Markovian projection onto a Heston model

A. Antonov*, T. Misirpashaev†, and V. Piterbarg‡

June 27, 2007

Abstract

We develop a systematic approach to the reduction of dimensionality of smile-enabled models by projecting them onto a displaced version of the two-dimensional Heston process. The projection is the key for deriving efficient, analytical approximations to European option prices in such models. This is a further development of the method of Markovian projection previously used for projecting on the displaced-diffusion process (with skew but without smile). The method is derived in a generic form and has a wide range of suitable applications. Examples for spread and basket options are given.

1 Introduction

Multi-dimensional stochastic models arise naturally in derivatives valuation. In a typical situation, the quantity of interest is a function of several underlying variables, each in isolation following relatively simple dynamics but joined together in a complex model. It is usually possible to derive an equation for the quantity of interest with an explicit, if cumbersome, expression for the stochastic volatility driven by multiple risk factors. Two characteristic examples are provided by equity basket models and forward LIBOR models. In the first case, the elementary underlying variables are individual stocks, S_i, for which one can assume, for example, a model with local or stochastic volatility. The value of the basket, which is a weighted average of the stocks, $B = \sum w_i S_i$, follows a process with a complicated stochastic volatility even in the simplest case of correlated Black-Scholes models for the individual stocks. In the second case, the elementary variables are the LIBOR rates, $L_i(t)$, and the quantity of interest can be, for example, a certain swap rate, R, the equation for which can be derived by applying Ito's

*NumeriX Software Ltd., 41 Eastcheap Street, 2nd Floor London EC3M 1DT UK; antonov@numerix.com
†Merrill Lynch, 4 World Financial Center, New York NY 10080 USA; Timur_Misirpashayev@ml.com
‡Barclays Capital, 5 The North Colonnade, London E14 4BB; vladimir.piterbarg@barcap.com

lemma to an explicit expression for the rate R in terms of the LIBOR rates. A simpler, yet still non-trivial, example is offered by a model in which the dynamics of the spread of two assets are obtained from a pair of models for the two assets.

The only generic way to solve a multi-dimensional stochastic model without approximations is a brute-force Monte Carlo simulation. However, in many applications it is necessary to develop efficient low-dimensional approximations which avoid a time-consuming simulation of all risk factors. One of the most important is model calibration, where a set of model parameters is found in a cycle of multiple recalculations of calibration instrument prices, repeated until the target prices are fitted.

We proceed to describe our strategy for the derivation of efficient low-dimensional approximations. Assume that we obtained an equation for the quantity of interest, S, in its martingale measure

$$dS(t) = \lambda(t) \cdot dW(t). \tag{1}$$

for some fixed initial value $S(0)$. Here and below, $W(t)$ is an F-component vector of independent Brownian motions, $\lambda(t)$ is an F-component stochastic volatility process, and the dot denotes a scalar product. As a simplest possibility, one can look for an effective local volatility process

$$dS^*(t) = \lambda^*(t, S^*(t))dZ(t), \quad S^*(0) = S(0), \tag{2}$$

driven by a single Brownian motion $dZ(t)$ and well-approximating $S(t)$ in a suitable sense. It turns out that one can achieve the identity of all one-dimensional marginal distributions of $S(t)$ and $S^*(t)$, which is ensured by a local volatility function satisfying the following equation

$$|\lambda^*(t, S^*)|^2 = \mathrm{E}[|\lambda(t)|^2 | S(t) = S^*]. \tag{3}$$

This result is due to Gyöngy [5]. Its early applications in the context of financial models ascend to Dupire [4]. A discrete-state version was analyzed in detail by Britten-Jones and Neuberger [3]. The application to the equity basket modeling developed in the current paper is similar in spirit to the work by Avellaneda et al [2] who derived an effective local volatility model for a basket of equities driven by individual local volatilities by evaluating the conditional expectation in Eq. (3) using the steepest descent method. More recently, numerous additional applications of the method were discovered by Piterbarg [7], [8] who coined for it the name of "Markovian projection" and achieved computational efficiency by restricting the space of approximating processes to the class of displaced diffusions with the local volatility[1] of the form

$$\lambda^*(t, S^*) = (1 + b(t)(S^* - S^*(0)))s(t). \tag{4}$$

Building on this work, Antonov and Misirpashaev [1] developed a systematic approach to the calculation of the optimal dependence for the shift $b(t)$ and effective volatility $s(t)$, leading to explicit formulas for European options.

[1]The case of stochastic volatility was also treated. See below for details.

Unlike the unrestricted local volatility model, an effective displaced diffusion cannot reproduce one-dimensional marginals of the original process exactly. Furthermore, while a displaced diffusion is sufficient to recover the skewness of implied Black volatilities, it is obviously inadequate if the initial model also has appreciable smile. Local volatility models are generally not well suited for smile modeling, in particular because there is no convenient, analytically tractable local volatility model leading to a curvature in the implied volatility smile. This is why smile modeling is usually achieved with stochastic volatility models, which also have an important advantage of more realistic dynamics. It is, therefore, desirable to extend the method of Markovian projection to cover the case where the target model is not a displaced diffusion but a stochastic volatility model. Piterbarg [8] gave an example of a projection onto a displaced Heston model for a spread of two underlying assets where each one follows a displaced Heston model. In this work we further develop the idea, continuing the systematic approach of [1] and deriving generic formulas applicable to a large class of initial models encountered in practice.

The difference in approaches between this work and [8] (as applied to stochastic volatility models) needs to be highlighted. The method of [8] is based on the fact that, given an arbitrary process for the underlying and an (arbitrary) exogenously-specified stochastic variance process, a local volatility component for the variance process could always be found to ensure a fit between the European option prices in the original and the projected (local-stochastic) volatility models. Hence, it leaves an open question concerning the selection of the process for the stochastic variance. While good choices could be made for many specific problems, a general approach is not available. In this paper, we look for a *systematic* method of finding the "best" stochastic variance process, by using a projection that not only tries to match the European option prices (ie, marginal distributions of the underlying), but also the one-dimensional distribution of the stochastic variance, and the joint underlying/variance distribution. The intention is that, by preserving more of the joint underlying/variance distribution information, the overall fit to European options *on the underlying* would be improved, and the need of a guess on the "right" stochastic variance is eliminated.

Our key technical tool is the multidimensional version of the Gyöngy lemma[2], which makes it possible to identify a two-dimensional Markovian process $(S^*(t), V^*(t))$ such that for any fixed $t = t_0$, the distribution of the pair $(S^*(t_0), V^*(t_0))$ is the same as that of the pair $(S(t_0), V(t_0))$ taken from the original, possibly non-Markovian, two-dimensional process. Here $V(t)$ is a measure of stochastic variance of the original process which will be properly defined below. As with the local volatility case, it is impossible to compute the required conditional expectations exactly without further projecting onto a specific form for the target process. In this paper, the target process is based on the Heston model enhanced with time-dependent coefficients and an additional shift parameter $\beta(t)$,

$$dx(t) \quad = \quad (1 + \beta(t)\,(x(t) - x(0)))\sqrt{z(t)}\,\sigma_H(t)\cdot dW(t), \quad x(0) = S(0),$$

[2]The extension of the original proof to the multi-dimensional case is trivial and we do not cover it here.

$$dz(t) \;=\; \alpha(t)\,(1-z(t))\,dt + \sqrt{z(t)}\,\sigma_z(t)\cdot dW(t), \quad z(0)=1. \tag{5}$$

We call this version of the Heston model the *shifted* (or *displaced*) Heston model. Note that the correlation ρ between the increments of the rate variable $x(t)$ and stochastic variance $z(t)$ can be non-trivial even though both variables are driven by the same multi-dimensional Brownian motion. The correlation is determined by the angle in the space of factors between the volatility vectors $\sigma_H(t)$ and $\sigma_z(t)$,

$$\rho(t) = \frac{\sigma_z(t)\cdot\sigma_H(t)}{|\sigma_H(t)|\,|\sigma_z(t)|}. \tag{6}$$

This model is sufficiently rich to generate both skew and smile, and is analytically tractable[3]. The target Markovian pair for the projection is related to the shifted Heston model by scaling, $S^*(t) = x(t)$, $V^*(t) = |\sigma_H(t)|^2 z(t)$.

This article is organized as follows. In Sect. 2, we review the main results for the Markovian projection on a displaced diffusion, based on the classic form of Gyöngy's lemma. In Sect. 3, we introduce a general approach to multi-dimensional Markovian projection based on a corresponding extension of Gyöngy's lemma. Sect. 4 presents a derivation of the parameters of the effective Heston model in terms of unconditional averages. Sect. 5 develops the perturbation technique for the calculation of the required averages, applicable to a large class of underlying processes. The technique is applied to a linear combination of Heston processes in Sect. 6. Numerical results for European spread and asset basket options are given in Sect. 7. Sect. 8 concludes and states the directions of further research.

2 One-dimensional Markovian projection

In this section we briefly review the one-dimensional Markovian projection on a displaced diffusion, which is the subject of Refs. [7], [8], and [1].

The technique is based on Gyöngy's lemma [5] according to which the Markovian local volatility process $S^*(t)$, following Eq. (2), has exactly the same marginal distributions as the initial process $S(t)$, following Eq. (1), given that the local volatility $\lambda^*(t,x)$ satisfies Eq. (3).

