Option Pricing with Orthogonal Polynomial Expansions*

Damien Ackerer† Damir Filipović‡

November 23, 2017

Abstract

We derive analytic series representations for European option prices in polynomial stochastic volatility models. This includes the Jacobi, Heston, Stein–Stein, and Hull–White models, for which we provide numerical case studies. We find that our polynomial option price series expansion performs as efficiently and accurately as the Fourier transform based method in the affine case.

Keywords: option pricing, polynomial diffusion models, stochastic volatility, orthogonal polynomials

MSC2010 Classification: 91G20, 91G60, 60H30, 60J60

JEL Classification: C32, G12, G13

1 Introduction

We study the class of stochastic volatility models for asset price log returns dX in which the squared volatility, $d\langle X\rangle/dt$, is given by a polynomial in an underlying univariate polynomial diffusion Y. This class nests many models of interest having bounded and unbounded volatility support, including the Jacobi, Heston, Stein–Stein, and Hull–White models. Such polynomial models are highly tractable because the conditional moments of the marginal distributions of the log price X_T are given in closed form in terms of a matrix exponential. We use this to derive the prices of options with discounted payoff $f(X_T)$ in terms of analytic orthogonal polynomial expansions. Observing that X_T has a density g given by a Gaussian mixture with an infinite number of components, we approximate g by a finite Gaussian mixture w. Specifically, under some technical conditions, we show that the likelihood ratio function $\ell = g/w$ lies in the weighted space L^2_w of square-integrable functions with respect to w. In this case, and when the discounted payoff function f also belongs to L^2_w, the corresponding option price has a series representation of the form $\pi_f = \langle f, \ell\rangle_w = \sum_{n\geq 0} f_n\ell_n$ where the Fourier coefficients f_n and ℓ_n of f and ℓ are with respect to an orthonormal basis of polynomials $\{H_n,\ n \geq 0\}$ of L^2_w. Thanks to the polynomial property of the model, the likelihood coefficients $\ell_n = \mathbb{E}[H_n(X_T)]$ are given in closed form. For important examples, including European call options, the payoff coefficients f_n are given in analytic form. The option price $\pi_f = \sum_{n\geq 0} f_n\ell_n$ is approximated by truncating the series at a finite order N. This option price approximation is accurate for a small truncation order N if the true density g is statistically close to the auxiliary density w. Even when ℓ does not belong to L^2_w, so that the sequence of option price approximations $\sum_{n=0}^N f_n\ell_n$ does not converge as $N \to \infty$, we show that a finite order N can provide an accurate approximation of the true price π_f. We also provide simple and efficient algorithms to construct the auxiliary mixture density w and the corresponding orthonormal polynomials. We derive recursive systems of equations for the payoff coefficients f_n in case of a European call option and for auxiliary Gaussian and Gamma mixture densities.

We validate our method on several use cases. In the Jacobi model with a single auxiliary Gaussian density, the accuracy of the European call option price approximation for a fixed N decreases rapidly as

*The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 307465-POLYTE.

†Swissquote Bank, 1196 Gland, Switzerland. *Email:* damien.ackerer@swissquote.ch

‡EPFL and Swiss Finance Institute, 1015 Lausanne, Switzerland. *Email:* damir.filipovic@epfl.ch

the upper bound of the volatility support increases. We therefore let the auxiliary density w be a mixture of two Gaussian distributions whose first two moments are matching the log price density. We show that the option price series converges significantly faster using this Gaussian mixture density in comparison to a single Gaussian density. This allows to approximate option prices rapidly and accurately for any parameters choice. We hence calibrate the Jacobi model on a volatility surface sample and report a twice smaller implied volatility root mean squared error than with the Heston model. We also approximate European call option prices in the Stein–Stein model using a Gaussian mixture with different number of components and parameters choices, and benchmark them to the option prices computed using Fourier transform techniques. We also price call options in the Hull–White model, which is shown to induce large log price kurtosis values. Despite the fact that the likelihood ratio function ℓ in these models does not belong to L^2_w when w is a Gaussian mixture density, we find that our approach still produces accurate option price approximations. We also price call options on the realized variance in the GARCH diffusion model.

Stochastic volatility models for asset returns are necessary to reproduce realistically the dynamical behavior of asset prices. Such models were first introduced by (Hull and White 1987), (Scott 1987), (Wiggins 1987) to overcome the assumption of constant volatility in the seminal works by (Black and Scholes 1973), and (Merton 1973). However, in the context of derivatives, a stochastic volatility model is practical only if tractable formulas are available to approximate option prices. The (Heston 1993) model therefore became a benchmark because it has a quasi-analytical formula to price European options. The Heston model belongs to the large class of affine models whose development lead to many new tractable models, see (Duffie, Filipović, and Schachermayer 2003). Indeed, the characteristic function of the log price can be computed numerically in affine models which enables the use of standard Fourier transform techniques to compute European option prices, see for details (Carr and Madan 1999; Fang and Oosterlee 2009; Lipton and Sepp 2008). Our paper adds to this literature by extending the range of tractable models beyond the affine class. The literature on option pricing in non-affine polynomial models contain (Ackerer, Filipović, and Pulido 2016) for the Jacobi model or (Filipović, Gourier, and Mancini 2016) for variance swaps. Our results improve the option pricing performance in (Ackerer, Filipović, and Pulido 2016). Notable progress has recently been made on the pricing of early-exercise and path-dependent options with stochastic volatility using recursive marginal quantization as studied in (Pagès and Sagna 2015; McWalter, Rudd, Kienitz, and Platen 2017; Callegaro, Fiorin, and Grasselli 2017a). In the context of affine and polynomial models, this approach has been shown to perform well when combined with Fourier transform techniques as in (Callegaro, Fiorin, and Grasselli 2017b), or polynomial expansion techniques as in (Callegaro, Fiorin, and Pallavicini 2017), whose results could be further improved with the new expansions presented in our paper.

The remainder of the paper is as follows. Section 2 discusses the density expansion and option price series representation with an auxiliary mixture density. Section 3 introduces the polynomial stochastic volatility models and describes some Gaussian mixture constructions for the auxiliary density. Section 4 contains the numerical applications. Section 5 concludes. The proofs are collected in Appendix A. Alternative moment based methods to construct orthonormal polynomial basis can be found in Appendix B.

2 Polynomial Price Series Expansions

We recap the density expansion approach described in (Filipović, Mayerhofer, and Schneider 2013) along with the option price series representation further developed in (Ackerer, Filipović, and Pulido 2016). We show how the elements of these option price series can be efficiently computed when the auxiliary density is a mixture density. We then give some examples of component densities which are convenient to work with.

2.1 European Option Price Series Representation

Fix a maturity $T > 0$ and assume that the distribution of the log price X_T of an asset has a density $g(x)$. Our goal is to compute the option price

$$\pi_f = \int_{\mathbb{R}} f(x)g(x)dx \tag{1}$$

for some discounted payoff function $f(x)$. Let $w(x)$ be an auxiliary density that dominates $g(x)$ with likelihood ratio $\ell(x)$ given by

$$g(x) = \ell(x)w(x).$$

We define the weighted Lebesgue space,

$$L^2_w = \left\{ f(x) \mid \|f\|^2_w = \int_{\mathbb{R}} f(x)^2 w(x)dx < \infty \right\},$$

which is a Hilbert space with the scalar product

$$\langle f, g \rangle_w = \int_{\mathbb{R}} f(x)g(x)w(x)dx.$$

Assume that the polynomials are dense in L^2_w and that the functions $\ell(x)$ and $f(x)$ are in L^2_w. Then, the option price has the following series representation

$$\pi_f = \sum_{n \geq 0} f_n \ell_n \tag{2}$$

with *likelihood coefficients*

$$\ell_n = \langle \ell, H_n \rangle_w = \int_{\mathbb{R}} H_n(x)g(x)dx \tag{3}$$

and *payoff coefficients*

$$f_n = \langle f, H_n \rangle_w = \int_{\mathbb{R}} f(x)H_n(x)w(x)dx, \tag{4}$$

and where $H_0 = 1, H_1, H_2, \ldots,$ denotes an orthonormal polynomial basis of L^2_w such that $\deg H_n(x) = n$ and $\langle H_m, H_n \rangle_w = 1$ if $m = n$ and zero otherwise. In practice we approximate the option price by truncating the option price series in (2) as follows

$$\pi^{(N)}_f = \sum_{n=0}^{N} f_n \ell_n \tag{5}$$

for some positive integer N. The accuracy of this approximation crucially depends on the statistical distance between the true density and the auxiliary density. Indeed, as $\ell_0 = 1$, the L^2-divergence of g from w equals $\|\ell - 1\|^2_w = \sum_{n \geq 1} \ell^2_n$. For example if $w(x) = g(x)$ then $\pi_f = f_0 \ell_0 = f_0$, as $\ell_n = 0$ for all $n \geq 1$ in this case. The efficiency of this approach also depends on how fast the operations required to compute the coefficients in (3) and (4) can be performed.

2.2 Auxiliary Mixture Density

We now assume that the auxiliary density $w(x)$ is a mixture density of the form

$$w(x) := \sum_{k=1}^{K} c_k v_k(x)$$

for some mixture weights $c_k > 0$ satisfying $\sum_{k=1}^{K} c_k = 1$, and some mixture components $v_k(x)$ which are also probability densities. To each density $v_k(x)$ is associated an orthonormal polynomial basis $H^k_n(x)$, $n \geq 0$. Let a^k_n and b^k_n denote the coefficients that define the three term recurrence relation for $H^k_n(x)$, which always holds for univariate orthonormal polynomial basis, see e.g. (Gautschi 2004, Section 1.3),

$$xH^k_n(x) = b^k_{n+1}H^k_{n+1}(x) + a^k_n H^k_n(x) + b^k_n H^k_{n-1}(x) \tag{6}$$

for all $n \geq 0$ with $H^k_{-1} = 0$ and $H^k_0 = 1$. We define the following tridiagonal matrices

$$J^k_N = \begin{pmatrix} a^k_0 & b^k_1 & & & \\ b^k_1 & a^k_1 & b^k_2 & & \\ & \ddots & \ddots & \ddots & \\ & & b^k_{N-1} & a^k_{N-1} & b^k_N \\ & & & b^k_N & a^k_N \end{pmatrix}, \quad \text{for } k = 1, \ldots, K.$$

The recurrence relation (6) used to construct the orthonormal polynomial basis is explicitly known for many densities taking values both on compact and unbounded supports. When $v_k(x)$ is a Gaussian density with mean μ_k and variance σ_k^2, the coefficients in (6) are given by $a_n^k = \mu_k$ and $b_n^k = \sqrt{n}\sigma_k$. We refer to (Schoutens 2012, Chapter 1) for an overview of orthonormal polynomial basis families.

Remark 2.1. *The coefficients a_n^k and b_n^k can be inferred from the basis $H_n^k(x)$. Denote $\alpha_{n,i}^k$ the coefficient in front of the monomial x^i of the polynomial $H_n^k(x)$. Then by inspection of the recurrence relation* (6) *we have*

$$b_{n+1}^k = \frac{\alpha_{n,n}^k}{\alpha_{n+1,n+1}^k} \quad \textit{and} \quad a_n^k = \frac{\alpha_{n,n-1}^k - b_{n+1}^k \alpha_{n+1,n}^k}{\alpha_{n,n}^k}.$$

The following proposition gives an algorithm to compute the coefficients a_n and b_n in the recursion of the orthonormal polynomial basis $H_n(x)$ associated with the mixture density $w(x)$,

$$xH_n(x) = b_{n+1}H_{n+1}(x) + a_n H_n(x) + b_n H_{n-1}(x)$$

for all $n \geq 0$ with $H_{-1} = 0$ and $H_0 = 1$.

