%	-2.7055
1.0	3	0.2408%	-2.2746	0.2437%	-2.2746	0.3009%	-2.2569
0.1	6	0.2205%	-3.4089	0.2205%	-3.4089	0.2205%	-3.3988
0.4	6	0.2232%	-3.2186	0.2232%	-3.2183	0.2232%	-3.1287
0.8	6	0.3460%	-1.9985	0.3511%	-1.9984	0.4509%	-1.9744
1.0	6	0.6527%	-1.4938	0.6587%	-1.4937	0.7856%	-1.4817
0.1	12	0.2205%	-3.4078	0.2205%	-3.4077	0.2206%	-3.3700
0.4	12	0.2292%	-2.7376	0.2292%	-2.7372	0.2293%	-2.6214
0.8	12	0.9123%	-1.2176	0.9227%	-1.2175	1.1267%	-1.2103
1.0	12	1.6355%	-0.7247	1.6495%	-0.7246	1.9166%	-0.7165

The comparisons are made in three different market conditions categorized by the value of β , where $\beta = 0.0001$ corresponds to “normal” market, $\beta = 0.5$ for “CIR” market, and $\beta = 0.9999$ for “lognormal” market. As the expansion accuracy mostly depends on the perturbation parameter $\varepsilon = \nu\sqrt{T}$, we vary volatility-of-volatility ν from 10% to 100% and maturity T from 1 month to 12 months and fix other parameters in (4.1) as $\rho = 0$, $F = 100$, $A = 0.1$. It takes about 30 milliseconds for one evaluation of (5.5) on a machine with 2.93 GHZ Xeron CPU while the computation time for one call of the finite difference solver is about 6–7 minutes on the same machine. The potential savings in computation time are thus enormous.

Results are reported in Table 6.1. The point-wise l^∞ error ranges from 0.2% to 2% in absolute values for all three β cases, and the global l^2 error ranges from -3.4 to -0.7 in base 10 logarithm for all β cases. While the errors are all small across different values of ν and T , they increase monotonically as the perturbation parameter $\varepsilon = \nu\sqrt{T}$ increases, which is consistent with the small total volatility-of-volatility assumption for the expansions.

6.2. Marginal Transition Density

Earlier work on the marginal transition density from f to F is available in Hagan, Lesniewski, and Woodward (2005). To evaluate the accuracy of our method, we take the marginal density result in Hagan, Lesniewski, and Woodward (2005), which is readily available, as the benchmark and report comparisons to ours across a wide range of underlying levels f , time-to-maturities $T - t$, and model parameters $(\alpha, \beta, \rho, \nu)$. In

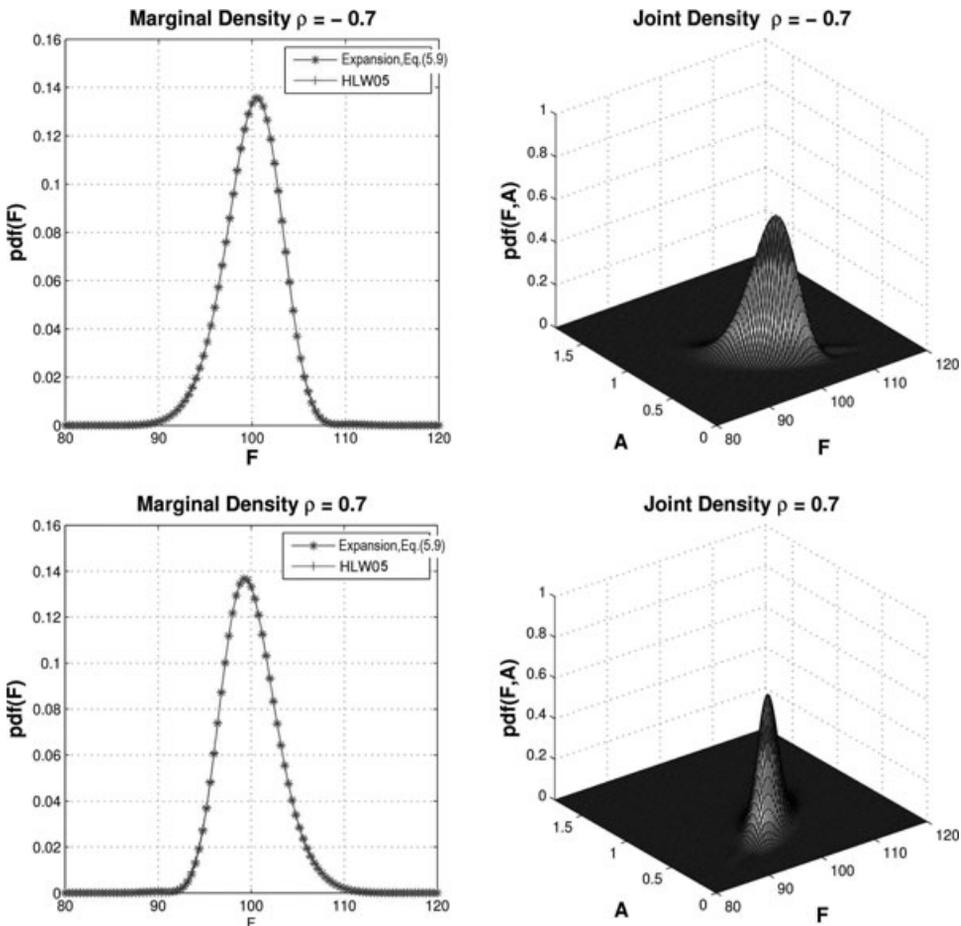


FIGURE 6.1. Marginal densities when correlation ρ equals to -0.7 and 0.7 . The upper two figures correspond to $\rho = -0.7$ and the lower two figures correspond to $\rho = 0.7$. Other parameters are $f = 100$, $\alpha = 60\%$, $\beta = 0.5$, $\nu = 40\%$, $T - t = 3$ months.

the case of Hagan, Lesniewski, and Woodward (2005), the marginal density is given by equation (90) which is analytically obtained by integrating the joint density expression from equation (83). In our case, we obtain the marginal density by numerically integrating the second-order joint density formula in equation (40) along the dimension of future volatility state A . Noted also that it is in the forward variable F that the marginal transition density $p_F(T - t, f, \alpha; F)$ is a probability distribution function, not the backward variables f, α .