For every fixed t, the conditional expectation in the right hand side of (3) can be characterized as a function of state $|\lambda^*(\cdot,x)|^2$ which minimizes the L_2-distance from the true variance,

$$\chi^2 = E\left[\left(|\lambda(t)|^2 - |\lambda^*(t,S(t))|^2\right)^2\right] \to \min. \tag{7}$$

To obtain a tractable model, instead of minimizing the functional (7) over the space of all local volatility functions, we minimize it only over a subspace of affine linear functions of state,

[3] One can use a shift averaging technique similar to [6] in order to obtain an effective time-independent shift β. After that the model becomes analytically solvable in a standard Heston way.

expressible in the form

$$\lambda^*(t, S(t)) = (1 + b(t)\,\Delta S(t))\;s(t), \tag{8}$$

where $\Delta S(t) = S(t) - S(0)$, $b(t)$ is a deterministic shift function, and $s(t)$ is a deterministic volatility, which is a vector in factor space. In other words, we look for an approximating Markovian process $S^*(t)$ in the class of displaced diffusions,

$$dS^*(t) = (1 + b(t)\,\Delta S^*(t))\;s(t)\cdot dW(t), \quad S^*(0) = S(0). \tag{9}$$

To find the optimal values of the parameters of displaced diffusion, we need to minimize the functional

$$\chi^2(s(t), b(t)) = E\left[\left(|\lambda(t)|^2 - (1 + b(t)\Delta S(t))^2\,|s(t)|^2\right)^2\right] \tag{10}$$

over $b(t)$ and $s(t)$ for every fixed t. Equating to zero the variations of $\chi^2(s(t), b(t))$ over $s(t)$ and $b(t)$, we get a pair of non-linear algebraic equations which only involve unconditional averages of the underlying processes. These equations can be solved numerically, however, it is more practical to use an expansion in volatilities. Keeping only the leading terms gives the following result,

$$|s(t)|^2 = E[|\lambda(t)|^2], \tag{11}$$

$$b(t) = \frac{E\left[|\lambda(t)|^2 \Delta S(t)\right]}{2E\left[|\lambda(t)|^2\right] E\left[(\Delta S(t))^2\right]}. \tag{12}$$

Ref. [1] contains a derivation of closed-form expressions for the optimal volatility $s(t)$ and shift $b(t)$ in the leading order in volatilities for the so-called *separable* process. (We cite these formulas at the end of Sect. 5.)

As was mentioned in the Introduction, a displaced diffusion cannot reproduce one dimensional marginals of the original process exactly. While it is sufficient to recover the skewness of implied Black volatilities, the smile cannot be obtained in this way. To address this deficiency, we now leave the class of local volatility models and extend the method of Markovian projection to include *stochastic volatility* targets.

3 Multi-dimensional Markovian projection

We proceed by stating the multi-dimensional version of Gyöngy's result. Consider an N-component, generally non-Markovian process $x(t) = \{x_1(t), \ldots, x_N(t)\}$ driven by F independent Brownian motion components, $W(t) = \{W_1(t), \ldots, W_F(t)\}$, with an SDE

$$dx_n(t) = \mu_n(t)\,dt + \sigma_n(t)\cdot dW(t).$$

(Recall that the dot denotes the scalar product in the factor space, so that $\sigma_n(t)\cdot dW(t) \equiv \sum_f \sigma_{nf}(t)\,dW_f(t)$.) The process $x(t)$ can be approximated ("mimicked" in Gyöngy's words) by a Markovian N-component process $x^*(t)$, such that the joint distribution of all components frozen at any time remains unchanged.[4] Specifically, the components of the approximating Markovian process $x^*(t)$ satisfy an SDE

$$dx_n^*(t) = \mu_n^*(t, x^*(t))\, dt + \sigma_n^*(t, x^*(t))\cdot dW(t), \quad x_n^*(0) = x_n(0),$$

with the coefficients being deterministic functions of time and achieved state of the process, such that

$$\begin{aligned} \mu_n^*(t,y) &= E[\mu_n(t)\,|\,x(t)=y], \qquad n = 1,2,\ldots,N, \\ \sigma_n^*(t,y)\cdot\sigma_m^*(t,y) &= E[\sigma_n(t)\cdot\sigma_m(t)\,|\,x(t)=y], \qquad n,m = 1,2,\ldots,N. \end{aligned}$$

Having in mind the desired form (5) of the projected process, we write the SDE for the quantity of interest $S(t)$ in the following form

$$dS(t) = (1 + \beta(t)\,\Delta S(t))\Lambda(t)\cdot dW(t). \tag{13}$$

Here $\Delta S(t) = S(t) - S(0)$, $\beta(t)$ is a deterministic function and $\Lambda(t) = \lambda(t)/(1 + \beta(t)\,\Delta S(t))$. The optimal choice of the function $\beta(t)$ will be discussed below. For now we only notice that the form (13) is as general as (1) because $\Lambda(t)$ can be an arbitrary adapted process. For the second component, we choose the quantity

$$V(t) = |\Lambda(t)|^2 = \frac{|\lambda(t)|^2}{(1+\beta(t)\,\Delta S(t))^2}, \tag{14}$$

which has the meaning of the (displaced-diffusion) variance of the $S(t)$. Given an explicit expression for $\lambda(t)$, an SDE for $V(t)$ can be derived using Ito's lemma, which will generally contain both a diffusion and a drift term. Thus we start with the following non-Markovian pair $\{S(t), V(t)\}$

$$dS(t) = (1 + \Delta S(t)\beta(t))\,\Lambda(t)\cdot dW(t), \tag{15}$$

$$dV(t) = \mu_V(t)\,dt + \sigma_V(t)\cdot dW(t). \tag{16}$$

By Gyöngy's lemma there exists a Markovian pair process $\{S^*(t), V^*(t)\}$ having the same distribution as the initial pair process $\{S(t), V(t)\}$ at every t. The mimicking pair $\{S^*(t), V^*(t)\}$ satisfies the SDEs

$$\begin{aligned} dS^*(t) &= (1 + \Delta S^*(t)\beta(t))\ \sigma_S(t; S^*(t), V^*(t))\cdot dW(t), \\ dV^*(t) &= \mu_V^*(t; S^*(t), V^*(t))\,dt + \sigma_V^*(t; S^*(t), V^*(t))\cdot dW(t), \end{aligned}$$

[4] In the single-component case, we can say without the risk of confusion that all one-dimensional marginals of the approximating process remain unchanged. In the multi-component case, we speak of the identity of joint distributions of process components frozen in time.

with the following constraints on the deterministic coefficient functions

$$\begin{aligned}
\mu_V^*(t;s,v) &= E[\mu_V(t)\,|\,S(t)=s,V(t)=v], &(17)\\
|\sigma_S^*(t;s,v)|^2 &= E[|\Lambda(t)|^2\,|\,S(t)=s,V(t)=v]=v, &(18)\\
|\sigma_V^*(t;s,v)|^2 &= E[|\sigma_V(t)|^2\,|\,S(t)=s,V(t)=v], &(19)\\
\sigma_S^*(t;s,v){\cdot}\sigma_V^*(t;s,v) &= E[\Lambda(t){\cdot}\sigma_V(t)\,|\,S(t)=s,V(t)=v]. &(20)
\end{aligned}$$

Exact computation of the conditional expectations is usually impossible, so that it is necessary to use an approximation based on a suitable choice of a functional basis. The simplest choice is a basis of linear functions, which naturally leads to models of affine type, including the shifted Heston model.

The Markovian projection procedure consists of two steps. In the first step, we fix the shift function $\beta(t)$ and calculate the conditional expectations (17)–(20) using a projection on the basis corresponding to the shifted Heston model. This gives us optimal parameters of the Heston model as functionals of $\beta(t)$. In the second step, we determine the optimal function $\beta(t)$ minimizing the approximation error from the first step. Intuitively, an optimal choice of the function $\beta(t)$ should make the approximating process for the variance $V^*(t)$ as close to Markovian as possible, as much as possible reducing the dependence of the coefficients μ_V^* and σ_V^* on the rate $S^*(t)$.

4 Optimal coefficients

In the first step of the procedure of Markovian projection, we assume a certain shift function $\beta(t)$. As discussed above, we look for a mimicking pair $\{S^*(t),V^*(t)\}$ closely related to the shifted Heston process (5). We define the rate $S^*(t)=x(t)$ and variance $V^*(t)=z(t)\,|\sigma_H(t)|^2$. This leads to the following SDEs for the pair $\{S^*(t),V^*(t)\}$

$$\begin{aligned}
dS^*(t) &= (1+\beta(t)\,\Delta S^*(t))\frac{\sqrt{V^*(t)}}{|\sigma_H(t)|}\,\sigma_H(t){\cdot}dW(t),\\
dV^*(t) &= \left(V^*(t)\left(\left(\log|\sigma_H(t)|^2\right)' - \theta(t)\right) + \theta(t)\,|\sigma_H(t)|^2\right)dt\\
&+ |\sigma_H(t)|\,\sqrt{V^*(t)}\,\sigma_z(t){\cdot}dW(t),
\end{aligned}$$

where $\Delta S^*(t)=S^*(t)-S^*(0)$, and prime $'$ denotes the derivative of a deterministic function of time.

The components of the volatility functions $\sigma_H(t)$ and $\sigma_z(t)$ together with the function $\theta(t)$ form our chosen basis for the calculation of the conditional expectations in the right-hand side of Eqs. (17)–(20). Because of the invariance with respect to rotations in the factor space, only the absolute values $|\sigma_H(t)$ and $|\sigma_z(t)|$ and the correlation $\rho(t)$ defined by Eq. (6) are constrained, but not the individual components. The quantities in the left-hand side of Eqs. (17), (19), and (20) are given by

$$
\begin{aligned}
\mu^*_\Lambda(t;s,v) &= v\left(\left(\log|\sigma_H(t)|^2\right)' - \theta(t)\right) + \theta(t)\,|\sigma_H(t)|^2, \\
|\sigma^*_V(t;s,v)|^2 &= v\,|\sigma_H(t)|^2\,|\sigma_z(t)|^2, \\
\sigma^*_S(t;s,v)\cdot\sigma^*_V(t;s,v) &= v\,\sigma_z(t)\cdot\sigma_H(t).
\end{aligned}
$$

(Eq. (18) is trivially satisfied by the choice of the volatility process and need not be considered.)