Proposition 2.2. *The recurrence relation coefficients of the mixture density $w(x)$ are given by $b_n = \sqrt{\psi_n/\psi_{n-1}}$ for $n = 1, \ldots, N$ and $a_n = \phi_n/\psi_n$ for $n = 0, \ldots, N-1$ where*

$$\begin{aligned}
\psi_n &= \sum_{k=1}^K c_k \left(z_n^k\right)^\top z_n^k \\
\phi_n &= \sum_{k=1}^K c_k \left(z_n^k\right)^\top J_N^k \, z_n^k \\
z_{n+1}^k &= (J_N^k - a_n I) z_n^k - (b_n)^2 z_{n-1}^k \quad \textit{for all } k = 1, \ldots, K
\end{aligned}$$

with $z_{-1}^k = 0$, and $z_0^k = \mathrm{e}_1$ is the vector whose first coordinate is equal to one and zero otherwise.

This algorithm is fast and performs well numerically. For example, with $N, K \sim 10^2$, it takes few milliseconds on a modern CPU to construct the orthonormal basis. There are other moments based approaches to construct the orthonormal basis, such as the Gram-Schmidt and the Mysovskikh algorithms, which are described in Appendix B. However these methods are typically subject to numerical problems and may be slow even for a relatively small order N.

The following proposition shows that the payoff coefficients with respect to the mixture density can be efficiently computed when the corresponding coefficients are known for the mixture components.

Proposition 2.3. *The payoff coefficients are equal to*

$$f_N = \langle f, H_N \rangle_w = \sum_{k=1}^K \sum_{n=0}^N c_k \, q_{N,n}^k \, f_n^k \quad \textit{with } f_n^k = \langle f, H_n^k \rangle_{v_k} \tag{7}$$

and where $q_N^k \in \mathbb{R}^N$ is the vector representation of $H_N(x)$ in the basis $H_n^k(x)$

$$H_N(x) = \sum_{n=0}^N q_{N,n}^k \, H_n^k(x), \quad k = 1, \ldots, K. \tag{8}$$

The usefulness of Proposition 2.3 lies on the premise that the coefficients f_n^k can be easily computed. In a situation where they are numerically costly to compute then one may directly integrate the payoff function with respect to the following density approximation

$$w^{(N)}(x) = \sum_{n=0}^N \ell_n H_n(x) \sum_{k=1}^K c_k v_k(x). \tag{9}$$

Remark 2.4. *The vectors q_N^k can be efficiently computed. Let $\mathbf{H}_N^k \in \mathbb{R}^{(N+1)\times(N+1)}$ denote the matrix whose (i,j)-th element is given by the coefficient in front of the monomial x^{j-1} in the $(i-1)$-th polynomial of the basis $H_n^k(x)$. Define similarly the matrix $\mathbf{H}_N$ with respect to the polynomial basis $H_n(x)$. The matrices $\mathbf{H}_N^k$ for $k = 1, \dots, K$, and $\mathbf{H}_N$ are upper triangular. We are interested in the upper triangular matrix $\mathbf{Q}_N^k \in \mathbb{R}^{(N+1)\times(N+1)}$ for $k = 1, \dots, K$ whose (i,j)-th element is equal to $q_{j,i}^k$. It is also given by $\mathbf{H}_N = \mathbf{H}_N^k \mathbf{Q}_N^k$ which forms a triangular system of equations and can thus be solved efficiently.*

2.3 Examples of Mixture Components

In practice, to ensure efficient option prices approximations we need to select mixture components $v_k(x)$ whose orthonormal polynomial bases $H_n^k(x)$ and payoff coefficients f_n^k can be efficiently computed. The Gaussian and uniform distributions have been successfully used in (Ackerer, Filipović, and Pulido 2016) and in (Ackerer and Filipović 2016) respectively. The logistic distribution whose tail decreases at an exponential-linear rate is used in (Heston and Rossi 2016). But the corresponding payoff coefficients are given by complicated expressions involving special functions. The bilateral Gamma distribution of (Küchler and Tappe 2008) was used in (Filipović, Mayerhofer, and Schneider 2013) to accurately approximate option prices by numerical integration of the discounted payoff function with respect to the density approximation. However an explicit recursion formula to construct the orthonormal polynomial basis was not provided.

We study hereinbelow the Gamma distribution whose tail decays at a polynomial-exponential-linear rate and for which simple recursive expressions can be derived for the payoff coefficients. By mixing two Gamma distributions one can obtain a distribution on the entire real line. The Gamma density on the half-line $(\xi, +\infty)$ for some $\xi \in \mathbb{R}$ is defined by

$$v_k(x) = 1_{\{x>\xi\}} \frac{\beta^\alpha}{\Gamma(\alpha)} (x-\xi)^{\alpha-1} \mathrm{e}^{-\beta(x-\xi)} \tag{10}$$

for some shape parameter $\alpha \geq 1$, rate parameter $\beta > 0$, and where $\Gamma(\alpha)$ is the upper incomplete Gamma function defined by

$$\Gamma(\alpha) = \Gamma(\alpha, 0) \quad \text{with} \quad \Gamma(\alpha, z) = \int_z^\infty x^{\alpha-1} \mathrm{e}^{-x} dx,$$

so that $\Gamma(n) = (n-1)!$ for any positive integer n. The Gamma density $v_k(x)$ admits an orthonormal polynomial basis $H_n^k(x)$ given by

$$H_n^k(x) = \sqrt{\frac{n!}{\Gamma(\alpha+n)}}\, \mathcal{L}_n^{\alpha-1}(\beta(x-\xi))$$

where $\mathcal{L}_n^{\alpha-1}(x)$ denotes the n-th order generalized Laguerre polynomial with parameter $\alpha - 1$ defined by

$$\mathcal{L}_n^{\alpha-1}(x) = \frac{x^{-\alpha+1}\mathrm{e}^x}{n!} \frac{\partial^n}{\partial x^n} (\mathrm{e}^{-x} x^{\alpha-1+n}).$$

The generalized Laguerre polynomials are recursively given by

$$\begin{aligned}
&\mathcal{L}_0^{\alpha-1}(x) = 1 \\
&\mathcal{L}_1^{\alpha-1}(x) = \alpha + x \\
&\mathcal{L}_{n+1}^{\alpha-1}(x) = \frac{2n+\alpha-x}{n+1} \mathcal{L}_n^{\alpha-1}(x) - \frac{n+\alpha-1}{n+1} \mathcal{L}_{n-1}^{\alpha-1}(x) \quad \text{for all } n \geq 1.
\end{aligned}$$

The following proposition shows that the payoff coefficients f_n^k of two specific discounted payoff functions can be recursively computed for the Gamma distribution $v_k(x)$.

Proposition 2.5. (i) *Consider the discounted payoff function of a call option with log strike k,*

$$f(x) = \mathrm{e}^{-rT} \left(\mathrm{e}^x - \mathrm{e}^k\right)^+.$$

Its payoff coefficients f_n are given by

$$f_n = \sqrt{\frac{n!}{\Gamma(\alpha+n)}}\frac{1}{\Gamma(\alpha)}\left(\mathrm{e}^{\xi} I_n^{\alpha-1}\left(\mu;\beta^{-1}\right) + \mathrm{e}^{k} I_n^{\alpha-1}\left(0;\beta^{-1}\right)\right), \quad n \geq 0$$

with $\mu = \max(0, \beta(k-\xi))$ and where the functions $I_n^{\alpha-1}(\mu;\nu)$ are recursively defined by

$$\begin{aligned} I_0^{\alpha-1}(\mu;\nu) &= (1-\nu)^{-\alpha}\,\Gamma(\alpha)\,\Gamma(\alpha,\mu(1-\nu)) \\ I_1^{\alpha-1}(\mu;\nu) &= \alpha I_0^{\alpha-1}(\mu;\nu) + I_0^{\alpha}(\mu;\nu) \\ I_n^{\alpha-1}(\mu;\nu) &= \left(2+\frac{\alpha-2}{n}\right) I_{n-1}^{\alpha-1}(\mu;\nu) - \left(1+\frac{\alpha-2}{n}\right) I_{n-2}^{\alpha-1}(\mu;\nu) \cdot \\ &\quad - \frac{1}{n}\left(I_{n-1}^{\alpha}(\mu;\nu) - I_{n-2}^{\alpha}(\mu;\nu)\right), \quad n \geq 2. \end{aligned} \tag{11}$$

(ii) *Consider the discounted payoff function with strike k,*

$$f(x) = \mathrm{e}^{-rT}\left(x-k\right)^{+}.$$

Its payoff coefficients f_n are given by

$$f_n = \sum_{j=0}^{n+2} p_{n,j}\,\frac{\Gamma(\alpha+j,\beta(k-\xi))}{\Gamma(\alpha)\beta^j}, \quad n \geq 0$$

where $p_n \in \mathbb{R}^{n+2}$ is the vector representation of $xH_n^k(x+\xi) + (\xi-k)H_n^k(x)$ in the monomial basis.

Note that the calculation of the term $I_n^{\alpha-1}(\mu;\nu)$ in part (i) requires calculations of the terms $I_{n-1}^{\alpha}(\mu;\nu)$ and therefore the dimension of the recursive system grows at the rate n^2. Part (ii) will be used to price realized variance derivatives, see example in Section 4.4.

The Gamma distribution on the half-line $(-\infty,\xi)$ for some $\xi \in \mathbb{R}$ together with its polynomial basis and Fourier coefficients can be similarly derived. A Gamma mixture on $\mathbb{R}$ can thus be constructed when $K \geq 2$ by letting, for example, the first component support be $(-\infty,\xi]$ and the second $[\xi,\infty)$. In particular, the coefficients in (6) when $v_k(x)$ is a Gamma density are given by

$$a_n^k = \frac{\delta 2n + \delta\alpha + \beta\xi}{\beta}, \quad n \geq 0, \quad \text{and} \quad b_n^k = -\frac{\delta\sqrt{\alpha+n}}{\beta}, \quad n \geq 1,$$

where $\delta = 1$ if the domain of the Gamma density is (ξ,∞) and $\delta = -1$ if its domain is $(-\infty,\xi)$.

3 Polynomial Stochastic Volatility Models

We present a class of stochastic volatility models for which the log price density $g(x)$ is given by a Gaussian mixture with an infinite number of components. We then describe a simple method to approximate the log price density by an auxiliary Gaussian mixture $w(x)$ with a finite number of components. The section terminates by studying the option price approximation error when the likelihood ratio $\ell(x) = g(x)/w(x)$ does not belong to the weighted space L_w^2.

3.1 Definition and Basic Properties

We fix a stochastic basis $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbb{Q})$ where $\mathbb{Q}$ is a risk-neutral measure, carrying a multivariate Brownian motion W_t. We first recall the notion and basic properties of a polynomial diffusion, see (Filipović and Larsson 2017). For $d,n \in \mathbb{N}$ and a state space $E \subseteq \mathbb{R}^d$, denote by $\mathrm{Pol}_n(E)$ the linear space of polynomials on E of degree n or less. Consider an E-valued *polynomial* diffusion

$$dZ_t = b(Z_t)\,dt + \sigma(Z_t)\,dW_t$$

in the sense that its generator $\mathcal{G} = b^\top \nabla + (1/2)\mathrm{Tr}(\sigma\sigma^\top \nabla^2)$ maps $\mathrm{Pol}_n(E)$ to $\mathrm{Pol}_n(E)$ for all $n \in \mathbb{N}$. It can easily be verified that this polynomial property holds if and only if $b \in \mathrm{Pol}_1(E)$ and $a \in \mathrm{Pol}_2(E)$ componentwise, see (Filipović and Larsson 2017, Lemma 2.2). Fix a degree n and a basis $q_0, \ldots, q_M$ of $\mathrm{Pol}_n(E)$, where $M+1 = \dim \mathrm{Pol}_n(E)$, and write $Q(z) = (q_0(z), \ldots, q_M(z))^\top$. Then $\mathcal{G}$ restricted to $\mathrm{Pol}_n(E)$ has a matrix representation G, so that $\mathcal{G}p(z) = Q(z)^\top G\vec{p}$ for any polynomial $p \in \mathrm{Pol}_n(E)$ and where $\vec{p}$ denotes its vector representation. A very useful consequence is that the conditional moments of Z_T are given in closed form by the moment formula

$$\mathbb{E}[p(Z_T) \mid \mathcal{F}_t] = Q(Z_t)^\top \mathrm{e}^{G(T-t)}\vec{p}, \tag{12}$$

see (Filipović and Larsson 2017, Theorem 2.4).