To illustrate an important point that relates to the impact of parameter changes on the shape of densities, we also plot the corresponding joint transition distributions along with the marginals for the same model parameters used. The point we want to make by adding the joint density next to the marginal is that a change in the model parameter can lead to a dramatic change in the shape of joint density function but not so in the shape of marginals. And one should not wrongly imply that Hagan, Lesniewski, and Woodward (2005) did not provide a joint density expression. In fact, it is given by equation (90) in Hagan, Lesniewski, and Woodward (2005).

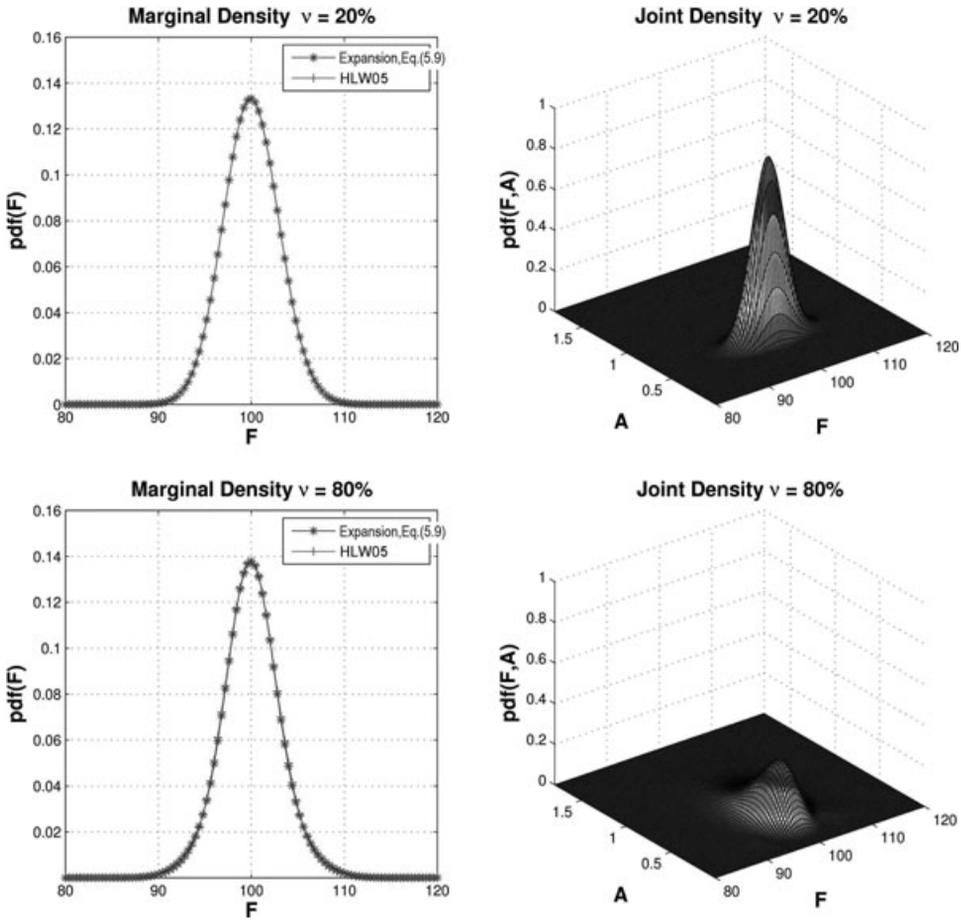


FIGURE 6.2. Marginal densities when volatility-of-volatility ν equals to 20% and 80%. The upper two figures correspond to $\nu = 20\%$ and the lower two figures correspond to $\nu = 80\%$. Other parameters are $f = 100$, $\alpha = 60\%$, $\beta = 0.5$, $\rho = 0$, $T - t = 3$ months.

Results are illustrated in Figures 6.1–6.3 in which HLW05 refers to results from Hagan, Lesniewski, and Woodward (2005). Figure 6.1 reports comparisons when the correlation parameter ρ change signs from -0.7 to 0.7 . Other parameters are fixed at $f = 100$, $\alpha = 60\%$, $\beta = 0.5$, $\nu = 0.4$, $T - t = 3$ months. Figure 6.2 reports comparisons when volatility-of-volatility ν increases from 20% to 80%. Other parameters are fixed at $f = 100$, $\alpha = 60\%$, $\beta = 0.5$, $\rho = 0$, $T - t = 3$ months. Figure 6.3 corresponds to time to maturities $T - t$ at 1 year and 5 years with $f = 100$, $\alpha = 20\%$, $\beta = 0.9999$, $\rho = -0.3$, $\nu = 10\%$.

The truncated domain for numerical evaluation in Figures 6.1 and 6.2 is $\Omega(F, A) \in [0.8f, 1.2f] \times [0.0001\alpha, 3\alpha]$ with 2001×2001 spatial grids for the joint density in (5.9) and $\Omega(F) \in [0.8f, 1.2f]$ with 2001 spatial grids for the marginal density from Hagan, Lesniewski, and Woodward (2005). For Figure 6.3, the same number of spatial grids are used and the truncated domain is $[0.0001f, 2.5f] \times [0.0001a, 2a]$ for the joint density in (5.9) and $[0.0001f, 2.5f]$ for the marginal density from Hagan, Lesniewski, and Woodward (2005).