Due to the L_2-measure minimizing property of the conditional expectation, our task reduces to joint minimization of the following three functionals,

$$
\chi_1^2(t) = E\left[\left(\mu_V(t) - V(t)\left(\left(\log|\sigma_H(t)|^2\right)' - \theta(t)\right) - \theta(t)\,|\sigma_H(t)|^2\right)^2\right], \tag{21}
$$

$$
\chi_2^2(t) = E\left[\left(|\sigma_V(t)|^2 - |\sigma_H(t)|^2\,|\sigma_z(t)|^2\,V(t)\right)^2\right], \tag{22}
$$

$$
\chi_3^2(t) = E\left[(\Lambda(t)\cdot\sigma_V(t) - \sigma_z(t)\cdot\sigma_H(t)V(t))^2\right], \tag{23}
$$

which leads to the following solution for the unknown functions $|\sigma_H(t)|$, $|\sigma_z(t)|$, $\theta(t)$, and $\rho(t)$,

$$
|\sigma_H(t)|^2 = E[V(t)], \tag{24}
$$

$$
\theta(t) = (\log E[V(t)])' - \frac{E[\delta\mu_V(t)\,\delta V(t)]}{E[\delta V^2(t)]}, \tag{25}
$$

$$
|\sigma_z(t)|^2 = \frac{E[V(t)|\sigma_V(t)|^2]}{E[V^2(t)]E[V(t)]}, \tag{26}
$$

$$
\rho(t) = \frac{E[V(t)\Lambda(t)\cdot\sigma_V(t)]}{\sqrt{E[V^2(t)]E[V(t)|\sigma_V(t)|^2]}}, \tag{27}
$$

where $\delta V(t) = V(t) - E[V(t)]$ and $\delta\mu_V(t) = \mu_V(t) - E[\mu_V(t)]$. The details of the calculations are given in Appendix A.

These formulas determine the optimal parameter functions of the projected shifted Heston model for a specific shift function $\beta(t)$. It remains to determine what function should be used, which can be done in different ways. A heuristic approach is to postulate a specific form based on the original model for the rate. We take a more systematic approach, which is to minimize the defects of the projection, i.e. the values of the functionals (21)–(23) after a substitution of the solutions (24)–(27). The defects are given by

$$
D_1(t) = E\left[(\delta\mu_V)^2\right] - \frac{E\left[\delta\mu_V(t)\,\delta V(t)\right]^2}{E\left[(\delta V(t))^2\right]}, \tag{28}
$$

$$D_2(t) = E\left[|\sigma_V(t)|^4\right] - \frac{E\left[|\sigma_V(t)|^2\, V(t)\right]^2}{E[V^2(t)]}, \tag{29}$$

$$D_3(t) = E\left[(\Lambda(t)\cdot\sigma_V(t))^2\right] - \frac{E\left[\Lambda(t)\cdot\sigma_V(t)\, V(t)\right]^2}{E[V^2(t)]}, \tag{30}$$

where $\delta\mu_V(t) = \mu_V(t) - E[\mu_V(t)]$.

We conclude this section with an intuitive interpretation of Eqs. (28)–(30). First, consider Eq. (28). The functional $D_1(t)$ is minimized when the mean of the variance process, μ_V, is maximally correlated with the variance process V. Obviously, $D_1(t)$ would be lower (closer to the optimal) for those processes V with the drift μ_V exhibiting only a weak dependence on S, in comparison with the processes with a strong dependence of μ_V on S. This means that the optimal choice of $\beta(t)$ is to make the drift of the variance process as independent of the underlying as possible, so that as much as possible of that dependence is concentrated in the term $1 + \beta(t)\,\Delta S(t)$. Same considerations apply to the variance of V, corresponding to Eq (29), and the covariance between the underlying and its variance, corresponding to Eq. (30).

5 Projection for a separable process

In the previous section we obtained generic expressions (24)–(27) for the parameters of the shifted Heston model which gives an optimal approximation to a given process among all shifted Heston models with a given shift function $\beta(t)$. Exact evaluation of the required averages is generally impossible but a perturbation theory in volatilities can be developed for a wide class of processes relevant in practical applications. In particular, closed form expressions in the leading order in volatilities can be derived for the case of a *separable* process introduced in [1].

Let σ be a parameter that characterizes the scale of the volatilities. We call a martingale process for the rate $S(t)$ *separable* if the volatility function $\lambda(t)$ of this process can be represented as a finite or countably infinite linear combination of several processes $X_n(t)$ which together form a Markovian multi-dimensional process $X(t) = (X_1(t), X_2(t), \ldots)$ with a drift of the order σ^2 and volatility of the order σ.

Thus we assume that [5]

$$dS(t) = \lambda(t)\cdot dW(t) = \sum_n X_n(t)\, a_n(t)\cdot dW(t), \tag{31}$$

$$dX_n = \mu_n\,(t; X(t))\; dt + \sigma_n\,(t; X(t))\cdot dW(t), \tag{32}$$

where $a_n(t)$, $\sigma_n(t;X)$ are deterministic vector functions[6] such that $a_n(t) = O(\sigma)$, $\sigma_n(t;X) = O(\sigma)$, and $\mu_n(t;X)$ are deterministic functions such that $\mu_n(t;X) = O(\sigma^2)$.

[5] This form differs from the one introduced in Ref. [1] by the omission of the multiplier $S(0) = S_0$.

[6] Recall that the driving Brownian motion $dW(t)$ is allowed to be multi-dimensional, in which case each coefficient $a_n(t)$ and $\sigma_n(t;X)$ becomes a vector with several components.

In the case of projection on a displaced diffusion form (9) considered in Ref. [1], we developed a technique for the calculation of the averages of the products of powers of $|\lambda(t)|^2$ and $\Delta S(t)$. For some of the averages, the leading order in volatilities is obtained simply by freezing all the basis processes $X_n(t)$ at $t = 0$ in the coefficients $\mu_n(t; X(t))$ and $\sigma_n(t; X(t)$. For example, this is sufficient for the evaluation of Eq. (11) for the volatility of the effective displaced diffusion, with the result

$$|s(t)| = |\hat{\sigma}(t)|, \tag{33}$$

where the "frozen" volatility[7] $\hat{\sigma}(t)$ is given by

$$\hat{\sigma}(t) = \sum_n X_n(0)\, a_n(t). \tag{34}$$

The effective shift (12) requires more work as the result of straightforward coefficient freezing would lead to an undetermined ratio 0/0. It is necessary to evaluate higher order corrections to the numerator and denominator and in each case find the first one which is not identically equal to zero. The results can be represented in a form

$$b(t) = \frac{\int_0^t \Phi(t,\tau)\cdot\hat{\sigma}(\tau)\, d\tau}{|\hat{\sigma}(t)|^2 \int_0^t |\hat{\sigma}(\tau)|^2\, d\tau}, \tag{35}$$

where $\hat{\sigma}_n(t)$ is another "frozen" volatility,

$$\hat{\sigma}_n(t) = \sigma_n(t; \{X_n(0)\}), \tag{36}$$

and $\Phi(t,\tau)$ is a kernel,

$$\Phi(t,\tau) = \sum_n a_n(t)\cdot\hat{\sigma}(t)\, \hat{\sigma}_n(\tau). \tag{37}$$

In order to calculate the Heston model optimal coefficients (24)–(27), we need to extend the calculations to include the drift and diffusion term of the stochastic variance

$$V(t) = \frac{|\lambda(t)|^2}{(1+\beta(t)\,\Delta S(t))^2} = \sum_{n,m} a_n(t)\cdot a_m(t) \left(X_n(t)\, X_m(t) \frac{1}{(1+\beta(t)\,\Delta S(t))^2} \right).$$

The answer can be given in a very compact form in terms of another kernel $\Omega(t,\tau)$,

$$\Omega(t,\tau) = 2\,(\Phi(t,\tau) - \beta(t)\,|\hat{\sigma}(t)|^2\, \hat{\sigma}(\tau)), \tag{38}$$

obtained from $\Phi(t,\tau)$ by adjusting for the shift $\beta(t)$. In the leading order in volatilities, the optimal coefficients (24)–(27) are given by the following expressions

$$|\sigma_H(t)| \;=\; |\hat{\sigma}(t)| + O(\sigma^3), \tag{39}$$

[7] We generally reserve the hat sign to indicate the result of a replacement of stochastic processes by their initial values in a function.

$$\theta(t) = \left(\log|\hat\sigma(t)|^2\right)' - \frac{\int_0^t \frac{\partial}{\partial t}|\Omega(t,\tau)|^2\, d\tau}{2\int_0^t |\Omega(t,\tau)|^2\, d\tau} + O(\sigma^2), \tag{40}$$

$$|\sigma_z(t)| = \frac{|\Omega(t,t)|}{|\hat\sigma(t)|^2} + O(\sigma^3), \tag{41}$$

$$\rho(t) = \frac{\Omega(t,t)\cdot\hat\sigma(t)}{|\Omega(t,t)|\,|\hat\sigma(t)|} + O(\sigma^2). \tag{42}$$

Note that the volatilities $\sigma_H(t)$ and $\sigma_z(t)$ can also be written in a vector form, which is consistent both with the absolute values (39), (41) and the correlation (42),

$$\begin{aligned}\sigma_H(t) &= \hat\sigma(t) + O(\sigma^3),\\ \sigma_z(t) &= \frac{\Omega(t,t)}{|\hat\sigma(t)|^2} + O(\sigma^3).\end{aligned}$$

The details of the calculation can be found in Appendix B.