Now let Y_t be a univariate polynomial diffusion of the form

$$dY_t = \kappa(\theta - Y_t)\,dt + \sigma(Y_t)\,dW_{1t} \quad \text{with} \quad \sigma(y)^2 = \alpha + ay + Ay^2 \tag{13}$$

for some real parameters $\kappa, \theta, \alpha, a, A$. We then define the dynamics of the log price X_t as follows

$$dX_t = (r-\delta)\,dt - \frac{1}{2}d\langle X\rangle_t + \Sigma_1(Y_t)\,dW_{1t} + \Sigma_2(Y_t)\,dW_{2t}$$

for the interest rate r, the dividend rate $\delta \geq 0$, and such that $\Sigma_1^2 + \Sigma_2^2 \in \mathrm{Pol}_m(\mathbb{R})$ and $\Sigma_1\sigma \in \mathrm{Pol}_{m+1}(\mathbb{R})$ for some $m \in \mathbb{N}$. The initial values $(X_0, Y_0) \in \mathbb{R}^2$ are deterministic. It can be verified that the $\mathbb{R}^{m+1}$-valued diffusion $Z_t = (X_t, Y_t, Y_t^2, \ldots, Y_t^m)$ has the polynomial property on $E = \{(x, y, y^2, \ldots, y^m) \mid (x,y) \in \mathbb{R}^2\}$, see (Filipović and Larsson 2017, Theorem 5.5). Fix the order N, the moment formula (12) implies that the payoff coefficients in (3) can be computed explicitly as follows

$$\ell_n = \mathbb{E}[H_n(X_T)] = Q(Z_0)^\top \mathrm{e}^{G\,T}\vec{H}_n, \quad n = 0, \ldots, N \tag{14}$$

for some chosen polynomials $Q(z) = (q_0(z), \ldots, q_M(z))^\top$ that form a basis of $\mathrm{Pol}_N(E)$. As above, G is the matrix representation of the generator of Z_t and $\vec{H}_n$ is the vector representation of the polynomial $H_n(x)$ in this basis. Some classical stochastic volatility models are nested in this setup, such as the Heston, Jacobi, Stein–Stein, and Hull–White models.

The spot variance of the log return dX_t is $V_t = d\langle X\rangle_t/dt = \Sigma_1(Y_t)^2 + \Sigma_2(Y_t)^2$. The leverage effect refers to the generally negative correlation between dV_t and dX_t and is given by

$$\begin{aligned}\mathrm{lev}(X_t) &= \frac{d\langle V, X\rangle_t}{\sqrt{d\langle V\rangle_t}\,\sqrt{d\langle X\rangle_t}} \\ &= \frac{\Sigma_1(Y_t)}{\sqrt{\Sigma_1(Y_t)^2 + \Sigma_2(Y_t)^2}}\,\mathrm{sign}\left[(\Sigma_1'(Y_t)\Sigma_1(Y_t) + \Sigma_2'(Y_t)\Sigma_2(Y_t))\,\sigma(Y_t)\right].\end{aligned} \tag{15}$$

The volatility of the volatility of the log return is

$$\mathrm{volvol}(X_t) = \sqrt{\frac{d\langle\sqrt{V}\rangle_t}{dt}} = \frac{1}{2\sqrt{V_t}}\sqrt{\frac{d\langle V\rangle_t}{dt}} = \sqrt{\frac{\left(\Sigma_1'(Y_t)\Sigma_1(.Y_t) + \Sigma_2'(Y_t)\Sigma_2(Y_t)\right)^2\sigma(Y_t)^2}{\Sigma_1(Y_t)^2 + \Sigma_2(Y_t)^2}} \tag{16}$$

The proofs of Equations (15) and (16) are given in Appendix A.

We fix a finite time horizon $T > 0$. The following proposition shows that the distribution of the log price is given by a Gaussian mixture density with an infinite number of components, see (McNeil, Frey, and Embrechts 2015, Chapter 6.2).

Proposition 3.1. *The distribution of X_T conditional on the trajectory of W_{1t} on $[0,T]$ is normally distributed with mean*

$$M_T = X_0 + (r-\delta)T - \frac{1}{2}\int_0^T \left(\Sigma_1(Y_t)^2 + \Sigma_2(Y_t)^2\right)dt + \int_0^T \Sigma_1(Y_t)dW_{1t} \tag{17}$$

and variance

$$C_T = \int_0^T \Sigma_2(Y_t)^2 dt. \tag{18}$$

Therefore, when it is well defined, the log price density is of the form

$$g(x) = \mathbb{E}\left[(2\pi C_T)^{-\frac{1}{2}} \exp\left(-\frac{(x-M_T)^2}{2C_T}\right)\right]. \tag{19}$$

Similar expressions for the log price density have been previously derived in (Lipton and Sepp 2008) and in (Glasserman and Kim 2011). In practice, we want to approximate the log price density $g(x)$ by a Gaussian mixture density $w_K(x)$ with K components that will in turn be used as auxiliary density to derive option price approximations,

$$w_K(x) = \sum_{k=1}^{K} c_k \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right) \tag{20}$$

for some mixture weights $c_k > 0$ such that $\sum_{k=1}^{K} c_k = 1$, and some constants $\mu_k \in \mathbb{R}$ and $\sigma_k > 0$ for $k = 1, \ldots, K$. In Section 3.2 we suggest a computationally efficient approach based on weighted simulations of the first Brownian motion W_{1t}.

We now give conditions such the likelihood ratio $\ell(x)$ belongs to the weighted Lebesgue space $L^2_{w_K}$.

Proposition 3.2. *Assume that there exist two constants $C_1, C_2 > 0$ such that*

$$|M_T| < C_1 \text{ and } C_T < C_2 \tag{21}$$

then $\ell \in L^2_{w_K}$ with w_K as defined in (20) if $\sigma_k > C_2/2$ for some k in $1, \ldots, K$. Property (21) holds if Y_t takes values in some compact interval $I \subset \mathbb{R}$ for all $t \geq 0$ and $\Sigma_1(y) = \phi(y)\sigma(y)$ for $y \in I$ for some $\phi \in C^1(I)$.

The Greeks of the option are computed by differentiating the option price with respect to one or multiple variables. For the sensitivity analysis we fix the auxiliary density $w(x)$, hence the basis $H_n(x)$ and the coefficients f_n, and let only $\ell(x)$ through $g(x)$ depend on the perturbed parameters. The sensitivity of π_f with respect to the variable y is hence given by

$$\partial_y \pi_f = \sum_{n \geq 0} (\ell_n \partial_y f_n + f_n \partial_y \ell_n) \tag{22}$$

with the partial derivative $\partial_y = \partial/\partial y$. The sensitivity of ℓ_n with respect to y is given by

$$\partial_y \ell_n = (0, \partial_y Q(Z_0))\, \mathrm{e}^{G\,T}\, \vec{H}_n + (1, Q(Z_0))\, \partial_y \mathrm{e}^{G\,T}\, \vec{H}_n.$$

The derivative of the exponential operator $\mathrm{e}^{G\,T}$ with respect to y is given by

$$\frac{\partial \mathrm{e}^{G\,T}}{\partial y} = \int_0^1 \mathrm{e}^{x\,G\,T}\, \frac{\partial\, G\,T}{\partial y}\, \mathrm{e}^{(1-x)\,G\,T}\, dx$$

as proved in (Wilcox 1967).

In particular for the Delta, which is the derivative of option price with respect to $y = \exp(X_0)$, we have that $\partial_y \mathrm{e}^{G\,T} = 0$. In practice, we can efficiently approximate the option greeks by truncation the series (22) at some finite order N. This may also prove to be valuable for the model calibration or estimation since the objective function derivatives with respect to the model parameters can be computed which enables the use of gradient based search algorithms.

3.2 Gaussian Mixture Specification

We present an efficient approach to specify Gaussian mixtures for the auxiliary density. It is based on the discretization of the single source of randomness affecting M_T and C_T in (17) and (18): the trajectory of W_1 on $[0, T]$. Fix a time grid $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$ with constant step size $t_{i+1} - t_i = \Delta t$. Then, apply a Monte-Carlo approach to obtain K weighted $\mathbb{R}^n$-valued vectors $Z^{(k)}$ of normally distributed Brownian increments

$$Z^{(k)} \sim (\Delta W_{1,t_1}, \ldots, \Delta W_{1,t_n}) \sim \mathcal{N}\left(\mathbf{0}_n,\, \mathrm{diag}(\Delta t\, \mathbf{1}_n)\right) \tag{23}$$

where $\Delta W_{1,t_i} = W_{1,t_i} - W_{1,t_{i-1}}$. The stochastic differential equations satisfied by (Y_t, M_t, C_t) can then be numerically integrated to obtain K triplets $(c_k, M_T^{(k)}, C_T^{(k)})$ where c_k is the weight associated to $Z^{(k)}$. We can then let these triplets be the parameters (c_k, μ_k, σ_k^2) of the Gaussian mixture in (20).

Standard i.i.d. simulations such that $c_k = 1/K$ would be costly because the number of triplets K required to obtain a good approximation of $g(x)$ may be large. In addition, a large K makes the computation of the coefficients f_n and ℓ_n computationally more demanding. Therefore, a weighted Monte-Carlo method appears to be a more sensible approach to parametrize the Gaussian mixture. For example, one can use an optimal K-quantization of the multivariate Gaussian distribution. This is a discrete probability distribution in $\mathbb{R}^n$ with K mass points that best approximates the multivariate normal (23) in the L^2 sense, we refer to (Pagès and Printems 2003) for more details. The K pairs $(c_k, Z^{(k)})$ can always be precomputed and loaded on demand since they do not depend on the stochastic volatility model at hand.

The discretization scheme used to numerically integrate the stochastic differential equations may also play an important role. We want a scheme that performs well with a large time step Δt since we are only interested in the log price density at time T. Numerical experiments on multiple models showed that good results can notably be obtained with the Interpolated-Kahl-Jäckel (IJK) scheme introduced in (Kahl and Jäckel 2006). The IJK scheme boils down to a scheme with linear interpolation of the drift and of the diffusion, consideration for the correlated diffusive terms, and with a higher order Milstein term. It bears little additional computational cost.

3.3 Nonconvergent Option Price Approximations

We provide some explanations why, even when $\ell \notin L^2_w$, the price approximation $\pi_f^{(N)}$ in (5) can be an accurate approximation of the true option price π_f. More precisely, we estimate the option price approximation error.