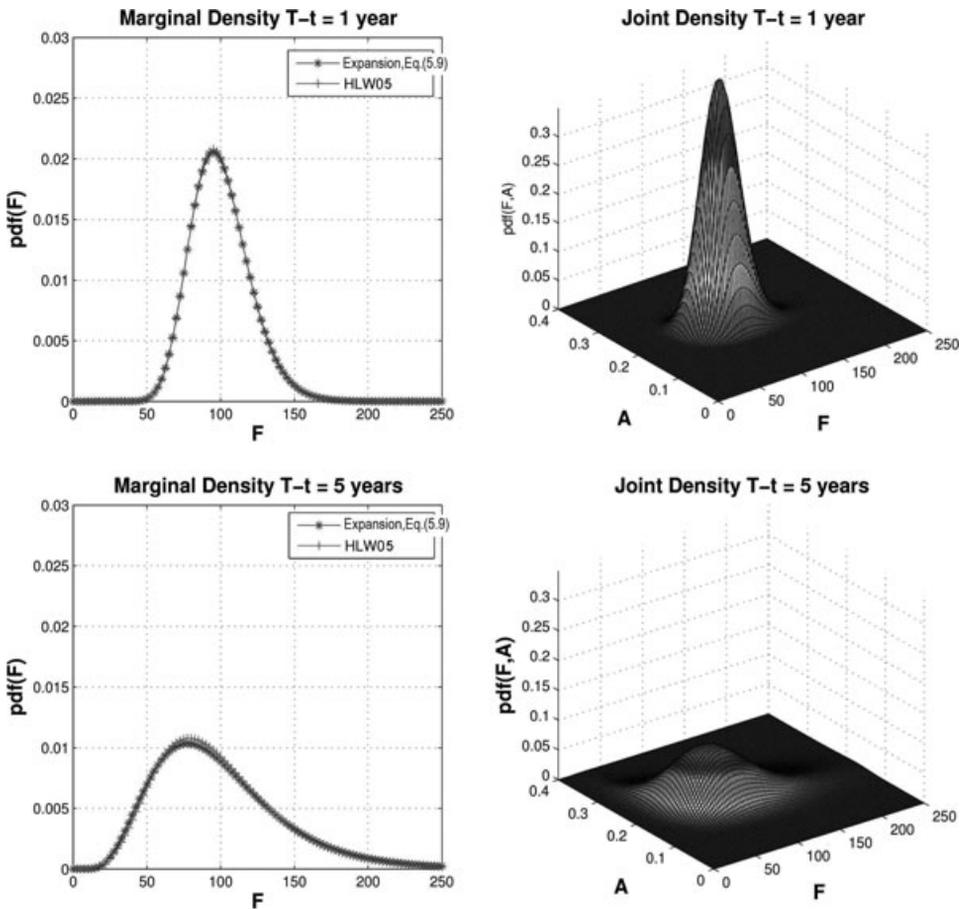


FIGURE 6.3. Marginal densities for time to maturity $T - t$ at 1 year and 5 years. The upper two figures correspond to $T - t = 1$ year and the lower two figures correspond to $T - t = 5$ years. Other parameters are $f = 100$, $\alpha = 20\%$, $\beta = 0.9999$, $\rho = -0.3$, $\nu = 10\%$.

Through Figures 6.1–6.3, our results match very well with Hagan, Lesniewski, and Woodward (2005) for different values of correlation ρ , volatility-of-volatility ν , and time to maturity $T - t$. What is particularly noticeable in Figures 6.1 and 6.2 is that when ρ and ν change, the drastic changes in the shape of the joint density are not obvious simply by looking at the changes in the marginal densities. Further, the correlation parameter ρ has a rotating effect on the joint density while increasing the value of volatility-of-volatility ν spreads out the joint density. This is an important feature when pricing derivatives with forward-starting and barrier features. Also for time to maturities as long as $T = 5$ years, marginal densities generated from both Hagan, Lesniewski, and Woodward (2005) and our equation (5.9) agree well for reasonable values of model parameters so that the perturbation parameter ε is smaller than 1.

6.3. Probability Mass

The probability mass of a density approximation is an important measure to gauge the accuracy of results obtained from expansions. We report total probability mass in the

second-order expansion formula (5.9) as $(T - t, f, \alpha, \beta, \rho, \nu)$ changes. For a comparison with the marginal formula in Hagan, Lesniewski, and Woodward (2005), we also include results obtained from equation (90) in Hagan, Lesniewski, and Woodward (2005). In terms of marginal density, the probability mass is numerically calculated as

$$\int_0^\infty p_F(T - t, f, \alpha, F) dF \approx \sum_{n=1}^{N_F} p_F(T - t, f, \alpha, F_n) \Delta F$$

and for joint density, the mass is

$$\int_0^\infty \int_0^\infty p(T - t, f, \alpha, F, A) dF dA \approx \sum_{i=1}^{N_F} \sum_{j=1}^{N_A} p(T - t, f, \alpha, F_i, A_j) \Delta F \Delta A.$$

The spatial support for $p_F(T - t, f, \alpha, F)$ and $p(T - t, f, \alpha, F, A)$ is \mathbb{R}_+ and \mathbb{R}_+^2 respectively, which are truncated for numerical integration such that the absolute values of densities outside the truncation are smaller than 10^{-6} . As different values of model parameters will result in drastically different distributions, so are the domains we choose for truncation to match the precision at the fixed number of grid points at $N_F = 2001$ and $N_A = 2001$.

Results are reported in Tables 6.2–6.4 in which HLW05 refers to results obtained from Hagan, Lesniewski, and Woodward (2005). Probability mass larger than 105% and smaller than 95% are shaded. In Table 6.2, we report results when time to maturity $T - t$ ranges from 1 month to 12 months and volatility-of-volatility ν ranges from 10% to 100%. In Table 6.3, we vary correlation ρ from -0.9 to 0.9 and volatility-of-volatility ν from 10% to 100%. Finally in Table 6.4, results are reported for different values of current underlying price f at 0.1, 1, 100 and current volatility α at 0.01, 0.1, 0.2, 0.3, 0.5.

What is remarkable to notice is that results from both Hagan, Lesniewski, and Woodward (2005) and equation (5.9) preserve probability mass very well across the wide ranges of parameters tested, except at the small forward case which corresponds to the numerical experiments at $f = 0.1$ for $\beta = 0.5$ and $\beta = 0.0001$ in Table 6.4 where most of the shaded values occurred. In the zero forward case we tested, the probability masses are very concentrated around the current underlying price f for both joint densities and marginal densities. However, one should note that the densities we tested throughout Tables 6.1–6.4 for equation (90) in Hagan, Lesniewski, and Woodward (2005) and for equation (5.9) in this paper are finite order asymptotics under free-boundary conditions, mass-losing at zero forward is a expected phenomenon. Refer to Hagan, Lesniewski, and Woodward (2005) for a detailed discussion of how to relate joint densities under various boundary conditions to the solution from free-boundary condition.