We observe that Eq. (39) does not have a dependence on the shift function $\beta(t)$ while Eqs. (40)–(42) contain a dependence on $\beta(t)$ in the kernel $\Omega(t,\tau)$. To fix the optimal choice of $\beta(t)$, we impose a condition that the projection defects $D_1(t)$, $D_2(t)$, and $D_3(t)$, given by Eqs. (28)–(30), must be minimized for every t.

We show first that the minimization of $D_1(t)$ in the leading order leads to an easily solvable first-order linear differential equation for the shift function $\beta(t)$. Indeed, plugging the results (74)–(76) from Appendix B into Eq. (28), we obtain the following expression

$$D_1(t) = \int_0^t \left|\Omega'_t(t,\tau)\right|^2\, d\tau - \frac{\left(\int_0^t \Omega'_t(t,\tau)\cdot\Omega(t,\tau)\, d\tau\right)^2}{\int_0^t |\Omega(t,\tau)|^2\, d\tau} + O(\sigma^8). \tag{43}$$

The variation $\delta D_1(t)$ under a change $\beta(t) \to \beta(t) + \delta\beta(t)$ is a linear combination of $\delta\beta(t)$ and $\delta\beta'(t)$ both coefficients in which should be set to 0. It turns out, however, that the resulting two conditions are identical, so that we get a single equation which expresses the extremality of $D_1(t)$,

$$\int_0^t |\Omega(t,\tau)|^2\, d\tau \int_0^t \Omega'_t(t,\tau)\cdot\hat\sigma(\tau)\, d\tau = \int_0^t \Omega'_t(t,\tau)\cdot\Omega(t,\tau)\, d\tau \int_0^t \Omega(t,\tau)\cdot\hat\sigma(\tau)\, d\tau. \tag{44}$$

Substitution of the definition (38) of $\Omega(t,\tau)$ leads to a simple ODE for $\tilde\beta(t) = \beta(t)|\hat\sigma(t)|^2$,

$$p_1(t)\tilde\beta'(t) + p_2(t)\tilde\beta(t) + p_3(t) = 0, \tag{45}$$

with the coefficients

$$p_1(t) = \left(\int_0^t \Phi(t,\tau)\cdot\hat\sigma(\tau)\, d\tau\right)^2 - \int_0^t |\Phi(t,\tau)|^2\, d\tau \int_0^t |\hat\sigma(\tau)|^2\, d\tau, \tag{46}$$

$$p_2(t) = \int_0^t \Phi'_t(t,\tau)\cdot\Phi(t,\tau)\,d\tau \int_0^t |\hat{\sigma}(\tau)|^2\,d\tau - \int_0^t \Phi'_t(t,\tau)\cdot\hat{\sigma}(\tau)\,d\tau \int_0^t \Phi(t,\tau)\cdot\hat{\sigma}(\tau)\,d\tau, \quad (47)$$

$$p_3(t) = \int_0^t \Phi'_t(t,\tau)\cdot\hat{\sigma}(\tau)\,d\tau \int_0^t |\Phi(t,\tau)|^2\,d\tau - \int_0^t \Phi'_t(t,\tau)\cdot\Phi(t,\tau)\,d\tau \int_0^t \Phi(t,\tau)\cdot\hat{\sigma}(\tau)\,d\tau. \quad (48)$$

The initial condition $\beta(0)$ remains unconstrained by the minimization of $D_1(t)$ in the leading order in volatilities. Indeed it is possible to show that any solution of Eq. (45) leads to the same value of $D_1(t)$. On the other hand, the choice of $\beta(0)$ definitely does influence the quality of the approximation. At present we are not able to propose a universal recipe for the optimal choice of the initial condition for the shift function, however we discuss this issue in a context of a specific application in Sect. 6.2.

The minimization of $D_2(t)$ and $D_3(t)$ can lead to additional constraints on the function $\beta(t)$, which in the general case will be impossible to satisfy simultaneously. We note, however, that $D_2(t)$ and $D_3(t)$ vanish in the highest order which can be expressed solely in terms of the values and time derivatives of the basic coefficients $a_n(t)$, $\mu_n(t;X)$, and $\sigma_n(t;X)$ of a separable process. Evaluation of the first non-vanishing order of $D_2(t)$ and $D_3(t)$ requires SDEs for $\mu_n(t;X)$ and $\sigma_n(t;X)$ and leads to significant technical complications which we chose to omit from our model.

6 Projection of a linear combination of shifted Heston models

We have shown that the leading order in volatilities for the coefficients of the optimal projection on the shifted Heston model is available in closed form for any process that can be cast in a separable form (31)–(32). In this section we specialize the result to a linear combination of shifted Heston processes, which is relevant for the modeling of options on interest rate spreads and asset baskets. In doing so, it is instructive to begin with a degenerate case of projecting a single shifted Heston model on itself.

6.1 Error cancellation for a single process

The sequence of approximations developed in this work has several sources of systematic errors even though the starting point (Gyöngy's lemma) is an exact result for theoretical equivalence of European option pricing. Indeed, we first drastically reduced the full space of all target Markovian processes by assuming a shifted Heston form. After that we introduced additional errors by computing only the leading order in volatilities[8] of the effective parameters.

It is therefore important to show that the shifted Heston process itself is restored exactly as a result of the projection procedure even though the intermediate steps involve approximations.

[8]Typical values of volatilities are usually low, however very large instantaneous values can be achieved with appreciable probability in the process of stochastic evolution if the volatility-of-volatility is significant.

This suggests that a massive error cancellation is possible with the right choice of the target process.

We assume an initial process

$$dS(t) = (1+\Delta S(t)\beta_D(t))\sqrt{z(t)}\lambda(t)\cdot dW(t), \quad (49)$$

$$dz(t) = \alpha(t)\,(1-z(t))\,dt + \sqrt{z(t)}\,\gamma(t)\cdot dW(t), \quad z(0)=1, \quad (50)$$

where $\lambda(t)$ and $\gamma(t)$ are deterministic vector functions in the space of Brownian motion components. As it is written, Eq. (49) does not correspond to the separable form of Eq. (32) because the drift is not small. However, it is easy to eliminate the drift by a change of variables. Indeed, writing

$$z(t) = 1 + y(t)\,e^{-\int_0^t \alpha(s)\,ds},$$

we replace $z(t)$ by a martingale process $y(t)$,

$$dy(t) = e^{\int_0^t \alpha(s)\,ds}\sqrt{1+y(t)\,e^{-\int_0^t \alpha(s)\,ds}}\,\gamma(t)\cdot dW(t), \quad y(0)=0.$$

Expanding the square root in powers of $y(t)$, we obtain the desired decomposition

$$(1+\Delta S(t)\beta_D(t))\sqrt{z(t)}\lambda(t) = \sum_n a_n(t) X_n(t), \quad (51)$$

with the following basis processes and coefficients in the first three terms,

$$\begin{aligned} X_1(t) &= 1, & a_1(t) &= \lambda(t), \\ X_2(t) &= \Delta S(t), & a_2(t) &= \beta_D(t)\,\lambda(t), \\ X_3(t) &= y(t), & a_2(t) &= \frac{1}{2}\,e^{-\int_0^t \alpha(s)\,ds}\lambda(t). \end{aligned}$$

Higher terms correspond to basis processes $X_{n+2}(t) = y^n(t)$ with $n \geq 2$. It is easy to check that these do not contribute in the leading order to the parameters of the projected process. Using the definitions (34), (36), and (37), we obtain

$$\hat{\sigma}(t) = \lambda(t), \quad \hat{\sigma}_1(t) = 0, \quad \hat{\sigma}_2(t) = \lambda(t), \quad \hat{\sigma}_3(t) = e^{\int_0^t \alpha(s)\,ds}\gamma(t), \quad (52)$$

$$\Phi(t,\tau) = \sum_{n=1}^{3} a_n(t)\cdot\hat{\sigma}(t)\,\hat{\sigma}_n(\tau) = |\lambda(t)|^2\left(\beta_D(t)\,\lambda(\tau) + \frac{1}{2}\,e^{-\int_\tau^t \alpha(s)\,ds}\,\gamma(\tau)\right), \quad (53)$$

$$\Omega(t,\tau) = 2\,|\lambda(t)|^2\left((\beta_D(t)-\beta(t))\,\lambda(\tau) + \frac{1}{2}\,e^{-\int_\tau^t \alpha(s)\,ds}\gamma(\tau)\right). \quad (54)$$

A substitution into Eqs. (39)–(42) leads to the following result for the parameters of the projected process,

$$
\begin{aligned}
|\sigma_H(t)| &= |\lambda(t)|, \\
|\sigma_z(t)| &= 2\sqrt{(\beta_D(t)-\beta(t))^2\,|\lambda(t)|^2 + (\beta_D(t)-\beta(t))\,\lambda(t)\cdot\gamma(t) + \frac{1}{4}|\gamma(t)|^2}, \\
\rho(t) &= \frac{(\beta_D(t)-\beta(t))|\lambda(t)|^2 + \frac{1}{2}\lambda(t)\cdot\gamma(t)}{|\lambda(t)|\sqrt{(\beta_D(t)-\beta(t))^2\,|\lambda(t)|^2 + (\beta_D(t)-\beta(t))\,\lambda(t)\cdot\gamma(t) + \frac{1}{4}|\gamma(t)|^2}}, \\
\theta(t) &= -\frac{\int_0^t \frac{\partial}{\partial t}\left|(\beta_D(t)-\beta(t))\lambda(\tau) + \frac{1}{2}\,e^{-\int_\tau^t \alpha(s)\,ds}\gamma(\tau)\right|^2 d\tau}{2\int_0^t \left|(\beta_D(t)-\beta(t))\lambda(\tau) + \frac{1}{2}\,e^{-\int_\tau^t \alpha(s)\,ds}\gamma(\tau)\right|^2 d\tau}.
\end{aligned}
$$

We see that the original model is recovered exactly if we choose the right shift function $\beta(t) = \beta_D(t)$. Furthermore, it is sufficient to require that $\beta(0) = \beta_D(0)$ because Eq. (45) implies the equality of $\beta(t)$ and $\beta_D(t)$ for $t > 0$.