Assume that instead of considering the true process X_t we consider the modified process X_t^τ whose dynamics is

$$dX_t^\tau = (r-\delta)dt - \frac{1}{2}d\langle X^\tau\rangle_t + 1_{\{\tau>t\}}\left(\Sigma_1(Y_t)\,dW_{1t} + \Sigma_2(Y_t)\,dW_{2t}\right)$$

where the stopping time τ is defined by

$$\tau = \inf\{t \geq 0 : |M_t| \geq C_1 \text{ or } C_t \geq C_2\}$$

for some positive constants C_1 and C_2. The event $\{\tau < T\}$ is thought be very unlikely, so that $\mathbb{P}[X_T = X_T^\tau] > 1-\epsilon_0$ for some small $\epsilon_0 > 0$. Note that the modified discounted cum-dividend asset price process $\mathrm{e}^{-(r-\delta)t+X_t^\tau}$ remains a martingale. As a consequence of this construction, Proposition 3.2 applies to the process X_t^τ which implies that the option price has a series representation for some Gaussian mixture density $w(x)$,

$$\pi_f^\tau = \mathbb{E}[f(X_T^\tau)] = \sum_{n=0}^{\infty} f_n \ell_n^\tau, \tag{24}$$

where the payoff coefficients f_n are defined as in (4) and the likelihood coefficients ℓ_n^τ are defined by

$$\ell_n^\tau = \mathbb{E}[H_n(X_T^\tau)]. \tag{25}$$

However, we do not know the moments of X_T^τ and we will use instead the moments of X_T. We hence obtain the option price approximation

$$\pi_f^{(N)} = \sum_{n=0}^{N} f_n \ell_n,$$

which is similar to (5) but with the important difference that it will not converge to the true price π_f. The pricing error with this approximation can be decomposed into three terms,

$$\left|\pi_f - \pi_f^{(N)}\right| \leq \left|\mathbb{E}[(f(X_T) - f(X_t^\tau))1_{\{\tau\leq T\}}]\right| + \left|\pi_f^\tau - \sum_{n=0}^{N} f_n\ell_n^\tau\right| + \left|\sum_{n=0}^{N} f_n(\ell_n - \ell_n^\tau)\right| =: \epsilon_1 + \epsilon_2 + \epsilon_3. \tag{26}$$

The first error term ϵ_1 in (26) is the difference between the option price with X_T and with X_T^τ. It can easily be controlled. For example if $|f(x)|$ is bounded by K on $\mathbb{R}$ then we have

$$\epsilon_1 \leq 2\,K\,\epsilon_0.$$

The difference between the call option prices can also be bounded by using the put-call parity given that $\mathbb{E}[\mathrm{e}^{X_T}] = \mathbb{E}[\mathrm{e}^{X_T^\tau}]$. The second error term ϵ_2 in (26) is the option price approximation error for the log price X_T^τ that converges to zero as $N \to \infty$. The third error term ϵ_3 in (26) is the difference between the option price approximations of X_T and X_T^τ, that we expect to be small for a low order N but will typically diverge as $N \to \infty$. This last error term can be further decomposed as

$$\begin{aligned}\epsilon_3 &= \left|\mathbb{E}\left[\sum_{n=0}^{N} f_n(H_n(X_T) - H_n(X_T^\tau))1_{\{\tau\leq T\}}\right]\right| \\ &\leq \left|\mathbb{E}\left[\sum_{n=0}^{N} f_n H_n(X_T)1_{\{\tau\leq T\}}\right]\right| + \left|\mathbb{E}\left[\sum_{n=0}^{N} f_n H_n(X_T^\tau)1_{\{\tau\leq T\}}\right]\right|.\end{aligned} \tag{27}$$

The first term on the right side of (27) is precisely the one that is not expected to converge as $N \to \infty$. Applying the Cauchy-Schwarz inequality we derive the upper bound

$$\left|\mathbb{E}\left[\sum_{n=0}^{N} f_n H_n(X_T)1_{\{\tau\leq T\}}\right]\right| \leq \sqrt{\mathbb{E}[p_N(X_T)]\,\epsilon_0}, \quad \text{where} \quad p_N(x) = \left(\sum_{n=0}^{N} f_n H_n(x)\right)^2$$

which can be computed explicitly. In practice, this bound may give an indication to whether the option price approximation $\pi_f^{(N)}$ is reasonable.

Remark 3.3. *A simple trick to stabilize the option price approximation when Proposition 3.2 does not apply is to match the N^*-th moment of the log price and the auxiliary density, $\mathbb{E}[X_T^{N^*}] = \int_{\mathbb{R}} x^{N^*} w(x)dx$. By doing so, the first N^* moments of $g(x)$ and $w(x)$ will be of the same magnitudes which should result in well-behaved values for the ℓ_n coefficients. This can be achieved, for example, by including a component with small mixture weight and which is used to tune the N^*-th moment.*

4 Numerical Applications

We show that a simple Gaussian mixture with two components can be used to improve the convergence rate of the option price approximations in the Jacobi model. We use this technique to swiftly calibrate the model on a sample of option prices. We then derive accurate option price approximations in the Stein–Stein and Hull–White models using Gaussian mixtures with many components for the auxiliary density. We also approximate the price of an option on the realized variance in the GARCH diffusion model.

4.1 Jacobi Model

The dynamics of (X_t, Y_t) in the Jacobi model has the form

$$\begin{aligned} dY_t &= \kappa(\theta - Y_t)\,dt + \sigma\sqrt{Q(Y_t)}\,dW_{1t} \\ dX_t &= (r - \delta - Y_t/2)\,dt + \rho\sqrt{Q(Y_t)}\,dW_{1t} + \sqrt{Y_t - \rho^2 Q(Y_t)}\,dW_{2t}\end{aligned}$$

where $Q(y) = (y - y_{min})(y_{max} - y)/(\sqrt{y_{max}} - \sqrt{y_{min}})^2$, $\kappa > 0$, and $\theta \in (v_{min}, v_{max}]$. The log return volatility $\sqrt{Y_t}$ takes values in $[\sqrt{v_{min}}, \sqrt{v_{max}}]$, the leverage is $\mathrm{lev}(X_t) = \rho\sqrt{Q(Y_t)/Y_t}$, and the volatility of volatility is $\mathrm{volvol}(X_t) = \frac{\sigma}{2}\sqrt{Q(Y_t)/Y_t}$. The Jacobi model was introduced in (Ackerer, Filipović, and Pulido 2016).

We illustrate the advantages of using a Gaussian mixture with the Jacobi model when $\mathbb{V}\mathrm{ar}[X_T] < y_{max}T/2$, so that the first two moments cannot be matched with a single auxiliary Gaussian density

$w_1(x) = v_1(x)$. We consider here a Gaussian mixture with two components $w_2(x)$ as auxiliary density given by

$$w_2(x) = c_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) + (1-c_1) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right)$$

for some weight $0 < c_1 < 1$, some mean parameters $\mu_1, \mu_2 \in \mathbb{R}$ and volatility parameters $\sigma_1, \sigma_2 > 0$. We match the the first two moments of X_T and $w_2(x)$, which gives the following underdetermined system of equations

$$\mathbb{E}[X_T] = c_1\,\mu_1 + (1-c_1)\,\mu_2 \tag{28}$$

$$\mathbb{E}[X_T^2] = c_1\,(\sigma_1^2 + \mu_1^2) + (1-c_1)\,(\sigma_2^2 + \mu_2^2). \tag{29}$$

We set $\mathbb{E}[X_T] = \mu_1 = \mu_2$, so that (28) is automatically satisfied and (29) rewrites

$$c_1 = \frac{\sigma_2^2 - \mathbb{V}\mathrm{ar}[X_T]}{\sigma_2^2 - \sigma_1^2}. \tag{30}$$

Hence it must be that $\left|\sigma_2^2 - \mathbb{V}\mathrm{ar}[X_T]\right| < \left|\sigma_2^2 - \sigma_1^2\right|$ and that $\left(\sigma_2^2 - \mathbb{V}\mathrm{ar}[X_T]\right)\left(\sigma_2^2 - \sigma_1^2\right) \geq 0$. We set $\sigma_2 = \sqrt{y_{max}T/2} + 10^{-4}$, so that Proposition 3.2 applies which ensures that the option price expansion will converge to the true price. Then we arbitrarily fix $c_1 = 95\%$ and solve (30) to get σ_1. The reason behind these choices is that, by doing so, the mixture component with large weight c_1 is almost a Gaussian approximation of the log price density since $\mathbb{E}[X_T] = \mu_1$ and $\mathbb{V}\mathrm{ar}[X_T] \approx \sigma_1^2$. In the following numerical example we use the parameters: $r = \delta = x_0 = 0$, $\kappa = 0.5$, $\theta = Y_0 = 0.04$, $\sigma = 1$, $y_{min} = 10^{-4}$, $y_{max} = 0.36$, and $T = 1/12$. The upper bound on the volatility support is therefore 60%.

The left panel of Figure 1 displays the Gaussian mixture with one and two components used as auxiliary density along with the log price density approximation (9) at the truncation order $N = 100$. We expect that $g(x) \approx \ell^{(100)}(x) w_2(x)$. It is clear from the figure that $w_1(x)$ is a poor approximation of the density $g(x)$ whereas $w_2(x)$ appears more sensible. As a consequence, the implied volatility of a call option with log strike $k = 0$ converges significantly faster using $w_2(x)$ as auxiliary density as can be seen on the right panel of Figure 1. The call (put) implied volatility is initially overestimated with $w_1(x)$ since it has significantly more weight in the right (left) tail than $g(x)$. This behavior is confirmed in Table 1 which reports the implied volatility error with respect to the approximation at the order $N = 100$ for call options with different moneyness and for different truncation order. More problematic, the option price approximation of the far OTM option is negative between $N = 2$ and $N = 18$ with $w_1(x)$.

Equipped with a Gaussian mixture with two components, we calibrate the Jacobi model on a sample of S&P500 options and compare its suitability to fit the implied volatility surface with the Heston model. We select all the call and put options available on March 30, 2017 with maturities in 1, 2, 3, or 4 weeks from the OptionMetrics database. With a linear regression we extract from the put-call parity the risk-free rate $r = 1.66\%$ and the dividend yield $\delta = 1.50\%$. For each maturity we select a sample of 25 call options with a Delta ranging from 5% to 95%. We denote here π_{ij}, σ_{ij}, and ν_{ij} the j-th option price, implied volatility, and Vega from the i-week maturity sample. Similarly $\hat{\pi}_{ij}$ and $\hat{\sigma}_{ij}$ denote the model, Jacobi or Heston, option price and implied volatility. We calibrate the two models to the implied volatility surface by minimizing the weighted root-mean-square-error (RMSE)

$$\sqrt{\frac{1}{100}\sum_{i=1}^{4}\sum_{j=1}^{25}\left(\frac{\pi_{ij} - \hat{\pi}_{ij}}{\nu_{ij}}\right)^2}.$$

This criterion is a computationally efficient approximation for the implied volatility surface RMSE criterion which follows by observing that

$$\sigma_{ij} - \hat{\sigma}_{ij} \approx \frac{\pi_{ij} - \hat{\pi}_{ij}}{\nu_{ij}} \quad \text{when} \quad \pi_{ij} \approx \hat{\pi}_{ij}.$$

Table 2 reports the fitted parameters and Figure 2 displays the corresponding implied volatility surfaces. We observe that the values of κ, θ, and ρ are relatively similar, but the vol-of-vol parameter σ is almost twice

larger for the Jacobi model which suggests that its volatility process may take large values, close to y_{max}, within little time. The fitted volatility support goes, roughly, from 5% to 40% which seems reasonable for this sample. With two additional parameters, the Jacobi model is able to better fit the implied volatility surface than the Heston model. Indeed, the resulting RMSE on the implied volatility is twice smaller for the Jacobi model. In particular, the Jacobi models seems to perform better in capturing the short-term skew and smile curvature.

4.2 Stein–Stein Model

The dynamics of (X_t, Y_t) in the Stein–Stein model has the form

$$\begin{aligned} dY_t &= \kappa(\theta - Y_t)dt + \sigma dW_{1t} \\ dX_t &= \left(r - \delta - Y_t^2/2\right)dt + \rho Y_t dW_{1t} + \sqrt{1-\rho^2}Y_t dW_{2t} \end{aligned} \tag{31}$$

for some positive parameters κ, θ, and σ, and with $\rho \in (-1, 1)$. The process Y_t follows an Ornstein-Uhlenbeck process. The asset return volatility is $|Y_t|$, the leverage is $\mathrm{lev}(X_t) = \rho\,\mathrm{sign}(Y_t)$, and the volatility of volatility is $\mathrm{volvol}(X_t) = \sigma$. This model has been introduced by (Stein and Stein 1991) and generalized by (Schöbel and Zhu 1999) to allow for non-zero leverage.

The Stein–Stein model has the particularity that the diffusion (X_t, Y_t, Y_t^2) is not only polynomial but also affine. This enables the use of standard Fourier transform techniques to compute option prices that we use as reference option prices, see for details (Carr and Madan 1999), (Duffie, Filipović, and Schachermayer 2003), and (Fang and Oosterlee 2009). We approximate option prices using a Gaussian mixture for the auxiliary density as described in Section 3.2.