6.4. Implied Volatility

As the SABR model is widely used to fit implied volatility curves in interest rate derivative market, we report results on two typical cases: futures options on Libor rate where the underlying is the forward Libor rate quoted on $100(1 - r_{\text{Libor}})$ and European swaptions with the underlying quoted on r_{Libor} . Technically, futures options correspond to a large forward level case and and European swaptions correspond to a small forward case.

TABLE 6.2
 Probability Mass of Second-Order Expansion in Equation (5.9) (Left Column) and Hagan, Lesniewski, and Woodward (2005) (Right Column) and When Time to Maturity $T - t$ and Volatility-of-Volatility ν Changes

Other Parameters: $f = 100, \alpha = 0.1, \rho = 0$											
Equation (5.9)	$\beta = 0.9999$					HLW05	$\beta = 0.9999$				
$T - t \setminus \nu$	10%	20%	40%	80%	100%	$T - t \setminus \nu$	10%	20%	40%	80%	100%
1 month	1.00	1.00	1.00	1.00	1.00	1 month	1.00	1.00	1.00	1.00	1.00
6 months	1.00	1.00	1.00	1.00	1.00	6 months	1.00	1.00	1.00	1.00	1.00
12 months	1.00	1.00	1.00	1.00	1.01	12 months	1.00	1.00	1.00	1.01	1.01
Other Parameters: $f = 100, \alpha = 0.1, \rho = 0$											
Equation (5.9)	$\beta = 0.5$					HLW05	$\beta = 0.5$				
$T - t \setminus \nu$	10%	20%	40%	80%	100%	$T - t \setminus \nu$	10%	20%	40%	80%	100%
1 month	1.00	1.00	1.00	1.00	1.00	1 month	0.99	0.99	0.99	0.99	0.99
6 months	1.00	1.00	1.00	1.00	1.00	6 months	0.99	0.99	0.99	0.99	0.99
12 months	1.00	1.00	1.00	1.00	1.01	12 months	1.00	0.99	0.99	1.00	1.00
Other Parameters: $f = 100, \alpha = 0.1, \rho = 0$											
Equation (5.9)	$\beta = 0.0001$					HLW05	$\beta = 0.0001$				
$T - t \setminus \nu$	10%	20%	40%	80%	100%	$T - t \setminus \nu$	10%	20%	40%	80%	100%
1 month	1.00	1.00	1.00	1.00	1.00	1 month	0.99	0.99	0.99	0.99	0.97
6 months	1.00	1.00	1.00	1.00	1.00	6 months	0.99	0.99	0.99	0.98	0.99
12 months	1.00	1.00	1.00	1.00	1.01	12 months	1.00	0.99	0.99	0.99	0.99

Comparisons from Monte Carlo simulation, Hagan et al. (2002), and equation (5.9) are plotted in Figure 6.4 in which HKLW02 refers to results from Hagan et al. (2002). The left figure corresponds to the swaption case with $f = 8\%$, $\alpha = 20\%$, $\beta = 0.9999$, $\rho = 0$, $\nu = 20\%$, $T - t = 1$ year. The right figure corresponds to the option case where we set $f = 95$, $\alpha = 80\%$, $\beta = 0.5$, $\rho = 0$, $\nu = 30\%$, $T - t = 1$ year. For Monte Carlo simulation, we generate one million SABR paths with 100 time steps per day using Euler discretization. To use our expansion result, we first obtain option prices by numerically integrating the second-order joint density (5.9) against the payoff of a European call on a truncated domain with 2001×2001 spacial grids and then invert the resulting option prices to implied volatilities. For large forward case $f = 95$, $\alpha = 80\%$, the truncated domain for integration is $(F, A) \in [0.0001f, 2f] \times [0.0001\alpha, 4\alpha]$. For small forward case $f = 8\%$, $\alpha = 20\%$, it is $(F, A) \in [0.001f, 3f] \times [0.001\alpha, 3\alpha]$. And finally to compare with Hagan et al. (2002), we take the implied volatility formula directly and plot the results against those obtained from Monte Carlo simulation and equation (5.9).

It should be stressed that the implied volatilities HKLW02 are provided by a closed-form expression, whereas our results require numerical integration and inversion of the Black formula. This is an important advantage of Hagan et al. (2002) and a reason for the

TABLE 6.3
 Probability Mass of Second-Order Expansion in Equation (5.9) (Left Column) and Hagan, Lesniewski, and Woodward (2005) (Right Column) and When Correlation ρ and Volatility-of-Volatility ν Changes

Other Parameters: $f = 100, \alpha = 0.1, T - t = 0.5$ year												
Equation (5.9)	$\beta = 0.9999$					HLW05	$\beta = 0.9999$					
	$\rho \setminus \nu$	10%	20%	40%	80%		100%	$\rho \setminus \nu$	10%	20%	40%	80%
-0.9	1.00	1.00	1.00	1.00	0.99	-0.9	1.00	1.00	1.00	1.00	1.00	1.00
-0.3	1.00	1.00	1.00	1.00	0.99	-0.3	1.00	1.00	1.00	1.00	1.00	1.00
0.0	1.00	1.00	1.00	1.00	0.99	0.0	1.00	1.00	1.00	1.00	1.00	1.00
0.3	1.00	1.00	1.00	1.00	0.99	0.3	1.00	1.00	1.00	1.00	1.00	1.00
0.9	1.00	1.00	1.00	1.00	0.99	0.9	1.00	1.00	1.00	1.00	1.00	1.00

Other Parameters: $f = 100, \alpha = 0.1, T - t = 0.5$ year												
Equation (5.9)	$\beta = 0.5$					HLW05	$\beta = 0.5$					
	$\rho \setminus \nu$	10%	20%	40%	80%		100%	$\rho \setminus \nu$	10%	20%	40%	80%
-0.9	1.00	1.00	1.00	1.00	1.00	-0.9	1.00	1.00	1.00	1.00	1.00	1.00
-0.3	1.00	1.00	1.00	1.00	1.00	-0.3	1.00	1.00	1.00	1.00	1.00	1.00
0.0	1.00	1.00	1.00	1.00	1.00	0.0	1.00	1.00	1.00	1.00	1.00	1.00
0.3	1.00	1.00	1.00	1.00	1.00	0.3	1.00	1.00	1.00	1.00	1.00	1.00
0.9	1.00	1.00	1.00	1.00	1.00	0.9	1.00	1.00	1.00	1.00	1.00	1.00