6.2 Several processes

We now turn to the case where the initial process $S(t)$ is a linear combination of two or more shifted Heston processes, that is

$$S(t) = \sum_i w_i\, S_i(t), \tag{55}$$

where each $S_i(t)$ is defined by the following pair of equations,

$$
\begin{aligned}
dS_i(t) &= (1 + \Delta S_i(t)\beta_i(t))\sqrt{z_i(t)}\lambda_i(t)\cdot dW, &\quad (56)\\
dz_i(t) &= \alpha_i(t)\,(1 - z_i(t))\,dt + \sqrt{z_i(t)}\,\gamma_i(t)\cdot dW(t), \quad z_i(0) = 1. &\quad (57)
\end{aligned}
$$

Applying the derivation of the previous subsection to each member of the sum, we obtain the following decomposition for the volatility of $S(t)$ after the omission of irrelevant higher order terms,

$$\sum_i \sum_{n=1}^{3} a_{ni}(t)\, X_{ni}(t) = \sum_i w_i \left(\lambda_i(t) + \Delta S_i(t)\,\beta_i(t)\lambda_i(t) + \frac{1}{2}\,y_i(t)\,e^{-\int_0^t \alpha_i(s)\,ds}\lambda_i(t)\right), \tag{58}$$

where the processes $y_i(t)$ follows the SDEs

$$dy_i(t) = e^{\int_0^t \alpha_i(s)\,ds}\sqrt{1 + y_i(t)\,e^{-\int_0^t \alpha_i(s)\,ds}}\,\gamma_i(t)\cdot dW(t), \quad y_i(0) = 0.$$

This immediately gives

$$\hat{\sigma}(t) = \sum_i w_i \lambda_i(t), \quad \hat{\sigma}_{1i}(t) = 0, \quad \hat{\sigma}_{2i}(t) = \sum_i w_i \lambda_i(t), \quad \hat{\sigma}_{3i}(t) = \frac{1}{2} w_i\, e^{-\int_0^t \alpha_i(s)\, ds}\, \gamma_i(t), \tag{59}$$

$$\Phi(t,\tau) = \sum_i w_i\, d_i(t) \left(\beta_i(t)\ \lambda_i(\tau) + \frac{1}{2}\, e^{-\int_\tau^t \alpha_i(s)\, ds}\, \gamma_i(\tau) \right), \tag{60}$$

$$\Omega(t,\tau) = 2 \sum_i w_i\, d_i(t) \left(\beta_i(t)\ \lambda_i(\tau) + \frac{1}{2}\, e^{-\int_\tau^t a_i(s)\, ds}\, \gamma_i(\tau) \right) - 2\beta(t)\, |\hat{\sigma}(t)|^2\, \hat{\sigma}(\tau), \tag{61}$$

where $d_i(t) = \lambda_i(t) \cdot \hat{\sigma}(t)$ (recall that dot denotes the scalar product in the factor space).

The parameters of the projection on an effective shifted Heston model are determined from Eqs. (39)–(42). The final result can be written in a vector form,

$$\sigma_H = \sum_i w_i \lambda_i(t), \tag{62}$$

$$\sigma_z(t) = 2 \frac{\sum_i w_i\, d_i(t) \left(\beta_i(t)\ \lambda_i(t) + \frac{1}{2}\, \gamma_i(t) \right)}{|\sigma_H(t)|^2} - 2\beta(t)\, \sigma_H(t), \tag{63}$$

$$\theta(t) = \left(\log |\sigma_H(t)|^2 \right)' - \frac{\int_0^t \frac{\partial}{\partial t} |\Omega(t,\tau)|^2\, d\tau}{2\, \int_0^t |\Omega(t,\tau)|^2\, d\tau}. \tag{64}$$

The shift function $\beta(t)$ is found as a solution of the linear ODE (45). As was mentioned above, the initial condition remains undetermined by our generic procedure and needs to be fixed from considerations of heuristic character. For the purpose of the numerical experiments presented in the next section, we chose $\beta(0)$ equal to the initial value of the shift function which corresponds to the the Markovian projection on a displaced-diffusion derived in Refs. [7] and [8],

$$\beta(0) = \frac{\sum_i w_i \beta_i(0)\, d_i(0)^2}{|\sigma_H(0)|^4}. \tag{65}$$

This formula immediately follows from Eqs. (35) and (60) after a substitution $\gamma_i(t) = 0$ and gives the right answer when the basket degenerate to a single Heston process. A more systematic investigation of the optimal choice of $\beta(0)$ is a subject for further research.

7 Numerical tests

We tested the accuracy of the projection formulas by comparing a simulation-based solution for European option in the original model with the analytical approximation. For the original model, we took a linear combination of several shifted Heston models with time-independent parameters, $\lambda_i(t) \equiv \lambda_i$, $\gamma_i(t) \equiv \gamma_i$, $\beta_i(t) \equiv \beta_i$, and $\alpha_i(t) \equiv \alpha_i$.

The general formulas for the optimal basket parameters (62)–(64) take the form

$$\sigma_H = \sum_i w_i \lambda_i, \tag{66}$$

$$\sigma_z(t) = 2\frac{\sum_i w_i\, d_i \left(\beta_i\ \lambda_i + \frac{1}{2}\,\gamma_i\right)}{|\sigma_H|^2} - 2\beta(t)\,\sigma_H, \tag{67}$$

$$\theta(t) = -\frac{\int_0^t \frac{\partial}{\partial t}|\Omega(t,\tau)|^2\, d\tau}{2\ \int_0^t |\Omega(t,\tau)|^2\, d\tau}, \tag{68}$$

where $d_i = \lambda_i{\cdot}\sigma_H$ and

$$\Omega(t,\tau) = 2\sum_i w_i\, d_i \left(\beta_i\ \lambda_i + \frac{1}{2}\,e^{-\alpha_i(t-\tau)}\,\gamma_i\right) - 2\beta(t)\,|\sigma_H|^2\,\sigma_H.$$

The optimal shift function $\beta(t)$ is found by solving the linear ODE (45) with the initial condition taken from the Markovian projection on a displaced diffusion (65)

$$\beta(0) = \frac{\sum_i w_i \beta_i\, d_i^2}{|\sigma_H|^4}.$$

We considered two cases: 1) European option on a spread of two rates and 2) European option on a basket of 5 assets.

7.1 Spread option

In this subsection, we present numerical results for a European option on a spread between a pair of rates $S_1(t)$, $S_2(t)$, which follow Heston processes with the following time-independent parameters

	Asset 1	Asset 2
initial value, $S_i(0)$	1	1
vol of rate, $\vert\lambda_i\vert$ (%)	10	9
shift parameter, β_i (%)	100	0
reversion, α_i (%)	10	10
vol-of-vol, $\vert\gamma_i\vert$ (%)	100	100
weight, w_i (%)	100	-90

The structure in the factor space is characterized by the following values of the correlations between the Brownian motions that drive the underlying rates and their stochastic volatilities,

$$\frac{\lambda_1{\cdot}\lambda_2}{|\lambda_1|\,|\lambda_2|} = 70\%,$$

$$\begin{aligned}\frac{\gamma_1 \cdot \gamma_2}{|\gamma_1|\,|\gamma_2|} &= 90\%, \\ \frac{\lambda_i \cdot \gamma_j}{|\lambda_i|\,|\gamma_j|} &= -25\%, \quad 1 \le i, j \le 2.\end{aligned}$$

The results for European option prices are given in terms of implied *normal* volatilities[9] computed analytically using the Markovian projection approximation (66)–(68), and compared with the direct simulation with a large number of paths. Strikes are expressed as percentages of spot.

[9]Let $\mathcal{C}$ be the price of a call option with strike K and maturity T. Its normal implied volatility σ_N is defined implicitly from a relationship $E[(\sigma_N\, W(T) + S(0) - K)^+] = \mathcal{C}$, where $W(T)$ is the standard one-dimensional Brownian motion. We use normal (Bachelier) rather than log-normal (Black) implied volatilities because the spread process can assume both positive and negative values.

strike (%)	sim vol (%)	MP vol (%)	error (%)
-100	7.95	7.92	-0.03
0	7.31	7.33	0.02
100	6.94	6.97	0.03
200	7.51	7.53	0.02
300	8.48	8.47	-0.01

Table 1: Normal implied volatilities for spread options with maturity 1Y.

strike (%)	sim vol (%)	MP vol (%)	error (%)
-100	7.33	7.28	-0.05
0	6.68	6.70	0.02
100	6.39	6.42	0.03
200	6.86	6.90	0.04
300	7.74	7.78	0.04

Table 2: Normal implied volatilities for spread options with maturity 5Y.

strike (%)	sim vol (%)	MP vol (%)	error (%)
-100	6.88	6.76	-0.12
0	6.41	6.36	-0.05
100	6.24	6.23	-0.01
200	6.59	6.59	0.00
300	7.28	7.27	-0.01

Table 3: Normal implied volatilities for spread options with maturity 10Y.