Applying the IJK scheme we obtain the following discretization for the dynamics of the process Y_t

$$Y_{t_{i+1}} = Y_{t_i} + \kappa(\theta - Y_{t_i})\Delta t + \sigma \Delta W_{1,t_{i+1}} \quad i = 0, \dots, n-1$$

for the conditional mean

$$M_T = X_0 - \frac{1}{2}\sum_{i=1}^{n} \frac{Y_{t_i}^2 + Y_{t_{i-1}}^2}{2}\Delta t + \rho \sum_{i=1}^{n} Y_{t_i}\Delta W_{1,t_i} + \frac{1}{2}\rho\sigma \sum_{i=1}^{n} (\Delta W_{1,t_i}^2 - \Delta t)$$

and for the conditional variance

$$C_T = (1-\rho^2)\sum_{i=1}^{n} \frac{Y_{t_i}^2 + Y_{t_{i-1}}^2}{2}\Delta t.$$

The time to maturity is $T = 1/12$ and we use a single step, $\Delta t = T$. We therefore use the optimal quantizers of the univariate normal distribution to approximate the Brownian increment $\Delta W_{1\Delta t}$. We also approximate option prices using an extended mixture obtained by including one additional component whose purpose is to adjust the moments of the auxiliary density as suggested in Remark 3.3. This additional component has a fixed weight equal to 5% while the other component weights are scaled down by 95%, its mean parameter is set equal to zero, and its variance parameter is computed such that the N^*-th moment of the auxiliary density and of the log price density are equal. A different number of components K for each auxiliary density and the first moment of the log price is always matched such that $\ell_1 = 0$. The parameters are $N^* = 20$, $r = \delta = X_0 = k = 0$, $\kappa = 0.5$, $\theta = Y_0 = 0.2$, $\sigma = 0.5$, and $\rho = -0.5$.

Table 3 reports the implied volatility errors for a call option with log strike $k = 0$ for different K and using the Gaussian mixture and the extended Gaussian mixture as auxiliary density, respectively denoted GM and GM$^+$. We can see that the implied volatility errors rapidly become small as the truncation order N increases for all K with the extended mixture and continue to converge toward π_f well after the matched moment N^*. On the other hand, the implied volatility errors do not appear to converge when K is small with the standard Gaussian mixture. Note that good option price approximations can be achieved by choosing a large number of components K even without moment matching.

4.3 Hull–White Model

The dynamics of (X_t, Y_t) in the Hull–White model has the form

$$\begin{aligned} dY_t &= \kappa(\theta - Y_t)dt + (\nu + \gamma Y_t)dW_{1t} \\ dX_t &= (r - \delta - Y_t^2/2)dt + \rho Y_t dW_{1t} + \sqrt{1-\rho^2} Y_t dW_{2t} \end{aligned} \tag{32}$$

for some positive parameters $\kappa, \theta, \gamma > 0$, a real-valued parameter ν, and with $\rho \in (-1, 1)$. The process Y_t takes values in $(-\nu/\gamma, \infty)$ if $\theta \geq -\nu/\gamma \geq 0$, so that $\nu + \gamma Y_t > 0$. The log return volatility is $|Y_t|$, the leverage is $\mathrm{lev}(X_t) = \rho$, the volatility of volatility is $\mathrm{volvol}(X_t) = \nu + \gamma Y_t$.

The Hull–White model has been introduced by (Hull and White 1987) with $\nu = 0$ and $\rho = 0$ and has been generalized by (Sepp 2016) for non-zero leverage and jumps but still with $\nu = 0$. The Stein–Stein model is obtained as a limit case for $\gamma \to 0$ and $\nu \neq 0$. The dynamics of the Hull–White model is extremely volatile as, for example, the stock price does not even admit a second moment for any finite horizon when the correlation parameter ρ is not sufficiently negative, see (Lions and Musiela 2007).

The diffusion (Y_t, Y_t^2, X_t) has the polynomial property but is not affine. Indeed, applying the infinitesimal generator of (32) to a monomial $Y_t^m X_t^n$ gives

$$\begin{aligned} \mathcal{G}\, y^m x^n = {} & m(\kappa\theta + (m-1)\nu\gamma) y^{m-1} x^n + m(0.5(m-1)\gamma^2 - \kappa) y^m x^n \\ & + 0.5m(m-1)\nu^2 y^{m-2} x^n + mn\rho\gamma y^{m+1} x^{n-1} + n(m\rho\nu + r - \delta) y^m x^{n-1} \\ & + 0.5n(n-1) y^{m+2} x^{n-2} - 0.5n y^{m+2} x^{n-1} \end{aligned} \tag{33}$$

for any integers m and n. Note that the last term in (33) is a monomial of order $m + n + 1$, so that the diffusion (Y_t, X_t) does not have the polynomial property on its own.

Applying the IJK scheme we obtain the following discretization for the dynamics of the process Y_t

$$Y_{t_{i+1}} = Y_{t_i} + \kappa(\theta - Y_{t_i})\Delta t + (\nu + \gamma Y_{t_i})\Delta W_{1,t_{i+1}} + \frac{1}{2}\gamma(\nu + \gamma Y_{t_i})(\Delta W_{1,t_{i+1}}^2 - \Delta t), \quad i = 0, \ldots, n-1$$

for the conditional mean

$$M_T = X_0 - \frac{1}{2}\sum_{i=1}^{n} \frac{Y_{t_i}^2 + Y_{t_{i-1}}^2}{2}\Delta t + \rho \sum_{i=1}^{n} Y_{t_i}\Delta W_{1,t_i} + \frac{1}{2}\rho \sum_{i=1}^{n} (\nu + \gamma Y_{t_i})(\Delta W_{1,t_i}^2 - \Delta t)$$

and for the conditional variance

$$C_T = (1-\rho^2)\sum_{i=1}^{n} \frac{Y_{t_i}^2 + Y_{t_{i-1}}^2}{2}\Delta t.$$

We follow the same procedure as in Section 4.2 and set a Gaussian mixture as auxiliary density, with and without N^* moment matching. The parameters are $r = \delta = X_0 = 0$, $\kappa = 0.5 = -\rho$, $\theta = Y_0 = 0.2$, $\nu = 0.25$, $\gamma = 0.5$, and $T = 1/12$. Table 4 reports the implied volatility values for a call option with log strike $k = 0$ for different K and using the Gaussian mixture and the extended Gaussian mixture as auxiliary density, respectively denoted GM and GM$^+$. As for the Stein–Stein model, the price series seems to converge faster when the number of Gaussian component K is large, and to stabilize for finite $N \leq N^*$ when the higher moment N^* is matched. The price series behavior beyond the matched moment N^* is however not as well behaved as for the Stein–Stein model. This seems to be a consequence of the exponentially fast growing higher moments in the Hull–White model which causes the third error term ϵ_3 in (26) to rapidly grow with N.

Figure 3 displays the kurtosis of the log price, $\mathbb{E}[(X_T - \mathbb{E}[X_T])^4]/\mathbb{V}\mathrm{ar}[X_T]^2$, for the Hull–White model for $\nu \in \{-0.05, 0, 0.25, 0.5\}$ and $\gamma \in \{0.5, 1\}$, and for maturities T ranging from one week and one year. The kurtosis in the limiting Stein–Stein model is obtained for $\gamma = 0$. We observe that the kurtosis grows with time T and in particular with γ and, for larger γ values, the kurtosis appears to explode at some point in time. This is in sharp contrast to the Black-Scholes model for which the kurtosis is constant, and to the Stein–Stein model for which the kurtosis range appears to take reasonable values. The erratic moments behavior renders polynomial option pricing techniques more delicate to apply for the Hull–White model than for the Jacobi or Stein–Stein models, partly because of numerical precision difficulties.

4.4 GARCH Model

The dynamics of (X_t, Y_t) in the GARCH diffusion model has the form

$$\begin{aligned} dY_t &= \kappa(\theta - Y_t)dt + (\nu + \gamma Y_t)dW_{1t} \\ dX_t &= (r - \delta - Y_t/2)dt + \rho\sqrt{Y_t}dW_{1t} + \sqrt{1-\rho^2}\sqrt{Y_t}dW_{2t} \end{aligned} \tag{34}$$

for some positive parameters $\kappa, \theta, \gamma > 0$, a non-positive parameter $\nu \leq 0$, and with $\rho \in (-1, 1)$. The log return volatility is $\sqrt{Y_t}$, the leverage is $\mathrm{lev}(X_t) = \rho$, the volatility of volatility is $\mathrm{volvol}(X_t) = \sigma\sqrt{Y_t}/2$.

When $\nu = 0$ the process Y_t follows the GARCH diffusion process introduced in (Nelson 1990). The link between discrete time and continuous GARCH processes has been studied in (Drost and Werker 1996). The diffusion (X_t, Y_t) is not polynomial (and hence not affine), unless $\rho = 0$, and there is no analytic expression to price options in general. A formula to approximate option prices in the GARCH diffusion model was derived in (Barone-Adesi, Rasmussen, and Ravanelli 2005) for the polynomial case $\rho = 0$. In this case, our method, as applied in Sections 4.2 and 4.3, can also be used to accurately approximate stock option prices.

We now show how to approximate the price of an option on the realized variance with discounted payoff given by

$$\mathrm{e}^{-rT}(I_T - k)^+, \quad \text{where } I_t = \frac{1}{T}\int_0^t Y_s ds.$$

The diffusion (Y_t, I_t) has the polynomial property but is not affine. Indeed, applying its infinitesimal generator to the monomial $Y_t^m X_t^n$ gives

$$\begin{aligned} \mathcal{G}\, y^m i^n = {} & m(\kappa\theta + (m-1)\nu\gamma)y^{m-1}i^n + m(0.5(m-1)\gamma^2 - \kappa)y^m i^n \\ & + 0.5m(m-1)\nu^2 y^{m-2}i^n + (1/T)ny^{m+1}i^{n-1} \end{aligned}$$

for any integers m and n.

The process I_t being lower bounded by $-\nu/\gamma$, we let the auxiliary density $w(x)$ be a Gamma density with parameter $\xi = -\nu/\gamma$. We fix the parameters α and β so that the first moment is matching and that the variance of the auxiliary density is twice large than the one from I_t. This gives,

$$\mathbb{E}[I_T] = \xi + \frac{\alpha}{\beta} \quad \text{and} \quad \mathbb{V}\mathrm{ar}[I_T] = \frac{\alpha}{2\beta^2} \quad \Leftrightarrow \quad \beta = \frac{\mathbb{E}[I_T] - \xi}{2\mathbb{V}\mathrm{ar}[I_T]} \quad \text{and} \quad \alpha = \frac{(\mathbb{E}[I_T] - \xi)^2}{2\mathbb{V}\mathrm{ar}[I_T]}.$$

We do not know whether g/w belongs to L^2_w. This motivates the arbitrary choice of setting the second moment of w be strictly larger than $\mathbb{E}[I_t^2]$, so that the right tail of the auxiliary density is likely to be sufficiently thick. The parameters are $r = \delta = X_0 = 0$, $\kappa = 0.5 = -\rho$, $\theta = Y_0 = 0.2$, $\nu = -0.025$, $\gamma = 0.5$, and $T = 1/2$.

The left panel of Figure 4 displays the auxiliary density w, the density approximation $\ell^{(20)}w$ at the order 20, and a simulated density g_{mc} with 100'000 paths. The simulations were performed with the time step $1/252$ and the Euler discretization scheme. As expected, because of the second moment specification, we observe that w is more dispersed than $g_{\mathrm{mc}} \approx g$. Interestingly, the density approximation at the order 20 is visually almost perfectly aligned with the simulated density. The right panel of Figure 4 displays the price series approximation of an option on the realized variance with $k = 0.2$. We notice that the price converges rapidly and stabilizes to the same value as the Monte-Carlo estimate.

5 Conclusion

We derived tractable option price series representations for polynomial stochastic volatility models when the auxiliary density is a mixture density. In particular, we presented efficient methods to specify an auxiliary Gaussian mixture density and indicated that accurate option price approximations are possible even when the likelihood ratio function of the auxiliary density does not belong to the corresponding weighted Lebesgue space. We provided several numerical examples that illustrate the good performance of our approach.

A Proofs

This Appendix contains the proofs of all theorems and propositions in the main text.