Other Parameters: $f = 100, \alpha = 0.1, T - t = 0.5$ year												
Equation (5.9)	$\beta = 0.0001$					HLW05	$\beta = 0.0001$					
	$\rho \setminus \nu$	10%	20%	40%	80%		100%	$\rho \setminus \nu$	10%	20%	40%	80%
-0.9	1.00	1.00	1.00	1.00	1.00	-0.9	1.00	1.00	1.00	1.00	1.00	1.00
-0.3	1.00	1.00	1.00	1.00	1.00	-0.3	1.00	1.00	1.00	1.00	1.00	1.00
0.0	1.00	1.00	1.00	1.00	1.00	0.0	1.00	1.00	1.00	1.00	1.00	1.00
0.3	1.00	1.00	1.00	1.00	1.00	0.3	1.00	1.00	1.00	1.00	1.00	1.00
0.9	1.00	1.00	1.00	1.00	1.00	0.9	1.00	1.00	1.00	1.00	1.00	1.00

success of the SABR formula. It is also the reason why we choose it as the benchmark. One could, in principle, apply numerical integration to either the joint density expression (equation (83)) or the marginal density expression (equation (90)) in Hagan, Lesniewski, and Woodward (2005) to calculate option prices and implied volatilities. We perceive the advantage of our joint density representation as more amenable to numerical integration in the sense that it is explicitly expressed in terms of current state variables f, α , future state variables F, A , and model parameters β, ρ, ν as well as the time-to-maturity $T - t$.

In both cases, equation (5.9) agrees well with Monte Carlo simulation and Hagan et al. (2002) across strikes around at-the-money region. Given the fact that implied volatility is a sensitive measure of both option strikes and model parameters, this is remarkable given the fact that both the joint density result in equation (5.9) and the implied volatility

TABLE 6.4
 Probability Mass of Second-Order Expansion in Equation (5.9) (Left Column) and Hagan, Lesniewski, and Woodward (2005) (Right Column) and When Current Forward Level f and Current Volatility α Changes. Numbers Greater Than 105% and Smaller Than 95% Probability Mass are Shaded

Other Parameters: $\rho = 0, \nu = 0.1, T - t = 1$ year											
Equation (5.9) $f \backslash \alpha$	$\beta = 0.9999$					HLW05 $f \backslash \alpha$	$\beta = 0.9999$				
	0.01	0.1	0.2	0.3	0.5		0.01	0.1	0.2	0.3	0.5
0.1	1.00	1.00	1.00	1.00	1.00	0.1	1.00	1.00	1.00	1.00	1.00
1	1.00	1.00	1.00	1.00	1.00	1	1.00	1.00	1.00	1.00	1.00
100	1.00	1.00	1.00	1.00	1.00	100	0.99	1.00	1.00	1.00	1.00

Other Parameters: $\rho = 0, \nu = 0.1, T - t = 1$ year											
Equation (5.9) $f \backslash \alpha$	$\beta = 0.5$					HLW05 $f \backslash \alpha$	$\beta = 0.5$				
	0.01	0.1	0.2	0.3	0.5		0.01	0.1	0.2	0.3	0.5
0.1	1.00	1.00	0.96	0.72	0.51	0.1	1.00	1.00	0.99	0.96	0.70
1	1.00	1.00	1.00	1.00	0.99	1	1.00	1.00	1.00	1.00	1.00
100	1.00	1.00	1.00	1.00	1.00	100	1.00	1.00	0.99	0.99	0.99

Other Parameters: $\rho = 0, \nu = 0.1, T - t = 1$ year											
Equation (5.9) $f \backslash \alpha$	$\beta = 0.0001$					HLW05 $f \backslash \alpha$	$\beta = 0.0001$				
	0.01	0.1	0.2	0.3	0.5		0.01	0.1	0.2	0.3	0.5
0.1	1.00	0.84	0.69	0.63	0.57	0.1	1.00	0.84	0.63	0.63	0.58
1	1.00	1.00	1.00	1.00	0.98	1	1.00	1.00	1.00	1.00	0.98
100	1.00	1.00	1.00	1.00	1.00	100	1.00	0.99	0.99	0.99	1.00

formula in Hagan et al. (2002) are obtained from small-time asymptotics which often fail in the tails of a model's probability distribution with finite-order terms. For the two sets of parameters reported, it is also interesting to notice that at small strikes, implied volatilities obtained from equation (5.9) agree well with those from Monte Carlo simulation while at large strikes, implied volatilities from Hagan et al. (2002) agree well with those from with Monte Carlo simulation. Work on the extreme strike behavior under the SABR model can be found in Morini and Mercurio (2006) and Obloj (2007). For general results of implied volatility at extreme strikes, please refer to Lee (2004) and Benaim, Friz, and Lee (2008).

7. CONCLUDING REMARKS

In conclusion, we have constructed a converging expansion series for the joint transition density of the SABR model and derived explicit results up to the first three orders. An existence result is then established to characterize the size of a finite order approximation.

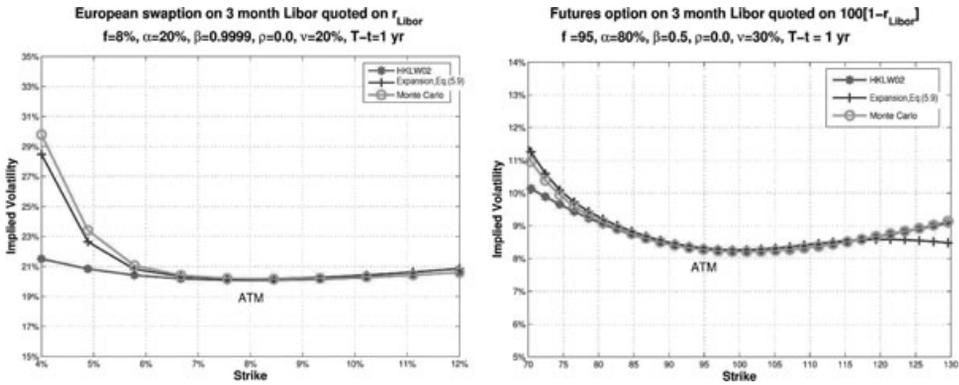


FIGURE 6.4. Implied volatilities. The left figure corresponds to European swaption on 3 month Libor rate quoted on $100(1 - r_{\text{Libor}})$ with $f = 8\%$, $\alpha = 20\%$, $\beta = 0.9999$, $\rho = 0$, $\nu = 20\%$, $T - t = 1$ year. The right figure corresponds to future option on 3 month Libor rate quoted on r_{Libor} with $f = 95$, $\alpha = 80\%$, $\beta = 0.5$, $\rho = 0$, $\nu = 30\%$, $T - t = 1$ year.