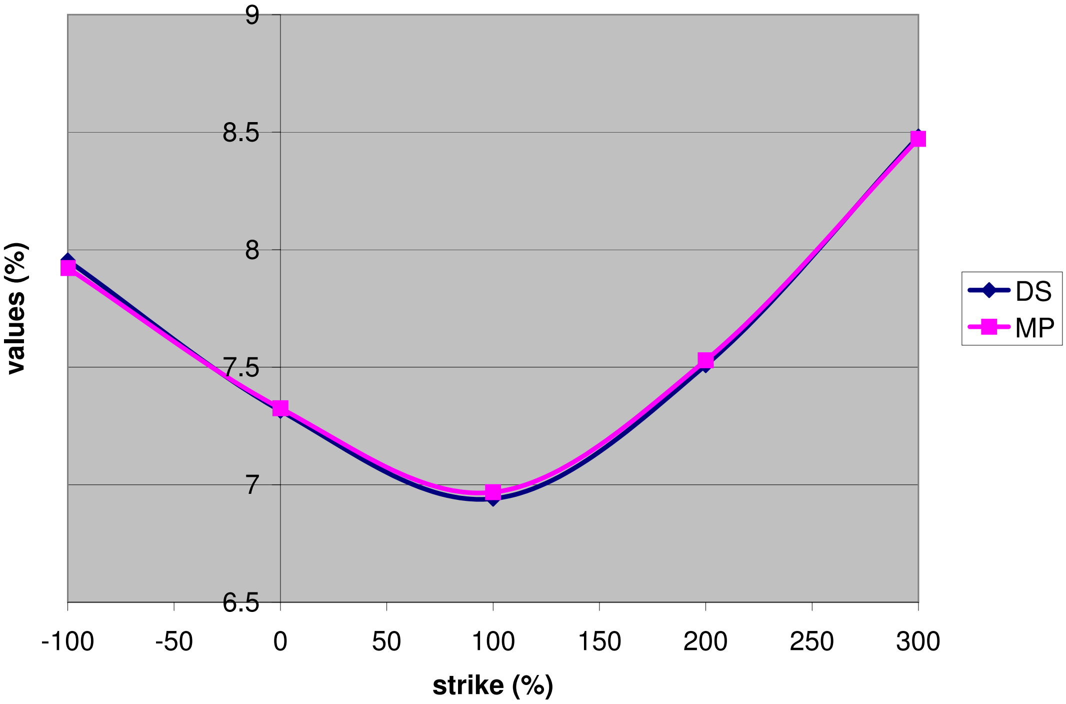

Figure 1: Normal implied volatilities for spread options with maturity 1Y, computed using a direct simulation (DS) and Markovian projection approximations (MP).

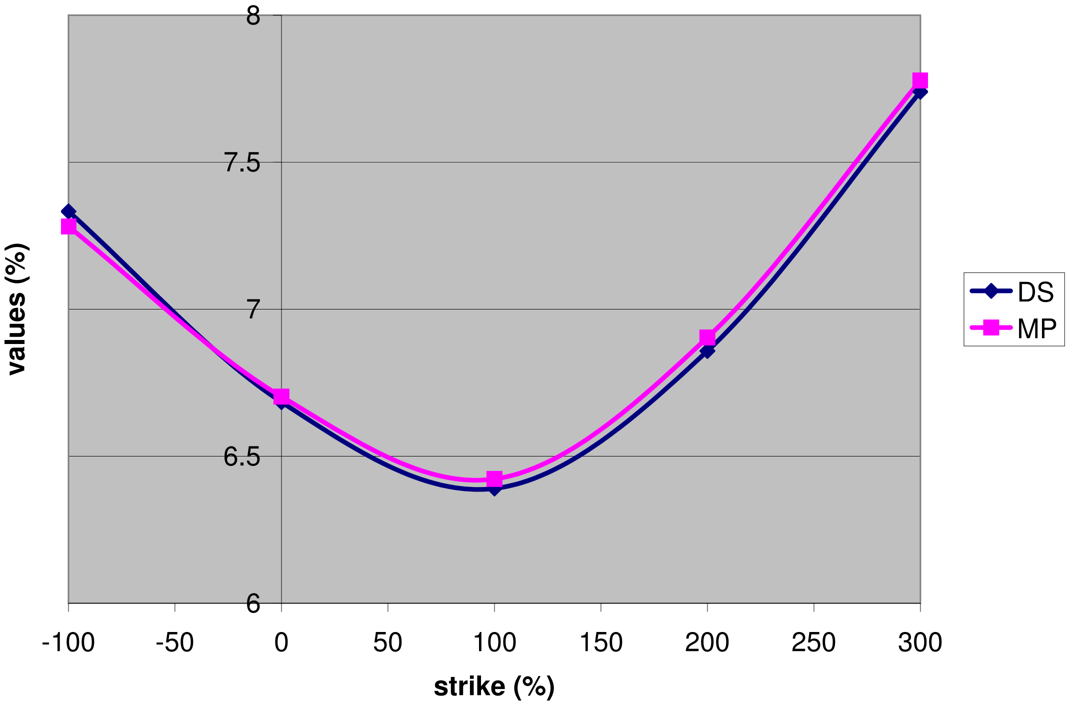

Figure 2: Normal implied volatilities for spread options with maturity 5Y, computed using a direct simulation (DS) and Markovian projection approximations (MP).

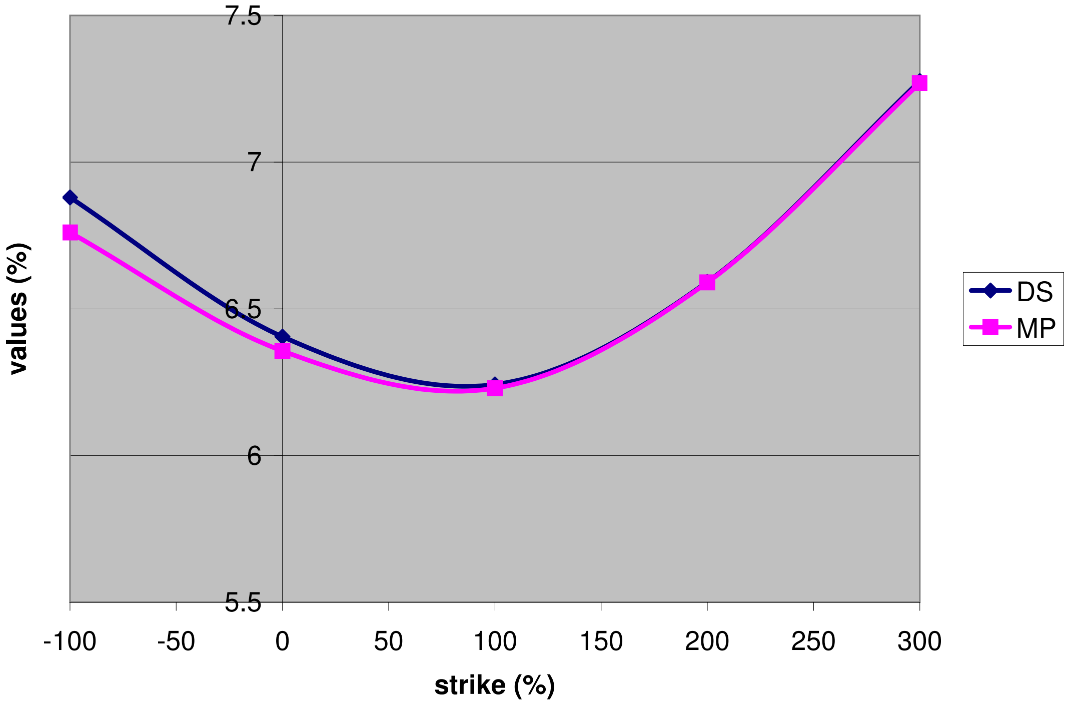

Figure 3: Normal implied volatilities for spread options with maturity 10Y, computed using a direct simulation (DS) and Markovian projection approximations (MP).

7.2 Basket of 5 assets

As a second example, we consider a basket of 5 assets with the following Heston model parameters

	Asset 1	Asset 2	Asset 3	Asset 4	Asset 5
initial value, $S_i(0)$	1	1	1	1	1
vol of rate, $\lvert\lambda_i\rvert$ (%)	14	15	16	17	18
beta skew, β_i (%)	30	40	50	60	70
reversion, α_i (%)	10	10	10	10	10
vol-of-vol, $\lvert\gamma_i\rvert$ (%)	70	75	80	85	90
weight, w_i (%)	20	20	20	20	20

The structure in the factor space is characterized by the following values of the correlations between the Brownian motions that drive the underlying rates and their stochastic volatilities,

$$
\begin{aligned}
\frac{\lambda_i \cdot \lambda_j}{|\lambda_i|\,|\lambda_j|} &= 70\%, \quad 1 \le i, j \le 5, \quad i \neq j, \\
\frac{\gamma_i \cdot \gamma_j}{|\gamma_i|\,|\gamma_j|} &= 90\%, \quad 1 \le i, j \le 5, \quad i \neq j, \\
\frac{\lambda_i \cdot \gamma_j}{|\lambda_i|\,|\gamma_j|} &= -20\%, \quad 1 \le i, j \le 5.
\end{aligned}
$$

The results for European option prices are given in terms of standard implied *log-normal* volatilities, computed analytically using the Markovian projection approximation (66)–(68), and compared with the direct simulation with a large number of paths. Strikes are expressed as percentages of spot.

strike (%)	sim vol (%)	MP vol (%)	error (%)
70	17.10	17.16	0.06
80	15.67	15.73	0.06
90	14.47	14.52	0.05
100	13.56	13.61	0.05
110	13.06	13.10	0.04
120	12.93	12.97	0.04
130	13.04	13.04	0.00

Table 4: Log-normal implied volatilities for basket options with maturity 1Y.

strike (%)	sim vol (%)	MP vol (%)	error (%)
50	19.38	19.51	0.13
60	17.47	17.63	0.16
80	14.51	14.69	0.18
100	12.59	12.77	0.18
120	12.00	12.16	0.16
150	12.45	12.56	0.11
190	13.34	13.37	0.03

Table 5: Log-normal implied volatilities for basket options with maturity 5Y.

strike (%)	sim vol (%)	MP vol (%)	error (%)
40	20.04	20.42	0.38
50	17.91	18.26	0.35
70	14.83	15.14	0.31
100	12.15	12.40	0.25
130	11.50	11.71	0.21
180	12.20	12.33	0.13
250	13.34	13.34	0.00

Table 6: Log-normal implied volatilities for basket options with maturity 10Y. Comparison between direct simulations (DS) and the Markovian Projection approximation (MP).