Proof of Proposition 2.2

This proof is based on the results in (Fischer and Golub 1992). We first aim to derive a series of orthogonal monic polynomial basis $\{h_n\}_{n\geq 0}$, that is whose leading order coefficient is equal to one, and then normalize it to obtain the desired basis $\{H_n\}_{n\geq 0}$. The recurrence relation for the orthogonal monic basis is given by

$$xh_n(x) = h_{n+1}(x) + \alpha_n h_n(x) + \gamma_n h_{n-1}(x) \tag{35}$$

for all $n \geq 0$ with $h_{-1} = 0$ and $h_0 = 1$, and where the coefficients α_n, γ_n are given by

$$\alpha_n = \frac{\langle h_n^*, h_n\rangle_w}{\langle h_n, h_n\rangle_w} \quad \text{and} \quad \gamma_n = \frac{\langle h_n, h_n\rangle_w}{\langle h_{n-1}, h_{n-1}\rangle_w} \tag{36}$$

for $n \geq 0$, and where $h_n^*(x) = xh_n(x)$. The orthonormal polynomial basis is then obtained by normalizing the orthogonal monic basis, that is

$$H_n(x) = \frac{h_n(x)}{\sqrt{\langle h_n, h_n\rangle_w}},$$

which in view of Equation (35) is equivalent to define the recurrence coefficients as follows

$$a_n = \alpha_n \quad \text{and} \quad b_n = \sqrt{\gamma_n}.$$

The inner products in Equation (36) are left to be computed. We show that one can actually compute effectively and accurately the integral $\langle p, 1\rangle_w$ for any polynomial of order less than $2N$. First, we recall the Gauss quadrature rule associated with the density v_k

$$\langle p, 1\rangle_{v_k} = \int_{\mathbb{R}} p(x)v_k(x)dx = \sum_{i=0}^{N} (\nu_{i1}^k)^2\, p(\lambda_i^k)$$

where ν_i^k is the eigenvector corresponding to the eigenvalue λ_i^k of the Jacobi matrix J_N^k. Those values however do not need to be computed explicitly. Observe that, the matrix J_N^k being Hermitian, there exists a unitary matrix U_N^k whose columns are the normalized eigenvectors of J_N^k and such that

$$\Sigma_N^k := \mathrm{diag}(\lambda_0^k, \ldots, \lambda_N^k) = (U_N^k)^\top J_N^k\, U_N^k.$$

Combining the above results we obtain

$$\begin{aligned}\langle p, 1\rangle_w &= \sum_{k=1}^{K} c_k \sum_{i=0}^{N} (\nu_{i1}^k)^2\, p(\lambda_i^k) = \sum_{k=1}^{K} c_k\, \mathrm{e}_1^\top\, U_N^k\, p(\Sigma_N^k)(U_N^k)^\top\, \mathrm{e}_1 \\ &= \sum_{k=1}^{K} c_k\, \mathrm{e}_1^\top\, p(J_N^k)\, \mathrm{e}_1.\end{aligned}$$

Define the vector z_n^k as follows

$$\begin{aligned}z_{n+1}^k &= h_{n+1}(J_N^k)\mathrm{e}_1 = (J_N^k - \alpha_n)h_n(J_N^k)\mathrm{e}_1 - \gamma_n h_{n-1}(J_N^k)\mathrm{e}_1 \\ &= (J_N^k - \alpha_n)z_n^k\mathrm{e}_1 - \gamma_n z_{n-1}^k\mathrm{e}_1\end{aligned}$$

where the second equality follows from Equation (35). The inner products then rewrites

$$\langle h_n, h_n\rangle_w = \sum_{k=1}^{K} c_k\, \mathrm{e}_1^\top\, h_n(J_N^k)^\top\, h_n(J_N^k)^\top\, \mathrm{e}_1 = \sum_{k=1}^{K} c_k\, (z_n^k)^\top\, z_k$$

and similarly

$$\langle h_n^*, h_n\rangle_w = \sum_{k=1}^{K} c_k\, \mathrm{e}_1^\top\, h_n(J_N^k)^\top\, (J_N^k)^\top\, h_n(J_N^k)^\top\, \mathrm{e}_1 = \sum_{k=1}^{K} c_k\, (z_n^k)^\top\, J_N^k\, z_k.$$

Proof of Proposition 2.3

The proof follows several elementary steps

$$\begin{aligned} f_N &= \int_{\mathbb{R}} f(x) H_N(x) w(x) dx = \int_{\mathbb{R}} f(x) H_N(x) \sum_{k=1}^{k} c_k \, v_k(x) dx \\ &= \sum_{k=1}^{K} c_k \int_{\mathbb{R}} f(x) H_N(x) v_k(x) dx = \sum_{k=1}^{K} \sum_{n=0}^{N} c_k \int_{\mathbb{R}} q_{N,n}^k \, f(x) H_n^k(x) v_k(x) dx \\ &= \sum_{k=1}^{K} \sum_{n=0}^{N} c_k \, q_{N,n}^k \, \langle f, H_n^k \rangle_{v_k} \end{aligned}$$

which proves (7) and where the third line results from (8) which gives the representation of the polynomial $H_N(x)$ in the polynomials basis $H_n(x)$.

Proof of Proposition 2.5

Part (i): We want to compute

$$\int_{\mathbb{R}} f(x) L_n(x) v_k(x) dx = \frac{\sqrt{n!}}{\sqrt{\Gamma(\alpha+n)}\Gamma(\alpha)} \int_{\mu}^{\infty} (\mathrm{e}^{\xi + \frac{x}{\beta}} - \mathrm{e}^k) \mathcal{L}_n^{\alpha-1}(x) \, x^{\alpha-1} \mathrm{e}^{-x} dx$$

by a change of variable $y = \beta(x - \xi)$, with $\mu = \max(0, \beta(k - \xi))$ and $v_k(x)$ as in (10). We first show that

$$I_n^{\alpha-1}(\mu; \nu) = \int_{\mu}^{\infty} \mathrm{e}^{\nu x} \mathcal{L}_n^{\alpha-1}(x) \, x^{\alpha-1} \mathrm{e}^{-x} dx$$

satisfies the recursive system (11). This directly follows from the recursive relations

$$\mathcal{L}_n^{\alpha-1}(x) = \left(2 + \frac{\alpha - 2}{n}\right) \mathcal{L}_{n-1}^{\alpha-1}(x) - \frac{1}{n} x \, \mathcal{L}_{n-1}^{\alpha-1}(x) - \left(1 + \frac{\alpha - 2}{n}\right) \mathcal{L}_{n-2}^{\alpha-1}(x)$$

and the three-point rule

$$\mathcal{L}_n^{\alpha-1}(x) = \mathcal{L}_n^{\alpha}(x) - \mathcal{L}_{n-1}^{\alpha}(x)$$

such that we obtain

$$I_n^{\alpha-1}(\mu; \nu) = \left(2 + \frac{\alpha - 2}{n}\right) I_{n-1}^{\alpha-1}(\mu; \nu) - \left(1 + \frac{\alpha - 2}{n}\right) I_{n-2}^{\alpha-1}(\mu; \nu) - \frac{1}{n} \left(I_{n-1}^{\alpha}(\mu; \nu) - I_{n-2}^{\alpha}(\mu; \nu) \right).$$

We conclude by computing

$$\begin{aligned} I_0^{\alpha-1}(\mu; \nu) &= \int_{\mu}^{\infty} x^{\alpha-1} \mathrm{e}^{-(1-\nu)x} dx = \frac{\Gamma(\alpha)}{(1-\nu)^{\alpha}} \int_{\mu}^{\infty} \frac{(1-\nu)^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-(1-\nu)x} dx \\ &= \frac{\Gamma(\alpha)}{(1-\nu)^{\alpha}} \left(1 - \int_0^{\mu} \frac{(1-\nu)^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-(1-\nu)x}) dx\right) = (1-\nu)^{-\alpha} \left(\Gamma(\alpha) - \Gamma(\alpha, \mu(1-\nu))\right) \end{aligned}$$

and since $\mathcal{L}_1^{\alpha-1}(x) = (\alpha + x)$ we get that

$$I_1^{\alpha-1}(\mu; \nu) = \int_{\mu}^{\infty} (\alpha + x) \mathrm{e}^{-(1-\nu)x} x^{\alpha-1} dx = \alpha I_0^{\alpha-1}(\mu; \nu) + I_0^{\alpha}(\mu; \nu),$$

which completes the proof of part (i).

Part (ii): The Fourier coefficient f_n is given by the expression,

$$\begin{aligned} \int_{\mathbb{R}} f(x) H_n^k(x) v_k(x) dx &= \int_{\mathbb{R}} (x-k)^+ H_n^k(x) \frac{\beta^{\alpha}}{\Gamma(\alpha)} (x-\xi)^{\alpha-1} \mathrm{e}^{-\beta(x-\xi)} dx \\ &= \int_{k-\xi}^{\infty} (y + \xi - k) H_n(y + \xi) \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} \mathrm{e}^{-\beta y} dy \end{aligned}$$

where the second equality follows from the change of variable $y = x - \xi$. We now derive an explicit for these integrals which depend on the incomplete Gamma function,

$$\int_{k-\xi}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{j+\alpha-1} \mathrm{e}^{-\beta y} dy = \frac{1}{\Gamma(\alpha)\beta^j} \int_{\beta(k-\xi)}^{\infty} x^{j+\alpha-1} \mathrm{e}^{-x} dx = \frac{\Gamma(\alpha+j, \beta(k-\xi))}{\Gamma(\alpha)\beta^j}, \quad j \geq 0$$

which completes the proof Proposition 2.5.

Proof of Equations (15)–(16)

The dynamics of the variance $V_t = d\langle X \rangle_t / dt = \Sigma_1(Y_t)^2 + \Sigma_2(Y_t)^2$ is of the form

$$dV_t = (\cdots)dt + 2\left(\Sigma_1'(Y_t)\Sigma_1(Y_t) + \Sigma_2'(Y_t)\Sigma_2(Y_t)\right)\, \sigma(Y_t)\, dW_{1t}.$$

The quadratic covariation between X_t and V_t is therefore given by

$$d\langle X, V \rangle_t = 2\left(\Sigma_1'(Y_t)\Sigma_1(Y_t) + \Sigma_2'(Y_t)\Sigma_2(Y_t)\right)\, \sigma(Y_t)\, \Sigma_1(Y_t) dt$$

and the quadratic variation of V_t by

$$d\langle V \rangle_t = 4\left(\Sigma_1'(Y_t)\Sigma_1(Y_t) + \Sigma_2'(Y_t)\Sigma_2(Y_t)\right)^2\, \sigma(Y_t)^2 dt.$$

Equation (15) directly follows by observing that

$$\frac{2\left(\Sigma_1'(Y_t)\Sigma_1(Y_t) + \Sigma_2'(Y_t)\Sigma_2(Y_t)\right)\, \sigma(Y_t)}{\sqrt{4\left(\Sigma_1'(Y_t)\Sigma_1(Y_t) + \Sigma_2'(Y_t)\Sigma_2(Y_t)\right)^2\, \sigma(Y_t)^2}} = \operatorname{sign}\left[\left(\Sigma_1'(Y_t)\Sigma_1(Y_t) + \Sigma_2'(Y_t)\Sigma_2(Y_t)\right)\sigma(Y_t)\right],$$

and Equation (16) follows from the above and

$$d\langle \sqrt{V} \rangle_t = \frac{d\langle V \rangle_t}{4V_t}.$$

Proof of Proposition 3.1

Conditional on the trajectory of W_{1t} on $[0, T]$, the trajectory of Y_t is observable and W_{2t} is the only source of randomness in the dynamics of X_T which is thus equivalent to the dynamics of a Gaussian process with time varying parameters. Hence, its conditional distribution is given by a normal distribution with mean M_T and variance C_T as in (17) and (18). Taking expectation gives (19)

Proof of Proposition 3.2

The first part of the proposition follows from similar arguments as in (Ackerer, Filipović, and Pulido 2016, Theorem 3.1). Note that it is sufficient to consider only the component of w with the largest variance parameter.