We finally test the accuracy of our results in terms of joint transition density, marginal transition density, probability mass, and implied volatility for European call options.

APPENDIX

In the following, we precise the difference between the notion of “standardization” and “decoupling” for general multidimensional diffusion process driven by Brownian motion and use only Itô’s lemma to show why the two-dimensional SABR process can neither be standardized or decoupled.

Standardization versus Decoupling. Consider a driftless vector process X_t^1, \dots, X_t^n driven by n -dimensional Brownian motion W_t^1, \dots, W_t^n with each of the component process given in terms of SDE as

$$dX_t^i = f_i(X_t^1, \dots, X_t^n) dW_t^i, \quad i = 1, 2, \dots, n,$$

then we find a set of change of variables $\psi_i(X_t^1, \dots, X_t^n)$ for every component i such that the resulting process $Y_t^i := \psi_i(X_t^1, \dots, X_t^n)$ has constant diffusion coefficient

$$dY_t^i = \mu^i(t, Y_t^i) dt + \sum_{j=1}^n c_{ij} dW_t^j, \quad i = 1, \dots, n,$$

where $\{c_{ij}\}_{j=1}^n$ are real-valued constants and $\mu^i(t, Y_t^i)$ is the resulting drift term from the transform $\psi_i(X_t^1, \dots, X_t^n)$. We call this operator a transformation to near Gaussian in the sense that with this set of change of variables, if they exist and further are invertible, the original process is transformed into one that is composed of linear combination of the driving Brownian motions. The term “near” is to emphasis the fact that the resulting drift term may be nontrivial.

We did not use the more common terms “standardization” and “decoupling” because there are subtle differences between them. More importantly, the SABR process cannot

be either standardized or decoupled which we will see later. A strict “standardization” is to seek the change of variables ψ_i so that

$$dY_t^i = \sum_{i=1}^n c_i dW_t^i$$

while a strict “decoupling” means we would like Y_t^i to be only function of W_t^i , i.e.,

$$dY_t^i = \mu_i(t, Y_t^i) + f_i(Y_t^i) dW_t^i.$$

The term “decoupling” emphasizes the fact that individual component process, after change of variables, does not enter into other dimensions while “standardization” emphasizes that the resulting component process is a linear combination of Brownian motions. However, as we will see later, neither of the above transformations can be obtained for the SABR process, which then motivates us to see a weaker form of transformation as in our notion of “near Gaussian deformation.”

If both standardization and decoupling can be achieved for the two-dimensional coupled SABR process $\hat{F}_t, \hat{\alpha}_t$ in the SABR model

$$d\hat{F}_t = \hat{\alpha}_t \hat{F}_t^\beta dW_t^1; \quad d\hat{\alpha}_t = \nu \hat{\alpha}_t dW_t^2, \quad \hat{F}_0, \hat{\alpha}_0, \nu > 0$$

then the change of variables $\phi(\hat{F}_t, \hat{\alpha}_t)$ and $\psi(\hat{F}_t, \hat{\alpha}_t)$ that we are seeking will result in the processes $U_t := \phi(\hat{F}_t, \hat{\alpha}_t)$ and $V_t := \psi(\hat{F}_t, \hat{\alpha}_t)$ of the form

$$dU_t = c_u dW_t^1; \quad dV_t = c_v dW_t^2, \quad c_u, c_v \in \mathbb{R}.$$

To answer the question that whether invertible functions $\phi(\cdot), \psi(\cdot)$ exist or not and further, if they exist, what is their explicit form, we use Itô’s lemma to derive a system of equations that $\phi(\cdot)$ and $\psi(\cdot)$ should satisfy.

Assume that $\phi(\cdot), \psi(\cdot)$ satisfy the required regularity conditions, by Itô’s lemma,

$$\begin{aligned} d\phi(\hat{F}_t, \hat{\alpha}_t) &= \frac{\partial \phi}{\partial \hat{F}_t} d\hat{F}_t + \frac{\partial \phi}{\partial \hat{\alpha}_t} d\hat{\alpha}_t + \frac{1}{2} \left(\frac{\partial^2 \phi}{\partial \hat{F}_t^2} d\hat{F}_t^2 + 2 \frac{\partial^2 \phi}{\partial \hat{F}_t \partial \hat{\alpha}_t} d\hat{F}_t d\hat{\alpha}_t + \frac{\partial^2 \phi}{\partial \hat{\alpha}_t^2} d\hat{\alpha}_t^2 \right) \\ &= \frac{1}{2} \left[a(\hat{F}_t, \hat{\alpha}_t) \frac{\partial^2 \phi}{\partial \hat{F}_t^2} + 2b(\hat{F}_t, \hat{\alpha}_t) \frac{\partial^2 \phi}{\partial \hat{F}_t \partial \hat{\alpha}_t} + c(\hat{F}_t, \hat{\alpha}_t) \frac{\partial^2 \phi}{\partial \hat{\alpha}_t^2} \right] dt \\ &\quad + \left[a(\hat{F}_t, \hat{\alpha}_t)^{\frac{1}{2}} \frac{\partial \phi}{\partial \hat{F}_t} \right] dW_t^1 + \left[c(\hat{F}_t, \hat{\alpha}_t)^{\frac{1}{2}} \frac{\partial \phi}{\partial \hat{\alpha}_t} \right] dW_t^2, \end{aligned}$$