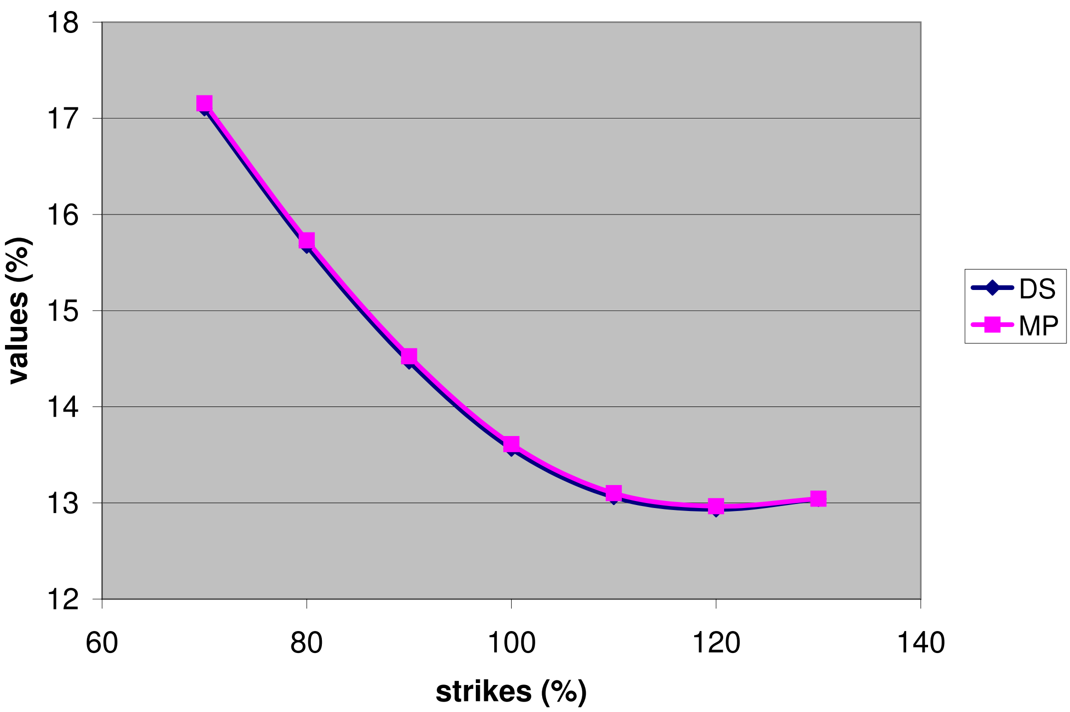

Figure 4: Log-normal implied volatilities for basket options with maturity 1Y, computed using a direct simulation (DS) and Markovian projection approximations (MP).

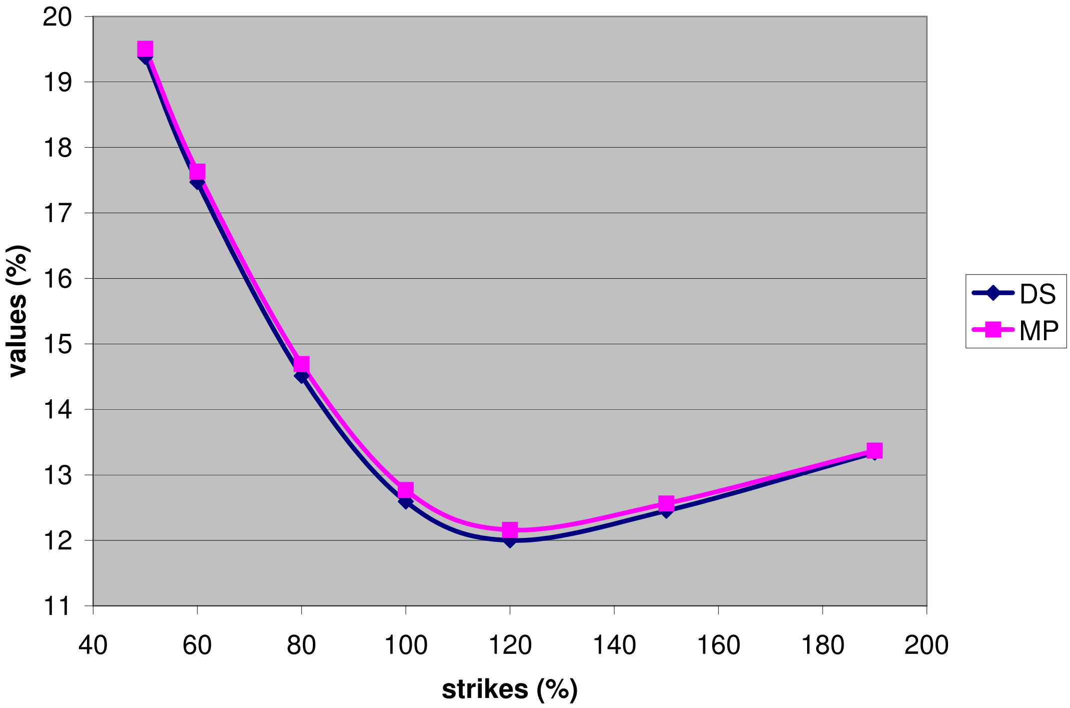

Figure 5: Log-normal implied volatilities for basket options with maturity 5Y, computed using a direct simulation (DS) and Markovian projection approximations (MP).

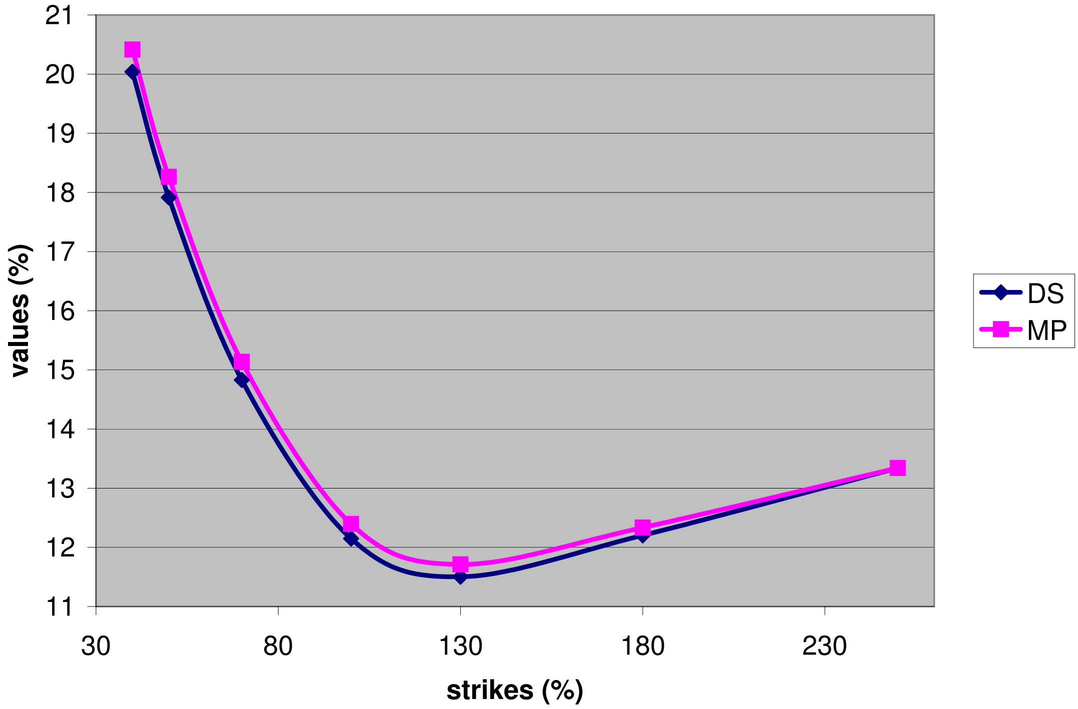

Figure 6: Log-normal implied volatilities for basket options with maturity 10Y, computed using a direct simulation (DS) and Markovian projection approximations (MP).

8 Conclusions and future directions

We presented a systematic approach to deriving accurate approximations for European option prices in models with complicated, multi-dimensional stochastic volatility dynamics. The method is based on Markovian projection on a shifted Heston model, preserving the joint distribution of the underlying and its stochastic variance. Closed-form results are obtained in the leading order in volatilities for a wide class of processes. Numerical examples for spread and asset basket options show good accuracy and faithful reproduction of implied volatility smiles. This work extends both the work on optimal projections onto displaced diffusion processes [1] and Markovian projection in the presence of an exogenous stochastic volatility process [8]. By incorporating stochastic volatility directly into the method, the values of European options across a wide range of strikes are obtained quickly and accurately.

Further possible applications of our methodology include the calibration of forward LIBOR models with multi-dimensional stochastic volatility and FX- or equity-linked interest rate hybrid models with stochastic volatility for multiple underlyings.[10]

There are several directions for the development of the method. As always, the calculation of conditional expected values presents a major numerical challenge of any method based on the Gyöngy result. Any advance in this regard would directly affect the quality of the approximations. In this work we compute conditional expectations by minimizing regression functionals in a special class of functions. A wider class of target functions can make an improvement.

AA is indebted to Serguei Mechkov for stimulating discussions and to his colleagues at NumeriX and especially to Gregory Whitten for supporting this work.

A Calculations for the optimal coefficients

In this appendix we derive the key formulas of Sect. 4.