For the second part, it is clear that C_T is bounded when Y_t takes values in a compact interval I since $\Sigma_1^2 + \Sigma_2^2$ is a polynomial, and thus Σ_1^2 is bounded on I. The random variable M_T is equivalently given by the expression

$$M_T = X_0 + (r - \delta)T - \frac{1}{2}\int_0^T \left(\Sigma_1(Y_t)^2 + \Sigma_2(Y_t)^2\right) dt + \int_0^T \phi(Y_t)(dY_t - \kappa(\theta - Y_t)dt).$$

Let $\Phi \in C^2(I)$ be a primitive function of ϕ, so that $\Phi' = \phi$, on I. Applying Ito's lemma to $\Phi(Y_t)$ we obtain

$$\int_0^T \phi(Y_t) dY_t = \Phi(Y_T) - \Phi(Y_0) - \frac{1}{2}\int_0^T \phi'(Y_t)\sigma(Y_t)^2 dt$$

which is uniformly bounded. Hence so is M_T.

B Basis Construction with Moments

In this Appendix we present moment-based constructions, alternative to Proposition 2.2, for the orthonormal basis (ONB) $H_n(x)$ of the space L^2_w which can also be used when the auxiliary density w is d-valued. Let $\pi : \mathcal{E} \to \{1, \dots, M\}$ be an enumeration of the set of exponents

$$\mathcal{E} = \left\{ n \in \mathbb{N}^d : |n| \leq N \right\}$$

for some positive integer N, with $\pi(0) = 1$, and such that $\pi(n) \leq \pi(m)$ if $|n| \leq |m|$. We denote $\pi_i = \pi^{-1}(i) \in \mathbb{N}^d$ where $\pi^{-1} : \{1, \dots, M\} \to \mathcal{E}$ is the inverse function of π.

A standard approach to construct the ONB is to apply the Gram-Schmidt algorithm outlined below. First ones constructs the orthogonal basis

$$\begin{aligned} u_0(x) &= 1 \\ u_i(x) &= x^{\pi_i} - \sum_{j=0}^{i-1} \frac{\langle x^{\pi_i}, u_j \rangle_w}{\langle u_j, u_j \rangle_w} u_j(x), \quad i \geq 1 \end{aligned}$$

and the ONB is obtained by normalization,

$$H_i(x) = \frac{u_i(x)}{\|u_i\|_w}, \quad i \geq 0.$$

Another interesting approach to construct an ONB of the space L^2_w is as follows. Let $\mathbf{M}$ denote the $(M \times M)$ Gram matrix defined by

$$\mathbf{M}_{i+1,\, j+1} = \langle x^{\pi_i}, x^{\pi_j} \rangle_w \tag{37}$$

which is thus symmetric and positive definite. Let $\mathbf{M} = \mathbf{L}\mathbf{L}^\top$ be the unique Cholesky decomposition of $\mathbf{M}$ where $\mathbf{L}$ is a lower triangular matrix, and defined the lower triangular $\mathbf{S} = \mathbf{L}^{-1}$.

Theorem B.1 ((Mysovskikh 1968))**.** *The polynomials*

$$H_i(x) = \sum_{j=0}^{i} \mathbf{S}_{i+1,j+1}\, x^{\pi_j}$$

form an ONB of L^2_w.

Remark B.2. *The orthonormal basis resulting from the classical Gram-Schmidt described above may not appear orthogonal numerically because of rounding errors, the procedure is said to be numerically unstable. To alleviate this issue the modified Gram-Schmidt implementation is often preferred in practice, the polynomial $u_i(x)$ is now computed in multiple steps*

$$u_i^{(j+1)}(x) = u_i^{(j)}(x) - \frac{\langle u_i^{(j)}(x), u_j \rangle_w}{\langle u_j, u_j \rangle_w} u_j(x), \quad j = 0, \dots, i-1$$

with $u_i^{(0)}(x) = x^{\pi_i}$ and such that $u_i(x) = u_i^{(i)}(x)$. Although the two algorithms are equivalent in exact arithmetic, significant difference can be observed in finite-precision arithmetic.

Remark B.3. *The Gram matrix in Equation* (37) *may numerically be singular because of rounding errors either in the computation in its eigenvalues or its moments. One approach to avoid this problem is to consider an approximately orthonormal basis in place of the monomial basis. By doing so, the Gram matrix would be already almost diagonal and thus more likely to be invertible. This may be achieved, for example, by implementing an algorithm that computes the ONB for the first j elements by using the ONB of the first $j-1$ elements enlarged with the monomial x^{π_j}.*

	$k=-0.1$		$k=0$		$k=0.1$	
N	$K=1$	$K=2$	$K=1$	$K=2$	$K=1$	$K=2$
0–1	23.59	3.67	26.67	1.02	25.97	3.25
2	3.77	3.67	7.47	1.02	2.88	3.25
3	2.63	1.89	7.36	0.87	-	0.17
4	7.38	1.86	4.03	0.77	-	0.03
5	5.55	1.01	4.00	0.72	-	2.47
6	5.18	0.88	2.57	0.58	-	1.80
7	4.22	0.55	2.55	0.55	-	3.00
8	3.23	0.38	1.78	0.42	-	2.10
9	2.73	0.28	1.77	0.40	-	2.32
10	1.91	0.12	1.30	0.29	-	1.66
11	1.64	0.12	1.29	0.28	-	1.57
12	1.02	0.00	0.98	0.19	-	1.15
13	0.88	0.02	0.97	0.19	-	1.02
14	0.42	0.06	0.76	0.12	-	0.78
15	0.37	0.04	0.75	0.12	-	0.67
16	0.04	0.09	0.60	0.07	-	0.53
17	0.03	0.06	0.59	0.07	-	0.44
18	0.21	0.09	0.48	0.04	5.12	0.38
19	0.18	0.07	0.47	0.03	4.43	0.31
20	0.35	0.08	0.39	0.01	3.12	0.28
30	0.39	0.00	0.15	0.01	0.49	0.04
40	0.15	0.04	0.06	0.01	0.16	0.09
50	0.02	0.04	0.02	0.01	0.32	0.10

Table 1: Implied volatility errors for the Jacobi model.

The reported values are absolute percentage errors with respect to implied volatility approximations obtained at the 100-th truncation order for call options with different log strikes k. The auxiliary density is a Gaussian mixture with two components whose two first moments match those of the log price. The "–" symbol indicates that the implied volatility was not retrievable because the option price approximation was negative.

	$\sqrt{\theta}$	κ	σ	ρ	$\sqrt{Y_0}$	$\sqrt{y_{min}}$	$\sqrt{y_{max}}$	RMSE
Jacobi	0.3035	1.8230	1.4591	-0.8177	0.0728	0.0546	0.3977	0.3102
Heston	0.2519	2.2532	0.7942	-0.6178	0.0732	–	–	0.6976

Table 2: Fitted parameters for the Heston and Jacobi models.

The table reports fitted parameters and the implied volatilities root-mean-squared-error (RMSE) in percentage. The models were calibrated on a subset of S&P500 options with maturity less than one month observed on March 30 2017.

	$K=3$		$K=10$		$K=50$	
N	GM	GM^+	GM	GM^+	GM	GM^+
0–1	1.26	0.64	0.01	1.85	0.16	1.95
2	0.75	0.12	0.20	0.24	0.10	0.24
3	0.26	0.02	0.15	0.19	0.16	0.18
4	0.43	0.07	0.09	0.05	0.15	0.05
5	0.08	0.03	0.06	0.03	0.08	0.03
6	0.68	0.06	0.04	0.01	0.03	0.02
7	0.46	0.03	0.01	0.00	0.04	0.01
8	1.27	0.04	0.01	0.01	0.01	0.02
9	1.56	0.03	0.03	0.01	0.00	0.01
10	2.94	0.03	0.04	0.02	0.01	0.03
11	6.06	0.02	0.06	0.01	0.01	0.02
12	8.02	0.02	0.04	0.02	0.01	0.03
13	25.63	0.01	0.05	0.02	0.01	0.02
14	23.63	0.01	0.18	0.02	0.02	0.03
15	–	0.01	0.10	0.02	0.01	0.03
16	–	0.01	0.43	0.02	0.02	0.03
17	–	0.01	0.03	0.02	0.01	0.03
18	–	0.01	1.30	0.02	0.02	0.02
19	–	0.01	0.63	0.02	0.02	0.02
20	–	0.00	3.82	0.02	0.02	0.02
30	–	0.00	–	0.00	0.00	0.00
40	–	0.00	–	0.00	4.65	0.00
50	–	4.34	–	4.60	–	9.96

Table 3: Implied volatility errors for the Stein–Stein model.

The reported values are absolute percentage errors with respect to the implied volatility computed with Fourier transform technique. The GM^+ column refers to option price approximations obtained with a Gaussian mixture auxiliary density whose 20-th moment is matching $\mathbb{E}[X_T^{20}]$. The "–" symbol indicates that the implied volatility was not retrievable either because the option price approximation was negative or because the implied volatility was larger than 99%.

	$K = 3$		$K = 10$		$K = 50$	
N	GM	GM$^+$	GM	GM$^+$	GM	GM$^+$
0–1	19.26	20.98	20.32	21.98	20.41	22.06
2	20.91	20.34	20.47	20.46	20.37	20.44
3	20.52	20.27	20.41	20.42	20.41	20.41
4	19.97	20.32	20.36	20.31	20.40	20.29
5	20.36	20.29	20.35	20.29	20.36	20.28
6	20.83	20.33	20.34	20.29	20.33	20.29
7	19.99	20.31	20.30	20.28	20.34	20.28
8	19.36	20.33	20.32	20.30	20.33	20.30
9	21.61	20.32	20.38	20.30	20.32	20.30
10	22.71	20.33	20.30	20.32	20.33	20.32
11	14.88	20.32	20.25	20.31	20.33	20.31
12	13.37	20.33	20.44	20.33	20.33	20.33
13	47.02	20.33	20.48	20.33	20.33	20.33
14	42.79	20.33	19.93	20.33	20.33	20.34
15	–	20.33	20.10	20.33	20.33	20.34
16	–	20.33	21.94	20.34	20.33	20.34
17	–	20.33	20.29	20.34	20.33	20.34
18	66.98	20.33	13.62	20.34	20.34	20.34
19	–	20.33	24.63	20.34	20.34	20.34
20	–	20.33	50.84	20.34	20.32	20.34
30	–	20.39	–	20.40	37.42	20.41
40	–	–	–	–	–	–
50	–	–	–	–	–	–

Table 4: Implied volatility for the Hull–White model.

The reported values are in percentage. The GM$^+$ column refers to option price approximations obtained with a Gaussian mixture auxiliary density whose 20-th moment is matching $\mathbb{E}[X_T^{20}]$. The "–" symbol indicates that the implied volatility was not retrievable either because the option price approximation was negative or because the implied volatility was larger than 99%.

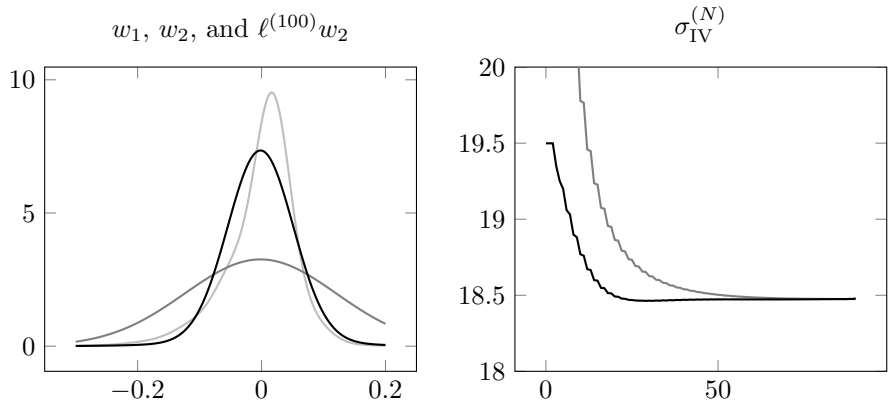

Figure 1: Auxiliary densities and implied volatility convergence.