where $a(\hat{F}_t, \hat{\alpha}_t), b(\hat{F}_t, \hat{\alpha}_t), c(\hat{F}_t, \hat{\alpha}_t)$ are entries of diffusion matrix of original SABR process as given by

$$a(\hat{F}_t, \hat{\alpha}_t) = \hat{\alpha}_t^2 \hat{F}_t^{2\beta}, \quad b(\hat{F}_t, \hat{\alpha}_t) = \rho \nu \hat{\alpha}_t^2 \hat{F}_t^\beta, \quad c(\hat{F}_t, \hat{\alpha}_t) = \nu^2 \hat{\alpha}_t^2$$

and $\psi(\cdot)$ has the same form as that of $\phi(\cdot)$

$$\begin{aligned} d\psi(\hat{F}_t, \hat{\alpha}_t) &= \frac{1}{2} \left[a(\hat{F}_t, \hat{\alpha}_t) \frac{\partial^2 \psi}{\partial \hat{F}_t^2} + 2b(\hat{F}_t, \hat{\alpha}_t) \frac{\partial^2 \psi}{\partial \hat{F}_t \partial \hat{\alpha}_t} + c(\hat{F}_t, \hat{\alpha}_t) \frac{\partial^2 \psi}{\partial \hat{\alpha}_t^2} \right] dt \\ &\quad + \left[a(\hat{F}_t, \hat{\alpha}_t)^{\frac{1}{2}} \frac{\partial \psi}{\partial \hat{F}_t} \right] dW_t^1 + \left[c(\hat{F}_t, \hat{\alpha}_t)^{\frac{1}{2}} \frac{\partial \psi}{\partial \hat{\alpha}_t} \right] dW_t^2. \end{aligned}$$

To transform $\hat{F}_t, \hat{\alpha}_t$ into U_t, V_t , it is to require that $\phi(\cdot), \psi(\cdot)$ satisfy (written as functions for change of variables $\phi(x, y), \psi(x, y)$)

$$yx^\beta \frac{\partial \phi}{\partial x} = c_u, \quad \nu y \frac{\partial \phi}{\partial y} = 0, \quad x^{2\beta} y^2 \frac{\partial^2 \phi}{\partial x^2} + 2\rho \nu x^\beta y^2 \frac{\partial^2 \phi}{\partial x \partial y} + \nu^2 y^2 \frac{\partial^2 \phi}{\partial y^2} = 0$$

$$yx^\beta \frac{\partial \psi}{\partial x} = 0, \quad \nu y \frac{\partial \psi}{\partial y} = c_v, \quad x^{2\beta} y^2 \frac{\partial^2 \psi}{\partial x^2} + 2\rho \nu x^\beta y^2 \frac{\partial^2 \psi}{\partial x \partial y} + \nu^2 y^2 \frac{\partial^2 \psi}{\partial y^2} = 0.$$

If the above six equality constraints satisfy simultaneously, we will have $\hat{F}_t, \hat{\alpha}_t$ transformed into U_t, V_t . However, we will see in the following this is not possible.

Transformation of $\hat{\alpha}_t$. $\nu y \frac{\partial \psi}{\partial y} = c_v$, we have the general form of solution to $\psi(x, y)$, up to a constant, as

$$\psi(x, y) = \frac{c_v}{\nu} \ln y + C_\psi g(x),$$

where C_ψ is an arbitrary constant and $g(x)$ is an arbitrary function of x only. We then have $\frac{\partial \psi}{\partial x} = C_\psi g'(x)$. Now consider $yx^\beta \frac{\partial \psi}{\partial x} = 0$, then following equality needs to be valid for all x, y

$$yx^\beta C_\psi g'(x) = 0, \quad \forall x, y$$

which means $g'(x) \equiv 0$, hence $g(x)$ is constant. Now $\psi(x, y)$ has the form of

$$\psi(x, y) = \frac{c_v}{\nu} \ln y + C'_\psi,$$

where C'_ψ is another constant. Then immediately, we can verify

$$x^{2\beta} y^2 \frac{\partial^2 \psi}{\partial x^2} + 2\rho \nu x^\beta y^2 \frac{\partial^2 \psi}{\partial x \partial y} + \nu^2 y^2 \frac{\partial^2 \psi}{\partial y^2} = 0$$

will not be valid if c_v is nonzero. This means it is not possible to find a change of variable $\psi(\cdot)$ such that the component process $\hat{\alpha}_t$ can be “standardized” into W_t^2 . There will be a nonzero drift term after the change of variable $\psi(\hat{F}_t, \hat{\alpha}_t) = \frac{c_v}{\nu} \ln \alpha_t + C'_\psi$. Hence, we could only arrive at

$$dU_t = \mu(t, U_t) dt + c_v dW_t^2.$$

Transformation of \hat{F}_t . From $yx^\beta \frac{\partial \phi}{\partial x} = c_u$, we have

$$\phi(x, y) = \frac{c_u}{1 - \beta} \frac{x^{1-\beta}}{y} + C_\phi f(y),$$

where C_ϕ is an arbitrary constant and $f(y)$ is an arbitrary function of y only. Plugging it into the second equation $\nu y \frac{\partial \phi}{\partial y} = 0$, we need the following equation to be valid for all x, y

$$\nu y \left[\frac{c_u x^{1-\beta}}{1 - \beta} \frac{-1}{y^{-2}} + C_\phi f'(y) \right] = 0, \quad \forall x, y,$$

thus, we need

$$f'(y) \equiv 0 \quad \text{and} \quad c_u \equiv 0.$$

This means it is not possible to find a change of variable $\phi(\hat{F}_t, \hat{\alpha}_t)$ to “decouple” the component process \hat{F}_t into one that is driven only by one Brownian motion, either W_t^1 or W_t^2 . Then

$$yx^\beta \frac{\partial \phi}{\partial x} = c_u, \quad \nu y \frac{\partial \phi}{\partial y} = 0$$

cannot be satisfied simultaneously. The best we will be able to do is to “standardize” the coefficient in front of dW_t^1 and the resulting V_t will be

$$dV_t = \mu(t, U_t, V_t) dt + c_u dW_t^1 + \sigma(t, U_t, V_t) dW_t^2.$$

There will be nonzero drift term $\mu(\cdot)$ and nonzero diffusion coefficient $\sigma(\cdot)$ for dW_t^2 . Not only we cannot decouple \hat{F}_t , but also we cannot completely standardize it.