Minimization of the regression functionals (22) and (23) gives a pair of equations

$$|\sigma_H(t)|^2\,|\sigma_z(t)|^2 = \frac{E[V(t)|\sigma_V(t)|^2]}{E[V^2(t)]}, \tag{69}$$

$$|\sigma_H(t)|\,|\sigma_z(t)|\rho(t) = \frac{E[V(t)\Lambda(t)\cdot\sigma_V(t)]}{E[V^2(t)]}, \tag{70}$$

which immediately lead to Eq. (27) for the correlation $\rho(t)$.

The functional (21) is minimized if

$$\begin{aligned} E[\mu_V(t)] &= A(t)\,E[V(t)] + B(t), \\ E[V(t)\mu_V(t)] &= A(t)\,E[V^2(t)] + B(t)\,E[V(t)], \end{aligned} \tag{71}$$

[10]For example, it can be used to justify the derivation of an effective Ornstein-Uhlenbeck process for the effective swaption volatility proposed in Ref. [9].

where

$$\begin{aligned}
A(t) &= \left(\log|\sigma_H(t)|^2\right)' - \theta(t), \\
B(t) &= \theta(t)\,|\sigma_H(t)|^2.
\end{aligned}$$

The system of linear equations (71) can be solved for $A(t)$ and $B(t)$, which subsequently allows to determine $|\sigma_H(t)|$ and $\theta(t)$. The solution reads

$$\begin{aligned}
A(t) &= \frac{E[V(t)\mu_V(t)] - E[\mu_V(t)]\,E[V(t)]}{E[V^2(t)] - E[V(t)]^2} = \frac{E[\delta V(t)\delta\mu_V(t)]}{E[(\delta V(t))^2]} \\
B(t) &= E[\mu_V(t)] - A(t)\,E[V(t)].
\end{aligned}$$

It is possible to verify Eqs. (24)–(25) for $|\sigma_H(t)|^2$ and $\theta(t)$ by a direct substitution, using the relationship $E[\mu_V(t)] = dE\left[V(t)\right]/dt$. Another useful equality

$$E[\delta V(t)\,\delta\mu_V(t)] = \frac{1}{2}\left(\frac{dE\left[(\delta V(t))^2\right]}{dt} - E[|\sigma_V(t)|^2]\right),$$

can be obtained by averaging the SDE for $(\delta V(t))^2$.

B Calculations for the separable process

The difficult part is to prove the expressions (39) and (40) for the volatility $\sigma_H(t)$ and the mean reversion $\theta(t)$ by computing the averages in Eqs. (24)–(25) in the leading order in volatilities.

We start by assuming a separable form for the stochastic variance process,

$$V(t) = \sum_n b_n(t)\,Y_n(t), \tag{72}$$

where $b_n(t)$ are deterministic functions such that $b_n(t) = O(\sigma^2)$, and the processes $Y_n(t)$ follow the SDEs

$$dY_n = \nu_n(t;Y(t))\,dt + \gamma_n(t;Y(t))\cdot dW(t)$$

with $\nu_n(t) = O(\sigma^2)$ and $\gamma_n(t) = O(\sigma)$. The volatility process $\sigma_V(t)$ can be expressed as

$$\sigma_V(t) = \sum_n b_n(t)\,\gamma_n(t;Y(t)).$$

We will show that this is not a new assumption because a separable form for the variance $V(t)$ follows from the definition of the separable underlying process $S(t)$. In fact, the basis processes $Y_n(t)$ can be identified with pairwise products $X_{n1}(t)X_{n2}(t)$.

To simplify the formulas for the averages, we introduce a kernel

$$\Omega(t,\tau) = \sum_n b_n(t)\, \hat{\gamma}_n(\tau), \tag{73}$$

with $\hat{\gamma}_n(t) \equiv \gamma_n(\tau; Y(0))$. It turns out that all necessary averages can be expressed in terms of the kernel,

$$\begin{aligned} E\left[(\delta V(t))^2\right] &= \int_0^t |\Omega(t,\tau)|^2\, d\tau + O(\sigma^8), & (74)\\ E[\delta V(t)\delta\mu(t)] &= \int_0^t \Omega_t'(t,\tau)\cdot \Omega(t,\tau)\, d\tau + O(\sigma^8), & (75)\\ E\left[(\delta\mu(t))^2\right] &= \int_0^t \left|\Omega_t'(t,\tau)\right|^2\, d\tau + O(\sigma^8). & (76) \end{aligned}$$

To prove these formulas, we notice that the increments $\delta Y_n = Y_n - E[Y_n]$ obey the following SDEs

$$d\,\delta Y_n = (\nu_n(t) - E[\nu_n(t)])\ dt + \gamma_n(t; Y(t))\cdot dW(t).$$

It follows that the increment of the variance $\delta V(t) = \sum_n b_n(t)\, \delta Y_n(t)$ satisfies the SDE

$$d\,\delta V(t) = \left(\sum_n b_n'(t)\, \delta Y_n(t) + b_n(t)\, (\nu_n(t) - E[\nu_n(t)])\right) dt + \sigma_V(t)\cdot dW(t),$$

which leads to the following expression for the increment of the drift,

$$\delta\mu_V(t) = \sum_n b_n'(t)\, \delta Y_n(t) + b_n(t)\, (\nu_n(t) - E[\nu_n(t)])\,. \tag{77}$$

To derive (74)–(76), we need to calculate the expectations $E[\delta Y_n(t)\delta Y_m(t)]$, $E[\delta Y_n(t)\, \nu_n(t)]$ and $E[(\nu_n(t) - E[\nu_n(t)])^2]$. The last two expectations are of the order $O(\sigma^4)$ and the first one is given by

$$E[\delta Y_n(t)\delta Y_m(t)] = \int_0^t \hat{\gamma}_n(\tau)\cdot\hat{\gamma}_m(\tau)\, d\tau + O(\sigma^4),$$

which leads to Eqs. (74)–(76). After a substitution into Eqs. (24)–(25), we get the optimal coefficients $|\sigma_H(t)|$ and $\theta(t)$

$$\begin{aligned} |\sigma_H(t)|^2 &= \sum_n b_n(t)\, Y_n(0) + O(\sigma^4), \\ \theta(t) &= \left(\log|\sigma_H(t)|^2\right)' - \frac{\int_0^t \frac{\partial}{\partial t}|\Omega(t,\tau)|^2\, d\tau}{2\int_0^t |\Omega(t,\tau)|^2\, d\tau} + O(\sigma^2). & (78) \end{aligned}$$

To finish the proof, we show that for a separable underlying process the variance can be expressed in the form (72). Indeed,

$$\begin{aligned} V(t) &= |\Lambda(t)|^2 = \sum_{n,m} a_n(t) \cdot a_m(t) \left(X_n(t)\, X_m(t) \frac{1}{(1+\beta(t)\,\Delta S(t))^2} \right) \\ &= \sum_{n,m} a_n(t) \cdot a_m(t) X_n(t)\, X_m(t) - 2\beta(t) \sum_{n,m} a_n(t) \cdot a_m(t) X_n(t)\, X_m(t)\, \Delta S(t) + \cdots . \end{aligned}$$

The omitted terms denoted by the dots contain higher powers in ΔS, with volatilities of the order higher than σ and drifts of the order higher than σ^2. For example, the volatility of $X_n(t)\, X_m(t)\, \Delta S^2(t)$ is $O(\sigma^3)$. In this way, we recover the form (72) for the variance. The kernel (73) can now be written in the form

$$\Omega(t,\tau) = \sum_{n,m} a_n(t) \cdot a_m(t) \left(X_n(0)\, \hat{\sigma}_m(\tau) + X_m(0)\, \hat{\sigma}_n(\tau) - 2\beta(t)\, X_n(0) X_m(0) \hat{\sigma}(\tau) \right) + O(\sigma^5), \quad (79)$$

and further simplified to Eq. (38) by a rearrangement of summations. Eq. (40) for $\theta(t)$ is obtained using the formula (78) and the expression for the effective volatility $|\sigma_H(t)|^2 = |\hat{\sigma}(t)|^2$.

To prove expressions (41) and (42) for the optimal volatility-of-volatility $\sigma_z(t)$ and the correlation $\rho(t)$, it is sufficient to evaluate Eqs. (26) and (27) in the leading order, that is to replace the underlying averages by their values corresponding to the initial values of X_n and S

$$V(t) \to |\hat{\sigma}(t)|^2, \qquad \sigma_V(t) \to \Omega(t,t).$$

References

[1] A. Antonov and T. Misirpashaev (2006) "Markovian Projection onto a Displaced Diffusion: Generic Formulas with Applications," available at SSRN: http://ssrn.com/abstract=937860.

[2] M. Avellaneda, D. Boyer-Olson, J. Busca, and P. Friz (2002) "Reconstructing Volatility," Risk, October, 87–91.

[3] M. Britten-Jones and A. Neuberger (2000) "Option Prices, Implied Price Processes, and Stochastic Volatility," *Journal of Finance*, 55, 839–866.

[4] B. Dupire (1997) "A Unified Theory of Volatility," Banque Paribas working paper, reprinted in "Derivatives Pricing: the Classic Collection," edited by P. Carr, Risk Books, London, 2004.

[5] I. Gyöngy (1986) "Mimicking the One-Dimensional Marginal Distributions of Processes Having an Ito Differential" *Probability Theory and Related Fields*, 71, 501–516.

[6] V. Piterbarg (2005) "Time to Smile," *Risk*, May, 18(5), 71–.75

[7] V. Piterbarg (2006) "Markovian Projection Method for Volatility Calibration," available at SSRN: http://ssrn.com/abstract=906473.

[8] V. Piterbarg (2007) "Markovian Projection for Volatility Calibration," *Risk*, April, 20(4), 84–89.

[9] J. Zhu (2007) "An Extended Libor Market Model with Nested Stochastic Volatility Dynamics," available at: http://workshop.mathfinance.de/2007/papers/zhu/paper.pdf