The left panel displays the Gausian mixture used as auxiliary density with one (grey) and two (black) components as well as the log price density approximation at the order $N = 100$ (light gray). The right panel displays the implied volatility series for the corresponding two auxiliary densities for a call option with strike $k = 0$.

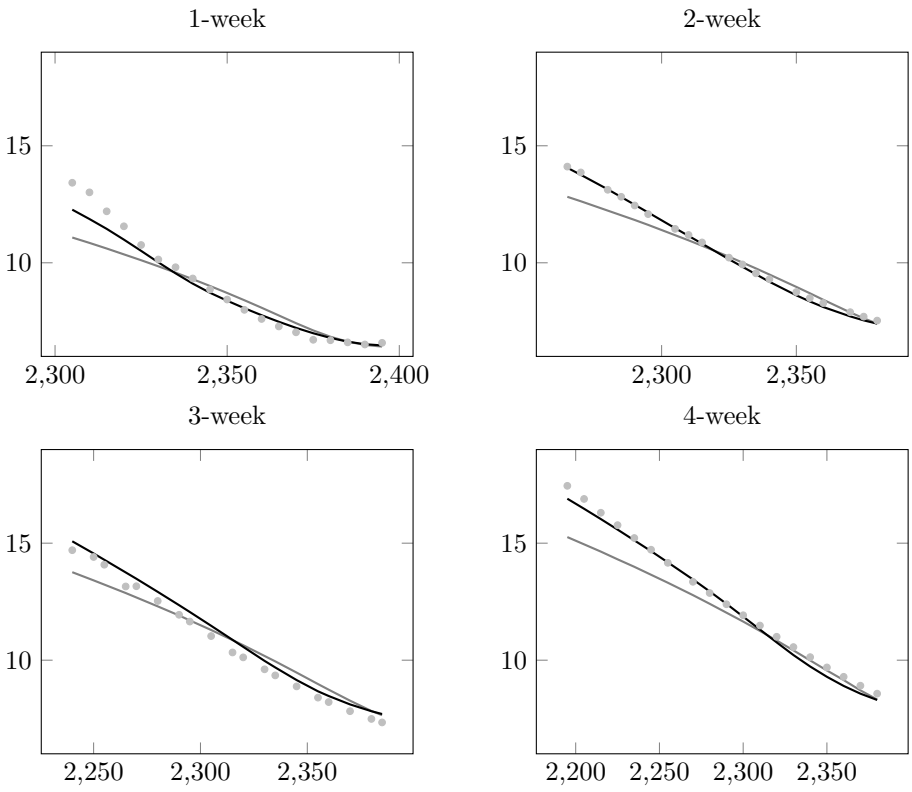

Figure 2: Fitted implied volatility surfaces for the Heston and Jacobi models.

The true (dotted light gray), the Heston (gray), and the Jacobi (black) implied volatility surfaces are displayed for each maturity as a function of the strike price. The data sample is a subset of S&P500 options with maturity less than one month observed on March 30 2017.

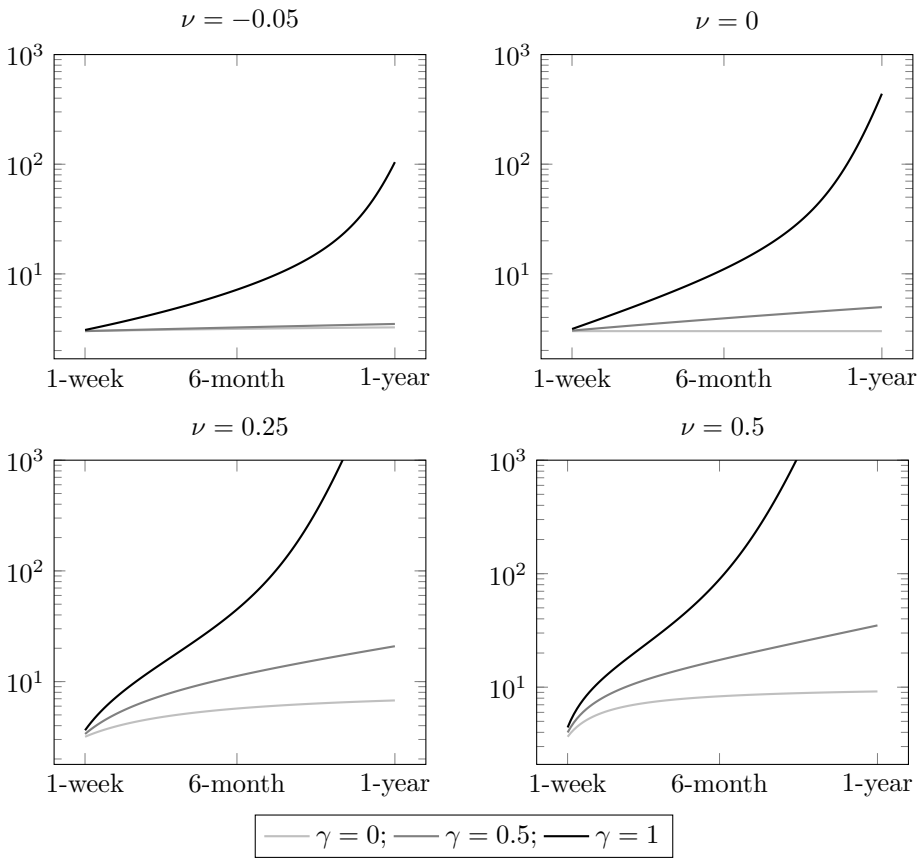

Figure 3: Kurtosis values for the Hull–White model.

The kurtosis is displayed for maturities between one week and on year, and for different combination of the ν and γ parameters. The light gray line $\gamma = 0$ gives the kurtosis of the nested Stein–Stein model.

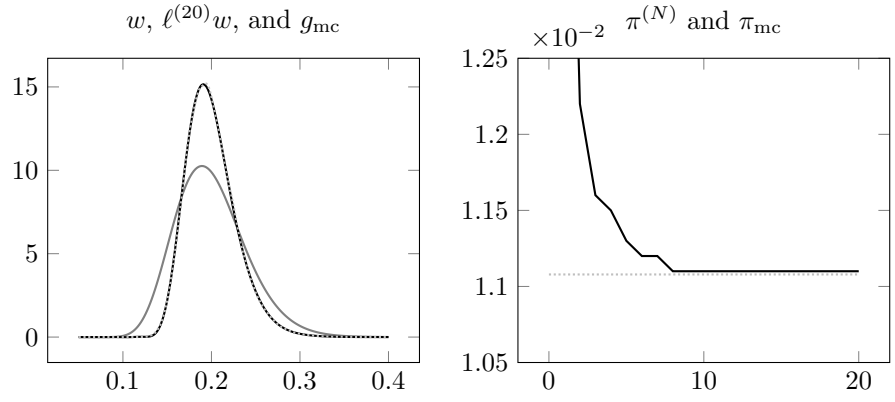

Figure 4: Realized variance density and option price approximation for the GARCH model.
The left panel displays the auxiliary density (gray), the density approximation at the order $N = 20$ (black), and the simulated density (dotted light gray) from 100'000 paths. Note that the last two densities almost perfectly coincide. The right panel displays the option price series for a call option with strike $k = 0.20$ (black) as well as the Monte-Carlo price (dotted light gray).

References

Ackerer, D. and D. Filipović (2016). Linear credit risk models. *Swiss Finance Institute Research Paper* (16-34).

Ackerer, D., D. Filipović, and S. Pulido (2016). The Jacobi stochastic volatility model. *Swiss Finance Institute Research Paper* (16-35).

Barone-Adesi, G., H. Rasmussen, and C. Ravanelli (2005). An option pricing formula for the GARCH diffusion model. *Computational Statistics & Data Analysis 49*(2), 287–310.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of political economy 81*(3), 637–654.

Callegaro, G., L. Fiorin, and M. Grasselli (2017a). Pricing via recursive quantization in stochastic volatility models. *Quantitative Finance 17*(6), 855–872.

Callegaro, G., L. Fiorin, and M. Grasselli (2017b). Quantization meets Fourier: A new technology for pricing options. *Working Paper*.

Callegaro, G., L. Fiorin, and A. Pallavicini (2017). Quantization goes polynomial. *Working Paper*.

Carr, P. and D. Madan (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance 2*(4), 61–73.

Drost, F. C. and B. J. Werker (1996). Closing the garch gap: Continuous time garch modeling. *Journal of Econometrics 74*(1), 31–57.

Duffie, D., D. Filipović, and W. Schachermayer (2003). Affine processes and applications in finance. *Annals of Applied Probability 13*(3), 984–1053.

Fang, F. and C. W. Oosterlee (2009). A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing 31*(2), 826–848.

Filipović, D., E. Gourier, and L. Mancini (2016). Quadratic variance swap models. *Journal of Financial Economics 119*(1), 44–68.

Filipović, D. and M. Larsson (2017). Polynomial jump-diffusion models. *Working Paper*.

Filipović, D., E. Mayerhofer, and P. Schneider (2013). Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics 176*(2), 93–111.

Fischer, B. and G. H. Golub (1992). How to generate unknown orthogonal polynomials out of known orthogonal polynomials. *Journal of Computational and Applied Mathematics 43*(1-2), 99–115.

Gautschi, W. (2004). *Orthogonal Polynomials: Computation and Approximation*. Numerical Mathematics and Scientific Computation. Oxford University Press, New York. Oxford Science Publications.

Glasserman, P. and K.-K. Kim (2011). Gamma expansion of the Heston stochastic volatility model. *Finance and Stochastics 15*(2), 267–296.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies 6*(2), 327–343.

Heston, S. L. and A. G. Rossi (2016). A spanning series approach to options. *The Review of Asset Pricing Studies 7*(1), 2–42.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance 42*(2), 281–300.

Kahl, C. and P. Jäckel (2006). Fast strong approximation Monte Carlo schemes for stochastic volatility models. *Quantitative Finance 6*(6), 513–536.

Küchler, U. and S. Tappe (2008). Bilateral Gamma distributions and processes in financial mathematics. *Stochastic Processes and their Applications 118*(2), 261–283.

Lions, P.-L. and M. Musiela (2007). Correlations and bounds for stochastic volatility models. In *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, Volume 24, pp. 1–16. Elsevier.

Lipton, A. and A. Sepp (2008). Stochastic volatility models and Kelvin waves. *Journal of Physics A: Mathematical and Theoretical 41*(34), 344012.

McNeil, A. J., R. Frey, and P. Embrechts (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

McWalter, T. A., R. Rudd, J. Kienitz, and E. Platen (2017). Recursive marginal quantization of higher-order schemes. *Working Paper*.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of economics and management science*, 141–183.

Mysovskikh, I. (1968). On the construction of cubature formulas with fewest nodes. In *Dokl. Akad. Nauk SSSR*, Volume 178, pp. 1252–1254.

Nelson, D. B. (1990). ARCH models as diffusion approximations. *Journal of Econometrics 45*(1-2), 7–38.

Pagès, G. and J. Printems (2003). Optimal quadratic quantization for numerics: The Gaussian case. *Monte Carlo Methods and Applications 9*(2), 135–165.

Pagès, G. and A. Sagna (2015). Recursive marginal quantization of the Euler scheme of a diffusion process. *Applied Mathematical Finance 22*(5), 463–498.

Schöbel, R. and J. Zhu (1999). Stochastic volatility with an Ornstein–Uhlenbeck process: An extension. *Review of Finance 3*(1), 23–46.

Schoutens, W. (2012). *Stochastic Processes and Orthogonal Polynomials*, Volume 146. Springer Science & Business Media.

Scott, L. O. (1987). Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative Analysis 22*(4), 419–438.

Sepp, A. (2016). Log-normal stochastic volatility model: Affine decomposition of moment generating function and pricing of vanilla options. *Working Paper*.

Stein, E. M. and J. C. Stein (1991). Stock price distributions with stochastic volatility: An analytic approach. *The Review of Financial Studies 4*(4), 727–752.

Wiggins, J. B. (1987). Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics 19*(2), 351–372.

Wilcox, R. (1967). Exponential operators and parameter differentiation in quantum physics. *Journal of Mathematical Physics 8*(4), 962–982.