Based on the above analysis, our choice is the following that we seek change of variables $\phi(\cdot)$, $\psi(\cdot)$ such that the resulting process U_t has constant diffusion coefficient in W_t^1 and V_t has constant diffusion coefficient in W_t^2 . As the resulting process U_t , V_t is neither strictly standardized nor decoupled with this choice, we will show in the following that only in the limit of $\varepsilon \downarrow$ we will have a standardized process which is a bivariate Brownian motion. Precisely, we will show that transition probability density function $\check{p}_\varepsilon(\tau, x, y)$ of scaled SABR process $\hat{F}_\tau^\varepsilon, \hat{\alpha}_\tau^\varepsilon$ converges to a bivariate normal distribution as $\varepsilon \downarrow 0$ and further $\hat{F}_\tau^\varepsilon, \hat{\alpha}_\tau^\varepsilon$ converges in distribution to a bivariate Brownian motion.

REFERENCES

- ANDERSEN, L., and V. PITERBARG (2007): Moment Explosions in Stochastic Volatility Models, *Finance Stoch.* 11, 29–50.
- ARONSON, D. G., and P. BESALA (1967): Parabolic Equations with Unbounded Coefficients, *J. Different. Equat.* 3, 1–14.
- BENAIM, S., P. FRIZ, and R. LEE (2008): On the Black-Scholes Implied Volatility at Extreme Strikes, *Frontiers in Quantitative Finance: Volatility and Credit Risk Modeling*, New York: Wiley.
- BENHAMOU, E., and O. CROISSAN (2007): Local Time for the SABR Model: Connection with the “Complex” Black Scholes and Applications to CMS and Spread Options, preprint.
- BERESTYCKI, H., J. BUSCA, and I. FLORENT (2004): Computing the Implied Volatility in Stochastic Volatility Models, *Commun. Pure Appl. Math.* LV II, 1352–1373.
- BOURGADE, P., and O. CROISSANT (2005): Heat Kernel Expansion for a Family of Stochastic Volatility Models: Delta-Geometry, preprint.
- CHEN, L. (1986): On the Behavior of Solutions for Large $|x|$ of Parabolic Equations with Unbounded Coefficients, *Tohoku Math. J.* 20, 589–595.
- ETHIER, S. N., and T. G. KURTZ (1986): *Markov Processes—Characterization and Convergence*, New York: Wiley.
- FABES, E. B. (1992): Gaussian Upper Bounds on Fundamental Solutions of Parabolic Equations: The Method of Nash, *Dirichlet Forms*, 1–20.

- FABES, E. B., and D. W. STROOCK (1984): The L^p Integrability of Green's Functions and Fundamental Solutions for Elliptic and Parabolic Equations, *Duke Math. J.* 51(4), 997–1016.
- FABES, E. B., and D. W. STROOCK (1986): A New Proof of Moser's Parabolic Harnack Inequality Using the Old Ideas of Nash, *Arch. Rational Mech. Anal.* 96(4), 327–338.
- FOUQUE, J. P., G. PAPANICOLUOU, and R. SIRCAR (2000): Stochastic Volatility: Calibrating Random Volatility, *Risk*, 13(2), 89–92.
- FOUQUE, J. P., G. PAPANICOLUOU, R. SIRCAR, and K. SOLNA (2003): Multiscale Stochastic Volatility Asymptotics, *SIAM J. Multiscale Model. Simulat.* 2(1), 22–42.
- FOUQUE, J. P., G. PAPANICOLUOU, R. SIRCAR, and K. SOLNA (2004): Timing the Smile, *Wilmott Mag.* March.
- FRIEDMAN, A. (1983): *Partial Differential Equations of the Parabolic Type*, New Jersey: Prentice Hall.
- GATHERAL, J. (2006): *The Volatility Surface: A Practitioner's Guide*, New York: Wiley.
- HAGAN, P., and A. LESNIEWSKI (2006): LIBOR Market Model with SABR Style Stochastic Volatility, preprint.
- HAGAN, P. S., D. KUMAR, A. S. LESNIEWSKI, and D. E. WOODWARD (2002): Managing Smile Risk, *Wilmott Mag.* September, 84–108.
- HAGAN, P. S., A. S. LESNIEWSKI and D. E. WOODWARD (2005): Probability Distribution in the SABR Model of Stochastic Volatility, preprint.
- HENRY-LABORDÈRE, P. (2005): A General Asymptotic Implied Volatility for Stochastic Volatility Models, preprint.
- HENRY-LABORDÈRE, P. (2007): Unifying the BGM and SABR Models: A Short Ride in Hyperbolic Geometry, preprint.
- LEE, R. (2004): The Moment Formula for Implied Volatility at Extreme Strikes, *Math. Finance* 14(3), 469–480.
- LIONS, P.-L., and M. MUSIELA (2005): Correlations and Bounds for Stochastic Volatility Models, *Ann. l'Institut Henri Poincaré (C) Non Linear Anal.* 24(1), 1–16.
- MORINI, M., and F. MERCURIO (2006): A Note on the SABR Model, preprint.
- MORINI, M., and F. MERCURIO (2007): No-Arbitrage Dynamics for a Tractable SABR Term Structure Libor Model, preprint.
- MORTON, K. W., and D. F. MAYERS (2005): *Numerical Solution of Partial Differential Equations*, Cambridge, New York: Cambridge University Press.
- MUSIELA, M., and M. RUTKOWSKI (1997): *Martingales Methods in Financial Modelling*, New York: Springer.
- OBLOJ, J. (2007): Fine-Tune Your Smile: Correction to Hagan et al., preprint.
- OSAJIMA, Y. (2007): The Asymptotic Expansion Formula of Implied Volatility for Dynamic SABR Model and FX Hybrid Model, preprint.
- REBONATO, R. (2007): A Time-Homogenous, SABR-Consistent Extension of the LMM: Calibration and Numerical Results, *Risk*.
- ROGERS, L. C. G., and L. A. M. VERAART (2008): A Stochastic Volatility Alternative to SABR, *J. Appl. Probab.* 45(4), 1071–1085.